RODOLPHO RAPHAEL

DEMOCRACIA HACKEADA

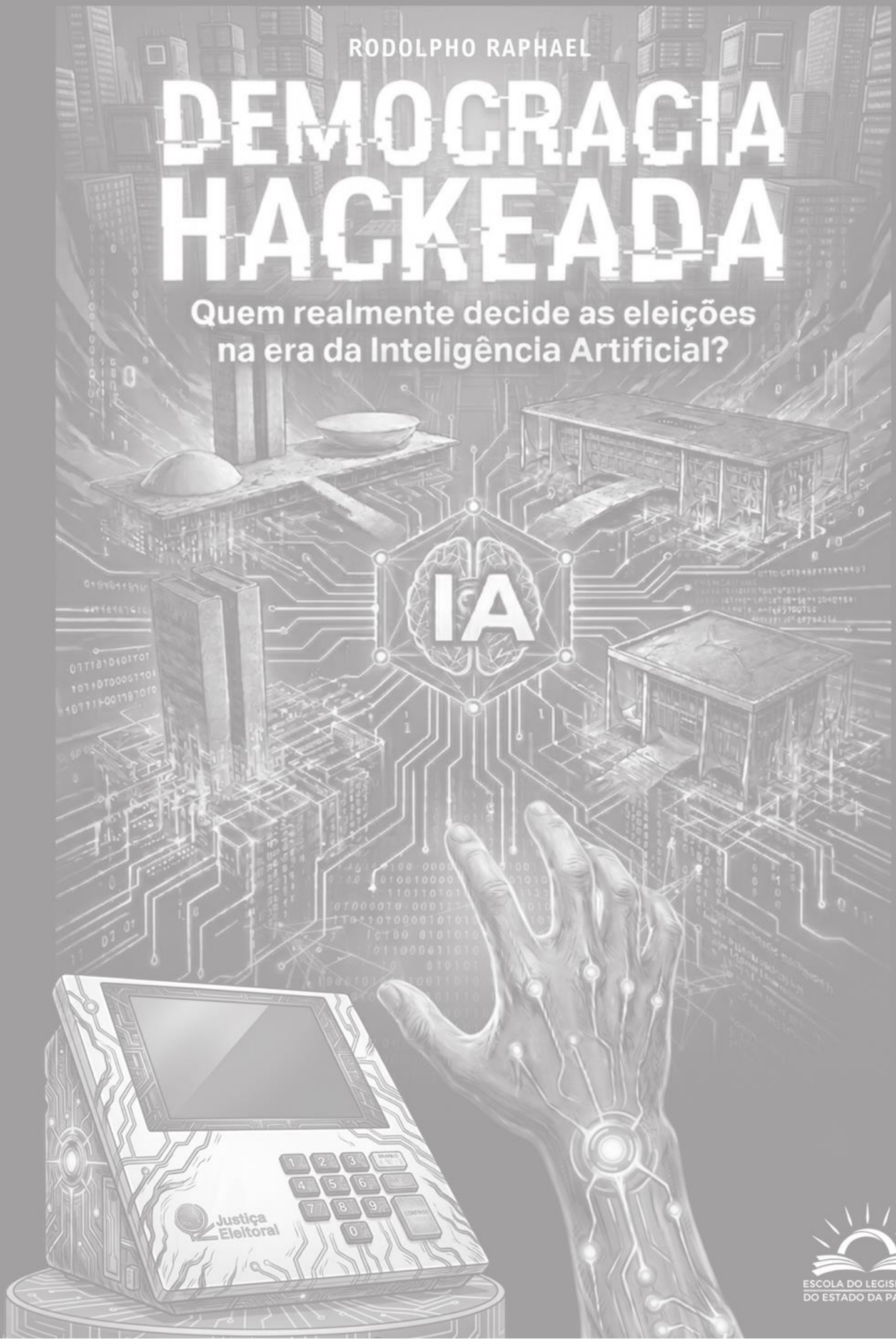
Quem realmente decide as eleições na era da Inteligência Artificial?

ESCOLA DO LEGISLATIVO
DO ESTADO DA PARAÍBA

RODOLPHO RAPHAEL

DEMOCRACIA HACKEADA

Quem realmente decide as eleições na era da Inteligência Artificial?



ESCOLA DO LEGISLATIVO
DO ESTADO DA PARAÍBA

2026 by Escola do Legislativo da Paraíba
Copyright © Escola do Legislativo da Paraíba
Copyright do Texto © 2026 Rodolpho Raphael
Copyright da Edição © 2026 Escola do Legislativo da Paraíba
Direção Geral: Prof. Maria Helena Toscano Moura de Caldas Barros

Diagramação: Prof. Rodolpho Raphael
Edição de Arte: Prof. Rodolpho Raphael

Revisão: O autor

Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição Creative Commons. Atribuição 4.0 Internacional (CC BY 4.0).

O conteúdo deste livro e seus dados em sua forma, correção e confiabilidade é de responsabilidade exclusiva dos autores. Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

A organização de referências, revisão de texto, apoio na análise de dados, programação e automação de tarefas foram realizadas por Inteligência Artificial Generativa conforme regulamenta a Portaria CNPq Nº 2.664 de 06 de março de 2026.

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Raphael, Rodolpho
Democracia hackeada [livro eletrônico] : quem realmente decide as eleições na era da inteligência artificial? / Rodolpho Raphael. -- 1. ed. -- João Pessoa, PB : Ed. do Autor, 2026.
PDF

ISBN 978-65-02-02203-0

1. Comunicação e política 2. Democracia 3. Política - Aspectos sociais 4. Redes sociais on-line - Aspectos políticos I. Título.

26-348479.0 CDD-321.8

Índices para catálogo sistemático:

1. Democracia : Ciências políticas 321.8

Aline Graziele Benitez - Bibliotecária - CRB-1/3129

LISTA DE NOMENCLATURAS, SIGLAS & ABREVIATURAS

Sigla / Termo	Definição
AIJE	Ação de Investigação Judicial Eleitoral
AIME	Ação de Impugnação de Mandato Eletivo
AI Act	Lei da Inteligência Artificial da União Europeia (Regulamento UE 2024/1689)
ANPD	Autoridade Nacional de Proteção de Dados
API	Application Programming Interface (Interface de Programação de Aplicativos)
BERTimbau	Modelo de linguagem BERT pré-treinado para o português brasileiro, desenvolvido pela USP
C2PA	Coalition for Content Provenance and Authenticity — padrão técnico de autenticação criptográfica de conteúdo audiovisual
CIEDDE	Centro Integrado de Enfrentamento à Desinformação e Defesa da Democracia Eleitoral (TSE)
CRM	Customer Relationship Management — sistema de gestão de relacionamento com o público
CSV	Comma-Separated Values — formato de arquivo de texto para dados tabulares
DALL-E	Sistema de geração de imagens por IA desenvolvido pela OpenAI
DPO	Data Protection Officer — Encarregado de Proteção de Dados (art. 41 da LGPD)
DSA	Digital Services Act — Lei de Serviços Digitais da União Europeia (Regulamento UE 2022/2065)
FTC	Federal Trade Commission — órgão regulador norte-americano de comércio e proteção do consumidor
GAN	Generative Adversarial Network — Rede Generativa Adversarial, arquitetura de IA para geração de conteúdo sintético
GPT	Generative Pre-trained Transformer — família de modelos de linguagem de grande escala desenvolvidos pela OpenAI
GPU	Graphics Processing Unit — unidade de processamento gráfico utilizada em treinamento de modelos de IA
IA	Inteligência Artificial
LGPD	Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018)
LLM	Large Language Model — Modelo de Linguagem de Grande Escala
MPE	Ministério Público Eleitoral
NER	Named Entity Recognition — Reconhecimento de Entidades Nomeadas, tarefa de Processamento de Linguagem Natural

Sigla / Termo	Definição
NLP / PLN	Natural Language Processing / Processamento de Linguagem Natural — área da IA voltada à análise e geração de texto em linguagem humana
PL	Projeto de Lei
POS	Part-of-Speech — classe gramatical de palavras, usada em tarefas de NLP
RCED	Recurso Contra Expedição de Diploma
rPPG	Remote Photoplethysmography — técnica biométrica que detecta padrão cardíaco por variação de cor da pele em vídeo
SMS	Short Message Service — serviço de mensagens de texto por telefonia celular
STF	Supremo Tribunal Federal
STJ	Superior Tribunal de Justiça
Token	Unidade básica de processamento em modelos de linguagem, aproximadamente equivalente a uma palavra ou parte de palavra
TRE	Tribunal Regional Eleitoral
TSE	Tribunal Superior Eleitoral
UE	União Europeia
VLOP	Very Large Online Platform — Plataforma Online de Grande Porte (classificação do DSA europeu, ≥ 45 milhões de usuários mensais na UE)

SUMÁRIO

12
A ELEIÇÃO QUE VOCÊ NÃO VIU ACONTECER: COMO A INTELIGÊNCIA ARTIFICIAL OPERA NOS BASTIDORES DAS DISPUTAS ELEITORAIS

40
COMO OS ALGORITMOS PASSARAM A DECIDIR O DEBATE PÚBLICO: BOLHAS DE FILTRO, CÂMARAS DE ECO E A ARQUITETURA DA POLARIZAÇÃO

69
O BRASIL ESTÁ PREPARADO PARA ISSO? CAPACIDADE INSTITUCIONAL, MARCO REGULATÓRIO E VULNERABILIDADES ESTRUTURAIS

95
A MÁQUINA DE PERSUADIR: MICROTARGETING, PSICOLOGIA COMPORTAMENTAL E OS LIMITES ÉTICOS E JURÍDICOS DA PERSONALIZAÇÃO ELEITORAL

121
CAMBRIDGE ANALYTICA: ANATOMIA DE UM ESCÂNDALO, PARADIGMA GLOBAL E IMPLICAÇÕES PARA O DIREITO ELEITORAL BRASILEIRO

144
BOTS, CHATBOTS E AVATARES DIGITAIS: AUTOMAÇÃO POLÍTICA, SIMULAÇÃO DE PRESENÇA E OS LIMITES DA IDENTIDADE DIGITAL NAS ELEIÇÕES

163
DESINFORMAÇÃO, DEEPFAKES E CONTEÚDO SINTÉTICO: TECNOLOGIAS DE SÍNTESE, CRISE DA AUTENTICIDADE E O DIREITO ELEITORAL NA ERA DA PÓS-AUTENTICIDADE

198
AS PLATAFORMAS QUE LUCRAM COM A POLARIZAÇÃO: MODELOS DE NEGÓCIO, INCENTIVOS ALGORÍTMICOS E RESPONSABILIDADE ESTRUTURAL

213
OS FINANCIADORES INVISÍVEIS DA DESINFORMAÇÃO: ATORES, MOTIVAÇÕES, FINANCIAMENTO E A ECOLOGIA DA DESINFORMAÇÃO ELEITORAL BRASILEIRA

226
O BRASIL NAS ELEIÇÕES DE 2018, 2022 E 2024: CASOS CONCRETOS, PADRÕES EMERGENTES E LIÇÕES PARA 2026

239
A RESOLUÇÃO TSE Nº 23.732/2024: ANÁLISE DO PRIMEIRO MARCO NORMATIVO MUNDIAL SOBRE IA EM ELEIÇÕES

265
CRIMES ELEITORAIS DIGITAIS E SANÇÕES: TIPIFICAÇÃO, RESPONSABILIZAÇÃO E A JURISPRUDÊNCIA EMERGENTE DO TSE

291
LGPD E PROTEÇÃO DE DADOS NAS ELEIÇÕES: TRATAMENTO DE DADOS PESSOAIS, CONSENTIMENTO E A ARTICULAÇÃO ENTRE ANPD E JUSTIÇA ELEITORAL

315
FERRAMENTAS PRÁTICAS DE DETECÇÃO E MONITORAMENTO: DEEPFAKES, DESINFORMAÇÃO, ANÁLISE DE SENTIMENTO E AUDITORIA DO AMBIENTE ELEITORAL DIGITAL

340
O QUE CANDIDATOS, ASSESSORES E PARTIDOS PRECISAM SABER: GUIA OPERACIONAL PARA CONFORMIDADE, ESTRATÉGIA E GESTÃO DE RISCOS NA ERA DA IA ELEITORAL

360
LITERACIA MIDIÁTICA COMO DEFESA DEMOCRÁTICA: EDUCAÇÃO, CIDADANIA DIGITAL E A CONSTRUÇÃO DE RESILIÊNCIA CONTRA A MANIPULAÇÃO ALGORÍTMICA

376
INTELIGÊNCIA ARTIFICIAL DE PRÓXIMA GERAÇÃO: TENDÊNCIAS TECNOLÓGICAS E SEUS IMPACTOS SOBRE O PROCESSO ELEITORAL

402
É POSSÍVEL SALVAR A DEMOCRACIA? PROPOSTAS, CAMINHOS E O IMPERATIVO DA AÇÃO COLETIVA NA ERA DA INTELIGÊNCIA ARTIFICIAL

APRESENTAÇÃO

A Democracia que não sabe que está sendo hackeada!

Este livro começa com uma pergunta que parece simples e é profundamente perturbadora: quem realmente decide as eleições na era da inteligência artificial? A resposta óbvia é os eleitores, naturalmente. O perturbador está na descoberta gradual, construída ao longo de dezoito capítulos, de que o processo pelo qual eleitores formam suas preferências, escolhem candidatos e exercem o voto foi transformado de maneira tão silenciosa quanto radical. Não por golpes abruptos nem por fraudes grosseiras detectáveis pela fiscalização eleitoral tradicional, mas por uma arquitetura de influência que opera precisamente abaixo do limiar da percepção consciente da maioria dos cidadãos que a experimenta.

Democracia Hackeada nasce de uma constatação que qualquer operador do direito eleitoral brasileiro, seja advogado, assessor de campanha, membro de tribunal ou pesquisador, que acompanhou de perto os ciclos eleitorais de 2018, 2022 e 2024 não pode mais ignorar: as categorias jurídicas, os instrumentos regulatórios e os processos institucionais disponíveis para proteger a integridade do processo eleitoral foram construídos para um mundo que já não existe. Um mundo em que produzir desinformação em escala exigia recursos significativos, em que deepfakes de candidatos eram privilégio de atores sofisticados com capacidade técnica especializada, em que a responsabilização por disparos em massa dependia de rastros financeiros rastreáveis, e em que a linha entre campanha legítima e manipulação ilícita podia ser traçada com os instrumentos analíticos disponíveis na legislação eleitoral de 1965.

Esse mundo acabou em novembro de 2022, quando o lançamento do ChatGPT tornou acessível a centenas de milhões de usuários capacidades que até então estavam confinadas a laboratórios de pesquisa e a empresas com orçamentos de IA de centenas de milhões de dólares. O que se seguiu, nos dezoito meses que antecederam as eleições municipais brasileiras de 2024, foi a democratização radical das ferramentas de produção de conteúdo sintético: vídeos que fazem candidatos dizer o que nunca disseram, áudios que clonam vozes com poucos segundos de referência, textos personalizados psicologicamente para explorar as vulnerabilidades específicas de cada segmento do eleitorado, e chatbots que simulam conversas com representantes de campanhas sem

qualquer indicação de que são sistemas automatizados. Tudo isso acessível a qualquer ator com conexão à internet e disposição para usá-lo.

O Tribunal Superior Eleitoral respondeu com o que é, ao tempo desta publicação, o marco normativo mais específico do mundo sobre inteligência artificial em eleições: a Resolução TSE nº 23.732/2024, editada em março de 2024 e testada nas eleições municipais de outubro daquele ano. Este livro analisa esse marco regulatório com o rigor que seu caráter pioneiro exige, reconhecendo seus avanços genuínos e identificando, com honestidade analítica que a urgência do tema demanda, as lacunas que permanecem. A Resolução foi um avanço necessário. Ela não é suficiente. E o Brasil precisará de uma revisão regulatória de segunda geração antes das eleições gerais de 2026, quando o escrutínio será incomparavelmente maior e as consequências de falhas regulatórias incomparavelmente mais graves.

Esta obra foi escrita para três públicos simultaneamente, e sua estrutura reflete essa escolha deliberada. Para profissionais de campanha e consultores de marketing político, ela oferece o mapa de riscos e oportunidades que permite operar com eficácia dentro dos limites que protegem o processo eleitoral, sem ingenuidade sobre a realidade das práticas disponíveis, mas também sem cumplicidade com condutas que comprometem a legitimidade do resultado. Para advogados eleitorais, ela constrói o arcabouço doutrinário e jurisprudencial que o campo ainda está elaborando: como tipificar condutas eleitorais habilitadas por IA usando tipos penais de 1965, como atribuir responsabilidade em cadeias de produção deliberadamente opacas, como provar em tempo útil algo que se disseminou em horas. E para cidadãos interessados na qualidade da democracia brasileira, ela oferece o enquadramento conceitual que permite compreender o que está verdadeiramente em jogo nas escolhas tecnológicas e regulatórias que determinam que tipo de democracia o Brasil terá em 2036.

A estrutura da obra segue a lógica de uma construção cumulativa, não de capítulos independentes. Os primeiros capítulos estabelecem os fundamentos: o campo da política computacional, os mecanismos de microtargeting, a arquitetura do capitalismo de vigilância, o caso Cambridge Analytica como estudo de caso paradigmático e as formas de automação política que vão do bot político simples ao avatar digital sintético. Os capítulos intermediários aprofundam os fenômenos específicos: desinformação e deepfakes, as plataformas que lucram com a polarização, os financiadores invisíveis da desinformação, os ciclos eleitorais brasileiros de 2018, 2022 e 2024. Os capítulos finais constroem a resposta institucional e prática: a análise detalhada da Resolução TSE nº 23.732/2024, os crimes eleitorais digitais e o sistema sancionatório, a articulação entre LGPD e direito

eleitoral, as ferramentas de detecção e monitoramento, o guia operacional para conformidade e gestão de riscos, a literacia midiática como defesa democrática, as tendências de IA de próxima geração e, por fim, a pergunta que percorre toda a obra: é possível salvar a democracia?

A tese central desta obra pode ser enunciada com precisão: a inteligência artificial generativa introduz uma descontinuidade qualitativa, não apenas quantitativa, nos riscos para a integridade eleitoral democrática. Não porque a manipulação política seja nova, pois a propaganda existe desde que a política existe, mas porque a combinação de escala industrial, personalização psicológica e imperceptibilidade que as ferramentas de IA tornam possíveis cria condições para a distorção do processo de formação de preferências políticas em magnitude que nenhuma tecnologia anterior havia alcançado. Essa descontinuidade qualitativa exige resposta regulatória, institucional e educacional que vai além do aperfeiçoamento incremental dos instrumentos existentes.

O método desta obra combina análise doutrinária com análise empírica e análise técnica, e essa combinação não é decorativa: é a condição para que qualquer uma das três dimensões produza diagnóstico útil. Uma análise jurídica que não compreende o que está regulando tende a ser simultaneamente excessiva em relação a práticas legítimas e insuficiente em relação às práticas que efetivamente ameaçam a integridade eleitoral. Uma análise técnica que não conhece os instrumentos jurídicos disponíveis produz diagnósticos sem resposta. E uma análise empírica que não articula o técnico com o jurídico descreve o problema sem orientar a solução. A interdisciplinaridade desta obra não é uma concessão ao ecletismo: é a única postura metodológica honesta diante de um campo em que as fronteiras entre tecnologia, política, direito e ética são constitutivamente porosas.

A publicação desta obra pela Escola do Legislativo da Paraíba não é circunstancial. É a expressão de uma compreensão institucional de que a formação jurídica e política adequada aos desafios da era digital é condição necessária para que o sistema representativo cumpra suas funções constitucionais. Legisladores que não compreendem o funcionamento dos algoritmos que moldam o ambiente informacional de seus eleitores, advogados que não conhecem o regime de responsabilidade solidária das plataformas digitais, assessores de campanha que ignoram as obrigações de rotulagem de conteúdo gerado por IA: todos esses profissionais estão operando com mapas desatualizados num território que se transformou radicalmente. A formação que esta obra pretende contribuir para construir não é luxo acadêmico. É requisito operacional para quem atua no campo eleitoral em 2026.

Uma advertência honesta ao leitor que chega a esta obra esperando encontrar soluções definitivas: elas não existem, e qualquer livro que as prometesse seria parte do problema que pretende resolver. O campo da IA eleitoral evolui mais rapidamente do que qualquer ciclo de produção acadêmica consegue acompanhar, e as ferramentas que serão usadas nas eleições de 2030 provavelmente ainda não foram desenvolvidas quando estas páginas foram escritas. O que esta obra oferece não é um conjunto de respostas permanentes, mas um conjunto de perguntas precisas e de instrumentos analíticos que permitem formular novas perguntas à medida que o campo evolui. A capacidade de fazer as perguntas certas sobre tecnologias e práticas que ainda não existem é mais valiosa, a longo prazo, do que o conhecimento sobre as que existem hoje.

O leitor encontrará nesta obra uma postura que pode ser descrita como crítica construtiva: o reconhecimento simultâneo do que o Brasil construiu, e construiu muito, mais do que a maioria das democracias equivalentes, e do que ainda precisa ser construído. A Resolução TSE nº 23.732/2024 é um avanço genuíno e pioneiro. O CIEDDE é uma iniciativa institucional relevante. A parceria do TSE com agências de fact-checking produziu resultados documentados. E ainda assim o sistema como um todo não está suficientemente preparado para o ambiente tecnológico de 2026, e muito menos para o de 2030. Nomear essa insuficiência com precisão não é pessimismo. É o primeiro passo para superá-la.

A democracia que o Brasil herdou de gerações que lutaram por ela em condições incomparavelmente mais difíceis merece ser entregue à geração seguinte em condição de fortalecimento, não apenas de sobrevivência. As eleições gerais de 2026 são o próximo teste. O tempo para preparação é finito. E a pergunta que intitula o capítulo de encerramento desta obra, se é possível salvar a democracia, tem resposta afirmativa. Mas essa resposta não está neste livro. Está nas escolhas que serão feitas por quem o leu.

Rodolpho Raphael

CAPÍTULO 1

A Eleição que Você Não Viu Acontecer

Como a Inteligência Artificial opera nos bastidores das disputas eleitorais

“A política sempre foi sobre narrativas. O que mudou é que agora as narrativas são fabricadas por máquinas e o eleitor não sabe disso."

Zeynep Tufekci

O CAMPO DA POLÍTICA COMPUTACIONAL

O campo de estudos que se convencionou denominar *política computacional* (*computational politics*) consolidou-se como área de investigação sistemática a partir de meados da primeira década do século XXI, em um contexto marcado pela convergência entre avanços significativos no aprendizado de máquina, a massificação das redes sociais digitais e a crescente digitalização dos processos políticos e eleitorais.

Ainda que suas raízes possam ser rastreadas em estudos anteriores sobre comunicação política mediada por tecnologia, é nesse período que se observa a formação de um campo propriamente estruturado, com agenda de pesquisa própria, métodos específicos e crescente institucionalização acadêmica. Trata-se de um domínio intrinsecamente interdisciplinar que articula, de maneira complexa e não hierárquica, contribuições das ciências da computação, da ciência política, dos estudos de comunicação, da psicologia social, da economia comportamental e do direito, especialmente no que se refere à regulação de dados e à integridade do processo democrático.

Os trabalhos fundadores desse campo desempenharam papel decisivo na definição de suas premissas empíricas e metodológicas. A análise de Ryan D. Nickerson e Todd Rogers (2010), ao explorar a aplicação de experimentos comportamentais em campanhas eleitorais, demonstrou a viabilidade de intervenções controladas capazes de produzir efeitos mensuráveis sobre o comportamento político. De modo complementar, a pesquisa de Michal Kosinski, David Stillwell e Thore Graepel (2013) evidenciou que traços de personalidade podem ser inferidos com elevado grau de precisão a partir de rastros digitais aparentemente triviais, como curtidas em redes sociais, inaugurando uma nova fronteira na modelagem psicográfica automatizada.

Já a descrição sistemática das técnicas de microtargeting eleitoral realizada por Sasha Issenberg (2012), em sua análise da campanha de reeleição de Barack Obama, forneceu uma síntese abrangente sobre a operacionalização estratégica desses dados em campanhas políticas reais. Em conjunto, esses trabalhos estabeleceram uma premissa central segundo a qual, a partir de volumes suficientes de dados comportamentais, é possível não apenas prever, mas também influenciar o comportamento eleitoral de indivíduos específicos com níveis estatisticamente significativos de acurácia.

No plano da teoria política, a problematização normativa dessas transformações tecnológicas foi desenvolvida de forma particularmente fecunda. Eli Pariser (2011), ao introduzir o conceito de “filtro invisível”, chamou atenção para os efeitos da personalização algorítmica sobre a autonomia informacional dos indivíduos e sobre a formação da opinião pública. Essa linha de análise foi aprofundada por Cass Sunstein (2017), cuja investigação sobre as câmaras de eco digitais demonstrou empiricamente como ambientes informacionais homogêneos tendem a intensificar a polarização ideológica e a fragmentação do espaço público.

Em uma perspectiva estrutural, Shoshana Zuboff (2019) formulou o conceito de “capitalismo de vigilância” para descrever o modelo econômico subjacente à coleta massiva e à mercantilização de dados comportamentais, modelo que constitui a infraestrutura material das práticas contemporâneas de microtargeting político.

No contexto brasileiro, a produção acadêmica tem avançado de forma consistente na análise empírica do ecossistema informacional eleitoral, com destaque para os estudos de Felipe Nunes de Araújo e João Guilherme Oliveira (2020), bem como para as contribuições do grupo NetLab da UFRJ, que têm se dedicado à investigação de dinâmicas de desinformação, coordenação inautêntica e circulação de conteúdo político em plataformas digitais. No plano internacional, os trabalhos de Yochai Benkler, Robert Faris e Hal Roberts (2018) também oferecem aportes relevantes ao demonstrar como a arquitetura do ecossistema midiático influencia a propagação de desinformação e a assimetria informacional entre diferentes grupos políticos.

Apesar do avanço significativo da literatura, persiste uma lacuna relevante na insuficiente integração entre a análise técnico-computacional dos sistemas de Inteligência Artificial, frequentemente desenvolvida em chave descritiva ou instrumental, e a análise jurídico-normativa do marco regulatório eleitoral, que tende a operar com categorias próprias do direito, nem sempre sensíveis às especificidades técnicas das tecnologias emergentes. Essa dissociação produz um duplo risco. De um lado, a regulação pode tornar-se ineficaz por desconhecer os mecanismos concretos de funcionamento das tecnologias que pretende disciplinar. De outro, a análise técnica pode desconsiderar os limites normativos e os valores democráticos que devem orientar sua aplicação.

É precisamente nesse ponto de interseção que esta obra se insere. Ao propor uma articulação sistemática entre as dimensões técnica e jurídica da política computacional, o livro busca não apenas descrever fenômenos, mas oferecer uma compreensão integrada capaz de informar tanto a prática das campanhas quanto a atuação regulatória e o debate público qualificado. Trata-se, em última instância, de contribuir para a construção de um arcabouço analítico que permita compreender, e quando necessário conter, os impactos da Inteligência Artificial sobre a integridade do processo democrático.

Conceito de Inteligência Artificial: Do Acadêmico ao Jurídico

O termo 'inteligência artificial' foi cunhado pelo matemático John McCarthy em 1956, na Conferência de Dartmouth, para descrever 'a ciência e a engenharia de fabricar máquinas inteligentes'. Desde então, a definição evoluiu junto com as capacidades técnicas e, nos últimos anos, também com as necessidades regulatórias. Para fins jurídicos, o mais relevante não é a definição filosófica se as máquinas 'realmente' pensam mas a definição funcional: o que esses sistemas fazem e quais são suas consequências sobre direitos e processos democráticos.

A Union Européenne AI Act (2024) adotou a definição hoje mais influente no mundo: sistema baseado em máquina que, para um conjunto dado de objetivos definidos por humanos, é capaz de fazer previsões, recomendações ou tomar decisões que influenciam ambientes reais ou virtuais. O AI Act adiciona um critério importante: o sistema deve ser capaz de gerar outputs como conteúdo, previsões, recomendações ou decisões com graus variáveis de autonomia.

📌 Definição — IA Generativa (TSE, 2024)

Conjunto de técnicas computacionais capazes de executar tarefas que normalmente requerem inteligência humana, incluindo, mas não se limitando a: processamento de linguagem natural, geração de texto, síntese de imagem, áudio e vídeo, reconhecimento de padrões, tomada de decisão automatizada e personalização de conteúdo, quando aplicadas em contexto de propaganda eleitoral, comunicação de campanha ou interação com eleitores. Nota: A definição é funcional ('capaz de executar tarefas que normalmente requerem inteligência humana') e não exaustiva ('incluindo, mas não se limitando a') — o que garante longevidade normativa mas exige interpretação caso a caso

Brasil. TSE. Resolução nº 23.732/2024. [Adaptação contextual do autor]

A distinção entre IA generativa e algoritmos tradicionais é a mais relevante para o direito eleitoral contemporâneo porque é ela que separa os sistemas que meramente selecionam e distribuem conteúdo existente (algoritmos de recomendação) daqueles que criam conteúdo novo textos, imagens, áudios e vídeos que nunca existiram com aparência crescentemente indistinguível do real.

IA GENERATIVA VERSUS ALGORITMOS TRADICIONAIS: A DISTINÇÃO QUE MAIS IMPORTA PARA O DIREITO ELEITORAL

A distinção entre IA generativa e algoritmos tradicionais é a mais relevante para o direito eleitoral contemporâneo, porque é ela que separa os sistemas que meramente selecionam e distribuem conteúdo existente daqueles que criam conteúdo novo, textos, imagens, áudios e vídeos que nunca existiram, com aparência crescentemente indistinguível do real.

Tabela 01. Algoritmo Tradicional versus IA Generativa no Ambiente Eleitoral

Dimensão	Algoritmo Tradicional	IA Generativa
O que faz	Processa, classifica, ordena e distribui dados ou conteúdo existente. Não cria nada novo.	Cria conteúdo novo: texto, imagem, áudio, vídeo, a partir de padrões aprendidos em dados massivos.
Exemplos eleitorais	Algoritmo do TikTok (decide o que cada usuário vê); sistema de segmentação do Facebook; modelo de scoring de eleitores.	ChatGPT gerando discurso de campanha; Midjourney criando imagem falsa de candidato; ElevenLabs clonando voz; Sora gerando vídeo sintético.
Risco eleitoral principal	Amplificação de conteúdo polarizador; bolhas de filtro; microtargeting psicográfico; supressão de voto por distribuição seletiva.	Criação de deepfakes eleitorais; fabricação de declarações falsas; desinformação personalizada em escala; clonagem de identidade digital.
Tratamento na Res. TSE 23.732/2024	Regulado indiretamente, via obrigações de transparência de impulsionamento e proibição de conteúdo enganoso.	Regulado diretamente: definição inclui geração de texto, síntese de imagem, áudio e vídeo; conteúdo sintético enganoso é vedação absoluta.
Detectabilidade	Difícil detectar individualmente: o dano está no padrão de distribuição, não num conteúdo específico.	Tecnicamente detectável por análise forense, mas custo de produção cai mais rápido que eficácia de detecção.

Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024.

Os Pilares Técnicos: Machine Learning, Deep Learning, NLP e Visão Computacional

Os termos técnicos que dominam o debate sobre IA eleitoral frequentemente intimidam profissionais com formação não técnica, e essa intimidação tem uma consequência política que precisa ser reconhecida: ela concentra a capacidade de participar do debate regulatório sobre tecnologias com profundas implicações democráticas nas mãos de quem tem formação técnica, que são frequentemente os próprios desenvolvedores das tecnologias que precisam ser reguladas. Esta seção apresenta os conceitos fundamentais com definições acessíveis e, crucialmente, com a explicação de por que cada um importa especificamente para o contexto eleitoral. O objetivo não é formar engenheiros de IA: é formar profissionais do direito, da política e da comunicação capazes de compreender, questionar e argumentar sobre sistemas técnicos cujas consequências jurídicas e democráticas são da mais alta relevância.

A distinção entre os quatro campos descritos a seguir não é apenas taxonômica. Ela tem implicações práticas para o advogado eleitoral que precisa enquadrar juridicamente uma conduta, para o regulador que precisa identificar a qual categoria de sistema uma norma se aplica e para o profissional de campanha que precisa avaliar quais ferramentas contratadas se enquadram nas obrigações da Resolução TSE nº 23.732/2024. Um deepfake de vídeo envolve tecnologia de visão computacional. Um chatbot eleitoral não identificado envolve NLP. Um sistema de scoring de probabilidade de voto envolve Machine Learning. E um sistema que produz, distribui e adapta conteúdo eleitoral de forma autônoma envolve a combinação de todos os quatro campos. Saber qual campo está em jogo é o primeiro passo para saber qual norma se aplica e qual argumento jurídico é mais sólido.

Machine Learning: a base do targeting eleitoral. Machine Learning é o ramo da inteligência artificial no qual sistemas aprendem padrões a partir de dados, sem serem explicitamente programados para cada tarefa. A diferença em relação à programação tradicional é estrutural: em vez de um programador escrever regras fixas que o computador executa, o sistema recebe exemplos e infere suas próprias regras a partir dos padrões que

identifica nesses exemplos. A qualidade do aprendizado depende diretamente da quantidade e da qualidade dos dados de treinamento, o que cria a dependência de dados em escala que é o fundamento do modelo de negócios das grandes plataformas digitais e, por extensão, do capitalismo de vigilância descrito por Zuboff (2019).

No contexto eleitoral, modelos de Machine Learning são a tecnologia subjacente a três aplicações com implicações diretas para a integridade democrática. A primeira são os sistemas de scoring eleitoral, que atribuem a cada eleitor um escore de probabilidade de voto em determinado candidato ou partido, de probabilidade de mudança de intenção de voto e de receptividade a mensagens específicas. Esses sistemas são a base do microtargeting moderno: sem o score, a personalização de mensagens não tem parâmetro de segmentação.

A segunda são os sistemas de detecção de bots em redes sociais, que identificam padrões de comportamento inautêntico por análise estatística de frequência, timing e conteúdo de publicações: são os mesmos algoritmos que, quando usados de forma ofensiva, geram comportamento de bot suficientemente plausível para escapar à detecção. E a terceira são os sistemas de recomendação de conteúdo que determinam o que cada eleitor vê em seu feed, criando o ambiente informacional personalizado que Pariser (2011) denominou bolha de filtro e que é o mecanismo pelo qual plataformas amplificam conteúdo emocionalmente ativador independentemente de sua autenticidade.

Deep Learning: a tecnologia dos deepfakes. Deep Learning é um subconjunto de Machine Learning que utiliza redes neurais artificiais com múltiplas camadas para aprender representações cada vez mais abstratas dos dados de treinamento. A diferença em relação ao Machine Learning convencional é a capacidade de aprender automaticamente quais características dos dados são relevantes para a tarefa, sem que humanos precisem definir essas características manualmente. Onde um sistema de Machine Learning convencional precisaria de um especialista que identificasse quais aspectos de um rosto são relevantes para o reconhecimento facial, uma rede de Deep Learning aprende sozinha, a partir de milhões de exemplos, quais padrões distinguem um rosto de outro.

É a base de todas as aplicações de IA generativa modernas, do ChatGPT aos sistemas de síntese de vídeo, e é a tecnologia subjacente a todos os deepfakes eleitorais, a todos os sistemas de clonagem de voz e a todos os modelos de linguagem que podem gerar conteúdo falso de campanha em escala. A arquitetura específica mais relevante para

deepfakes eleitorais é a das redes adversariais generativas, ou GANs, descritas por Goodfellow et al. (2016): dois sistemas de rede neural em competição, um gerador que produz conteúdo sintético e um discriminador que tenta identificar o sintético do autêntico, que melhoram mutuamente até que o gerador produza conteúdo indistinguível ao discriminador. Essa arquitetura é o mecanismo da corrida armamentista entre geração e detecção analisada no Capítulo 14: cada avanço no discriminador é incorporado ao treinamento do gerador na próxima iteração.

Processamento de Linguagem Natural: das fake news aos chatbots. O Processamento de Linguagem Natural é o campo da IA dedicado à compreensão e à geração de linguagem humana por computadores. Inclui análise de sentimento, tradução automática, geração de texto, reconhecimento de fala, síntese de voz e, em suas aplicações mais avançadas, raciocínio sobre conteúdo textual complexo. Os grandes modelos de linguagem, os LLMs, como GPT-4, Claude e Gemini, são a manifestação mais avançada de NLP disponível em 2026 e a tecnologia que mais radicalmente transformou o ecossistema de desinformação eleitoral a partir de 2022.

NLP alimenta as principais aplicações de desinformação textual automatizada que o direito eleitoral precisa regular: chatbots eleitorais que simulam interação humana sem identificação, violando o artigo 15 da Resolução TSE nº 23.732/2024; geradores de fake news em escala que produzem textos individualizados para cada perfil de eleitor; sistemas de análise de sentimento que monitoram em tempo quase real a resposta emocional do eleitorado a eventos de campanha; e ferramentas de microtargeting que personalizam mensagens com linguagem calibrada para o perfil psicográfico específico de cada receptor. Para o advogado eleitoral, a distinção entre um chatbot baseado em NLP que simula interação humana e um chatbot que se identifica adequadamente como sistema automatizado é a distinção entre conduta vedada pelo artigo 9º, inciso II, da Resolução e conduta regulada pelo artigo 15: a mesma tecnologia, usos com enquadramentos jurídicos radicalmente distintos.

Visão Computacional: da detecção ao deepfake. Visão Computacional é o campo da IA dedicado ao processamento e à análise de imagens e vídeos por sistemas computacionais. Inclui reconhecimento de faces, rastreamento de movimento, análise de cenas, geração de imagens sintéticas por GANs e modelos de difusão, e síntese de vídeo. É a mesma tecnologia que permite ao smartphone desbloquear com o rosto do usuário e

que permite criar um vídeo convincente de um candidato fazendo uma declaração que nunca fez: a diferença está na direção do problema, reconhecimento versus geração, e nos dados de treinamento utilizados.

Visão Computacional é a tecnologia central dos deepfakes eleitorais de vídeo e de imagem analisados no Capítulo 7. É também a tecnologia utilizada nos sistemas de detecção de deepfakes descritos no Capítulo 14: modelos treinados para identificar os artefatos estatísticos que sistemas de síntese deixam nas imagens e vídeos gerados, na distribuição de frequências, nos padrões de textura e nas inconsistências de iluminação que escapam à percepção humana mas são detectáveis por análise computacional. A ironia fundamental desse campo, que os sistemas de detecção e os sistemas de geração utilizam a mesma tecnologia e frequentemente os mesmos dados de treinamento, é o mecanismo que torna a corrida armamentista estruturalmente assimétrica em favor dos sistemas de geração: cada melhoria nos sistemas de detecção é informação disponível para os sistemas de geração na iteração seguinte.

A Evolução da Comunicação Política

A evolução da comunicação política desde o início do século XX até o presente segue uma trajetória que, analisada em retrospecto, revela um padrão recorrente de suficiente regularidade para ser tratado como lei empírica: cada nova tecnologia de comunicação de massa amplia o alcance das campanhas eleitorais, redistribui o poder entre atores com acesso diferenciado à tecnologia e cria novos vetores de manipulação que os marcos regulatórios construídos para a tecnologia anterior não haviam previsto. Essa regularidade não é acidental: ela reflete a estrutura do processo político democrático, em que campanhas têm incentivos permanentes para adotar qualquer vantagem comunicacional disponível e em que reguladores chegam sistematicamente depois, porque só podem regular o que já observaram.

O rádio, que se tornou meio de comunicação política de massa nas décadas de 1920 e 1930, criou pela primeira vez a possibilidade de que um candidato alcançasse milhões de eleitores simultaneamente com a mesma mensagem personalizada pela voz e pela narrativa. Também criou, pela primeira vez, o problema da propaganda política de massa com capacidade de exploração emocional que nenhuma regulação de imprensa havia

previsto: as experiências de manipulação radiofônica na Europa do período entreguerras demonstraram que a tecnologia que democratizava o acesso à informação também democratizava o acesso à desinformação em escala.

A televisão, que se tornou o meio dominante de comunicação política nas décadas de 1950 e 1960, adicionou a imagem ao alcance de massa do rádio, criando o fenômeno da personalização política pela aparência e pelo desempenho televisivo que transformou irreversivelmente a natureza da competição eleitoral. O debate televisivo Kennedy-Nixon de 1960 é frequentemente citado como o momento inaugural dessa transformação: pela primeira vez na história, candidatos com posições políticas similares podiam ter desempenhos radicalmente distintos perante o eleitorado com base em características de apresentação que nada tinham a ver com a qualidade de suas propostas. A regulação do horário eleitoral gratuito, que o sistema brasileiro desenvolveu para democratizar o acesso a esse meio, foi a resposta institucional mais significativa a esse desafio, mas chegou décadas depois dos primeiros problemas.

A internet e as redes sociais, que se tornaram meios políticos relevantes na virada do século e dominantes na década de 2010, adicionaram a fragmentação e a personalização ao alcance de massa das mídias anteriores, criando o ecossistema informacional descrito por Pariser (2011) em que cada eleitor habita uma versão ligeiramente diferente da realidade política. Cambridge Analytica, em 2016, tornou globalmente visível o problema do microtargeting político baseado em dados comportamentais que o marco regulatório eleitoral não havia antecipado: a regulação disponível havia sido construída para meios de comunicação de massa identificáveis e uniformes, e não para sistemas que entregam mensagens diferentes a públicos diferentes sem que nenhuma autoridade externa tenha acesso ao conjunto da comunicação.

A IA generativa, que se tornou tecnologia de acesso popular a partir de 2022, adicionou a síntese e a autonomia às capacidades das tecnologias anteriores, criando a descontinuidade qualitativa que esta obra procurou documentar e analisar. O que distingue a IA generativa de todas as tecnologias de comunicação política anteriores não é apenas a escala, o alcance ou a personalização: é a capacidade de fabricar realidade com fidelidade suficiente para enganar eleitores sem treinamento específico de detecção, e de fazê-lo de forma autônoma, em escala industrial e a custo marginal próximo de zero. Nenhuma das tecnologias anteriores havia combinado essas três características simultaneamente, e é

essa combinação que torna a resposta regulatória, jurídica e educacional que esta obra analisa estruturalmente diferente de todas as respostas às tecnologias anteriores.

A regularidade histórica que esse padrão revela tem uma implicação normativa que precisa ser enunciada com clareza: não existe tecnologia de comunicação política neutra. Cada tecnologia que amplia as capacidades de comunicação de campanhas eleitorais também amplia as capacidades de manipulação, e o equilíbrio entre as duas dimensões depende das escolhas de design, de governança e de regulação que são feitas quando a tecnologia ainda está em fase de adoção inicial. As escolhas sobre como governar a IA eleitoral que estão sendo feitas em 2026 seguirão o padrão histórico de todas as tecnologias anteriores: elas definirão o equilíbrio entre capacidade comunicacional legítima e capacidade de manipulação para as próximas décadas, e as gerações que herdarem esse equilíbrio terão muito menos poder de alterá-lo do que a geração que o está estabelecendo agora.

A linha do tempo a seguir materializa o argumento desenvolvido na seção anterior em marcos concretos e datados, permitindo visualizar com precisão o que a narrativa histórica tende a abstrair: a aceleração do ciclo de inovação tecnológica e seus efeitos sobre a comunicação política.

A leitura dos intervalos entre marcos é tão reveladora quanto a leitura dos marcos em si. Do UNIVAC de 1952 ao microtargeting moderno de 2004 passaram-se cinquenta anos. Do microtargeting de 2004 ao Cambridge Analytica de 2016 passaram-se doze anos. Do Cambridge Analytica ao ChatGPT passaram-se seis anos. Do ChatGPT ao primeiro ciclo eleitoral global com IA generativa passaram-se dois anos.

Essa compressão temporal não é circunstancial: é a manifestação empírica da curva exponencial que Brynjolfsson e McAfee identificaram como o mecanismo estrutural das tecnologias digitais. A implicação mais direta para os reguladores é a que esta obra desenvolveu ao longo de seus capítulos: se o ciclo de inovação tecnológica continuou acelerando nessa proporção, o próximo marco de ruptura equivalente ao ChatGPT chegará antes que qualquer resposta regulatória ao ciclo atual esteja completamente implementada. A preparação para esse próximo marco começa com a compreensão precisa de onde estamos no ciclo atual, que é o que a linha do tempo entrega.

Tabela 02. Da Análise Computacional à IA Generativa: Marcos da Tecnologia na Comunicação Política (1952–2024)

1952	**Primeiros computadores em análise eleitoral**	A CBS utilizou o computador UNIVAC I para prever o resultado da eleição presidencial americana. A previsão correta de Eisenhower surpreendeu os apresentadores ao vivo — foi a primeira demonstração pública do poder computacional na análise de dados eleitorais.
1960s–70s	**Databanks e segmentação eleitoral**	Primeiros bancos de dados de eleitores criados por partidos americanos para segmentação de envio de correspondência. Início do que viria a ser o 'voter file' — a base de dados de eleitores que é o coração de toda campanha moderna.
1994–1996	**Internet e primeiras campanhas online**	As eleições presidenciais americanas de 1996 (Clinton vs. Dole) foram as primeiras a usar websites de campanha sistematicamente. Início da comunicação política direta sem intermediários jornalísticos.
2004	**Micro-targeting moderno — Bush vs. Kerry**	A campanha de reeleição de George W. Bush utilizou, pela primeira vez em escala nacional, modelos preditivos sofisticados para identificar eleitores indecisos persuasíveis. Base do que se tornaria o microtargeting moderno.
2008	**Obama e a primeira campanha 2.0**	Barack Obama usa redes sociais (Facebook, YouTube) para mobilização em massa de voluntários e arrecadação de pequenas doações. Primeiro uso sistemático de dados de comportamento digital para segmentação eleitoral.
2012	**Obama — Project Narwhal e Big Data**	Campanha de reeleição de Obama integra dados de 180 milhões de americanos em uma plataforma unificada. 66.000 simulações computacionais por noite para alocar recursos. Marco do data-driven campaigning.
2016	**Cambridge Analytica, Trump, Brexit**	Dados de 87 milhões de usuários do Facebook coletados sem consentimento para microtargeting psicográfico. Primeiro uso documentado de perfis OCEAN para manipulação eleitoral em escala. Marco da crise de dados eleitorais.
2018	**Brasil — WhatsApp e disparo em massa**	Eleições brasileiras: primeiro caso documentado de uso do WhatsApp para disparo em massa de desinformação eleitoral em escala industrial. Introdução de novo vetor de manipulação: plataformas de mensageria criptografadas.
2022 2023	**IA generativa — ChatGPT muda o jogo**	Lançamento do ChatGPT (novembro 2022) democratiza o acesso a modelos de linguagem de alta qualidade. Início da era da desinformação industrial automatizada: qualquer pessoa pode gerar conteúdo falso convincente em escala.
2024	**Primeiro ciclo eleitoral global com IA generativa**	Mais de 60 países realizaram eleições em 2024 — o ano eleitoral mais intenso da história — em contexto de IA generativa amplamente disponível. Primeiros deepfakes eleitorais de alto impacto documentados (EUA, Eslováquia, Bangladesh, Argentina).

Fontes: Oxford Internet Institute; MIT Election Lab; TSE; relatórios eleitorais do IFES.

A Eleição Paralela: O que Acontece Enquanto Você Assiste ao Debate

Era noite de debate presidencial. Nas telas de televisão, dois candidatos trocavam argumentos sobre educação, segurança e economia, num ritual democrático conhecido, previsível e, em certa medida, confortável na sua familiaridade. Nas telas de smartphones, contudo, uma outra disputa estava em curso: invisível, silenciosa, distribuída por milhões de dispositivos simultaneamente e, em muitos aspectos, mais determinante para o resultado eleitoral do que aquela que se desenrolava sob os holofotes da transmissão oficial.

Enquanto o debate transcorria, sistemas automatizados de análise de sentimento monitoravam em tempo quase real as reações do eleitorado a cada declaração dos candidatos, processando milhões de interações digitais para identificar, com precisão estatística, quais temas produziam maior ressonância emocional em quais segmentos demográficos.

Esses dados não eram apenas observados: eram consumidos em tempo real por sistemas de otimização que recalibravam, minuto a minuto, quais mensagens seriam entregues a quais eleitores nas plataformas digitais, priorizando indivíduos classificados como persuadíveis em zonas geográficas eleitoralmente estratégicas. Em paralelo, redes coordenadas de contas automatizadas amplificavam artificialmente determinadas narrativas, elevando hashtags específicas às listas de tendências e produzindo a aparência de mobilização orgânica onde havia operação algorítmica.

Em grupos privados de WhatsApp, conteúdos cuidadosamente construídos, combinando elementos factuais e desinformação calibrada para cada perfil de receptor, circulavam com velocidade, capilaridade e personalização impossíveis em qualquer ambiente comunicacional pré-digital.

Essa dinâmica não é exercício de imaginação nem projeção futurista. É síntese empiricamente fundamentada de práticas documentadas em sucessivos ciclos eleitorais, no Brasil e em outras democracias contemporâneas. As eleições presidenciais brasileiras de 2018 e 2022, analisadas em detalhe nos Capítulos 9 e 10 desta obra, evidenciaram a consolidação progressiva de uma arquitetura informacional na qual a disputa política se desloca para camadas técnicas que escapam à percepção direta do eleitor comum e, com frequência, à capacidade de fiscalização dos sistemas regulatórios construídos para o ambiente eleitoral anterior.

A eleição que o cidadão vê, composta por debates televisivos, comícios, entrevistas e propaganda eleitoral oficial, corresponde apenas à superfície visível de um processo decisório muito mais complexo. Em níveis subjacentes, operam sistemas de classificação, predição e otimização que estruturam a circulação de informação política, definem

prioridades de exposição e modulam continuamente as condições sob as quais as preferências eleitorais são formadas. O resultado é um ambiente informacional assimétrico no qual diferentes eleitores são expostos a versões distintas e, por vezes, incompatíveis da realidade política, sem que nenhum deles tenha consciência da extensão dessa assimetria nem do mecanismo que a produz.

A dimensão mais perturbadora dessa transformação não é a sofisticação tecnológica dos sistemas envolvidos: é sua opacidade estrutural. Como observa Tufekci (2017), a força desses sistemas reside não apenas em sua capacidade de amplificação, mas na sua fragilidade institucional e na sua baixa transparência operacional. Diferentemente de um anúncio veiculado na televisão, cujo conteúdo é público e passível de escrutínio coletivo, a publicidade política mediada por algoritmos é, por definição, individualizada e invisível ao conjunto do eleitorado.

Um eleitor pode ser exposto reiteradamente a mensagens sobre segurança pública; outro, com perfil demográfico distinto, recebe conteúdos voltados a emprego e políticas sociais; um terceiro, identificado como potencial opositor, pode ser deliberadamente excluído da comunicação oficial da campanha ou direcionado a conteúdos projetados para reduzir sua propensão ao comparecimento eleitoral. Nenhum desses eleitores tem acesso ao universo informacional ao qual os demais estão submetidos.

Essa fragmentação radical da experiência informacional compromete um dos pressupostos normativos mais fundamentais da democracia deliberativa: a existência de um espaço público minimamente compartilhado, no qual argumentos possam ser confrontados sob condições de relativa simetria epistêmica. Quando cada eleitor habita um ambiente informacional personalizado, dinâmico e invisível aos demais, a própria ideia de debate público sofre uma transformação estrutural que vai além de uma mudança nos meios de comunicação: é uma reconfiguração das condições de possibilidade da deliberação democrática que Habermas (1997) identificou como pressuposto de legitimidade das decisões coletivas. É essa reconfiguração que este livro analisa, e é para endereçá-la que as propostas regulatórias, jurídicas e educacionais dos capítulos seguintes foram construídas.

Estamos, portanto, diante de um ponto de inflexão histórico. Pela primeira vez, as ferramentas mais eficazes de persuasão política não se encontram predominantemente nos discursos públicos, nos meios de comunicação tradicionais ou na mobilização presencial: estão em sistemas algorítmicos que operam de forma distribuída, adaptativa e opaca, sobre os quais o eleitor médio tem compreensão mínima e sobre os quais o sistema regulatório eleitoral tem, como os capítulos seguintes demonstrarão, alcance ainda insuficiente.

O Data Stack Eleitoral Moderno

O coração operacional da campanha digital moderna é o que os profissionais do setor denominam data stack eleitoral: um conjunto integrado de sistemas de coleta, armazenamento, processamento e análise de dados que alimenta cada decisão estratégica da campanha, desde a alocação de verbas publicitárias até a definição de quais eleitores merecem visita pessoal de um voluntário, desde a calibração de mensagens por perfil demográfico até a identificação de quais municípios têm maior potencial de reversão de voto.

O data stack eleitoral moderno não é uma inovação surgida com a IA generativa: é a infraestrutura de dados que campanhas sofisticadas foram construindo progressivamente desde o microtargeting de 2004 descrito na linha do tempo do apêndice anterior. O que a IA generativa transformou não foi a existência dessa infraestrutura, mas sua escala, sua velocidade de processamento e, mais significativamente, sua capacidade de agir sobre os dados em vez de apenas analisá-los: onde o data stack de 2012 produzia recomendações para decisões humanas, o data stack de 2026 pode executar as decisões autonomamente, sem intervenção humana no ato específico de cada intervenção comunicacional.

Para o advogado eleitoral e para o regulador, o data stack eleitoral é relevante por três razões que percorrem toda esta obra. A primeira é que ele é o mecanismo pelo qual dados pessoais de eleitores, coletados com base legal variável, são transformados em instrumento de persuasão política, com as implicações para a LGPD e para a integridade eleitoral analisadas no Capítulo 13. A segunda é que ele é a infraestrutura que torna o microtargeting de precisão psicológica possível em escala, com as implicações para a autonomia política dos eleitores analisadas nos Capítulos 4 e 17. E a terceira é que ele é o objeto sobre o qual a maioria das obrigações de transparência e auditoria do PL 2.338/2023 e da Resolução TSE nº 23.732/2024 incidem, tornando sua compreensão técnica condição necessária para a aplicação adequada dessas normas.

Tabela 03. Componentes do Data Stack Eleitoral Moderno

Componente	Função	Tecnologia Central	Relevância Jurídica
CRM Eleitoral	Mapeamento individualizado de eleitores: histórico de contato, predisposição ao voto, atividade de voluntário.	Banco de dados relacional, integração de fontes múltiplas	Tratamento de dados pessoais sujeito à LGPD

Data Lake Eleitoral	Repositório centralizado de dados heterogêneos: registros eleitorais, redes sociais, dados de consumo. Pode conter centenas de variáveis por eleitor.	Cloud storage, ETL pipelines, Big Data	Dados sensíveis (opinião política) exigem consentimento específico (LGPD art. 11)
Publicidade Programática	Compra automatizada de espaço publicitário em tempo real (RTB). Segmentação ultra-refinada por localização, comportamento e perfil.	Real-Time Bidding, algoritmos de leilão	Registro obrigatório no TSE; impulsionamento irregular é infração eleitoral
A/B Testing Político	Comparação simultânea de versões distintas de mensagens para identificar qual produz maior engajamento ou mudança de intenção de voto.	Experimentos controlados, análise estatística	Conteúdo gerado com auxílio de IA exige rotulagem (Res. TSE 23.732/2024)
Ground Game Digital	Coordenação tecnológica de voluntários para ações presenciais orientadas por dados preditivos sobre quais domicílios visitar.	Plataformas de field organizing, geolocalização	Coleta de dados presencial sujeita à LGPD
GOTV	Mobilização para comparecimento eleitoral. Foca em eleitores simpáticos com baixo score de comparecimento.	Modelos preditivos de turnout, sistemas de contato	Uso de dados pessoais para mobilização requer base legal adequada

Fonte: Elaborado pelo autor com base em Issenberg (2012), Matejic (2018)

Anatomia de uma Campanha de IA: As Cinco Camadas Invisíveis

Para compreender como a inteligência artificial opera concretamente em uma campanha eleitoral contemporânea, é útil desagregar o processo em suas camadas constitutivas, porque a análise da campanha como unidade indivisível obscurece precisamente o que precisa ser visível para fins de regulação, responsabilização e conformidade: a especificidade técnica, jurídica e ética de cada etapa do ciclo de vida da IA eleitoral.

Identificamos cinco camadas funcionais que, juntas, constituem a arquitetura operacional de uma campanha digital avançada. Cada camada tem tecnologias específicas, atores com responsabilidades distintas e riscos jurídicos que precisam ser analisados em

sua especificidade antes de serem integrados na análise sistêmica. A separação analítica das cinco camadas não é apenas exercício descritivo: é instrumento de diagnóstico que permite ao advogado eleitoral identificar em qual ponto da cadeia uma violação ocorreu, ao regulador determinar qual norma se aplica a qual conduta e ao profissional de campanha compreender onde estão as exposições de risco que uma análise superficial deixa invisíveis.

Há uma dimensão adicional que a análise em camadas revela e que a análise sistêmica tende a obscurecer: a dissociação deliberada entre as camadas é frequentemente uma escolha estratégica de atores que buscam criar distância entre a instrução original e a conduta irregular que dela resulta. Quando o financiador de uma operação de desinformação contrata um intermediário que contrata uma agência que utiliza ferramentas de IA que produzem conteúdo que é distribuído por contas automatizadas, cada elo da cadeia pode argumentar que não sabia o que os demais estavam fazendo. A análise em camadas permite identificar essa dissociação deliberada pela estrutura, não apenas pelo resultado, o que é o passo que transforma a negação plausível de cada elo em responsabilização da cadeia como um todo.

A primeira camada é a coleta e estruturação de dados, que constitui o alicerce sobre o qual todas as outras camadas operam. Inclui a agregação de dados de registros eleitorais públicos do TSE, redes sociais, históricos de compra, dados de geolocalização e outras fontes comerciais; a limpeza e estruturação desses dados em formatos compatíveis com os modelos analíticos das camadas seguintes; e a construção do perfil individual de cada eleitor com todas as variáveis disponíveis. É nessa camada que as questões de conformidade com a LGPD são mais urgentes e mais frequentemente negligenciadas: as bases legais para o tratamento de cada categoria de dado precisam ser identificadas antes da coleta, não depois, e a distinção entre dados pessoais comuns, dados comportamentais e dados sensíveis inferidos discutida no Capítulo 13 tem suas consequências mais práticas precisamente aqui.

A segunda camada é a modelagem e segmentação, que transforma dados brutos em inteligência acionável. Inclui a construção de modelos preditivos de comportamento eleitoral, como probabilidade de voto, probabilidade de persuasão e identificação de vulnerabilidades cognitivas específicas; a segmentação do eleitorado em grupos com características e necessidades comunicacionais distintas; e o desenvolvimento de perfis

psicográficos que permitem a personalização de mensagens. É nessa camada que o problema do microtargeting de precisão psicológica analisado no Capítulo 17 tem sua origem técnica: a sofisticação dos modelos preditivos disponíveis em 2026 permite inferências sobre estado emocional e vulnerabilidade cognitiva que a regulação de 2024 não foi projetada para endereçar.

A terceira camada é a produção de conteúdo, que utiliza as ferramentas de IA generativa analisadas nos capítulos técnicos desta obra para criar material de campanha em escala e com personalização impossível manualmente. Inclui a geração de variantes de texto para diferentes perfis de eleitor, a produção de imagens e vídeos com ou sem síntese facial e vocal, a criação de áudios com síntese de voz e a personalização de conteúdo em tempo quase real com base no feedback das camadas de monitoramento. É nessa camada que as obrigações de rotulagem da Resolução TSE nº 23.732/2024 têm sua aplicação mais direta: todo conteúdo gerado ou significativamente modificado por IA nessa camada está sujeito à identificação obrigatória, e é nessa camada que a maioria das violações documentadas nas eleições de 2024 teve origem.

A quarta camada é a distribuição e amplificação, que utiliza as plataformas digitais como canais de entrega do conteúdo produzido na camada anterior. Inclui a compra de publicidade segmentada nas plataformas, o impulsionamento orgânico de conteúdo por redes de apoiadores, a distribuição por aplicativos de mensageria e, em operações mais sofisticadas, a coordenação de redes de contas que amplificam determinadas narrativas para criar a aparência de mobilização espontânea. É nessa camada que a responsabilidade das plataformas digitais analisada nos Capítulos 8 e 12 tem seu ponto de incidência mais concreto: as plataformas não são apenas canal passivo de distribuição, mas atores ativos que decidem quais conteúdos amplificar e para quais públicos, com os incentivos estruturais para amplificar conteúdo emocionalmente ativador documentados extensivamente na literatura sobre o modelo de negócios das plataformas.

A quinta camada é o monitoramento e otimização, que fecha o ciclo ao medir os resultados das intervenções das camadas anteriores e realimentar o processo com informação para ajuste contínuo. Inclui o monitoramento em tempo real da resposta do eleitorado a diferentes mensagens, o rastreamento de métricas de engajamento por segmento demográfico, a identificação de narrativas adversárias em ascensão que requerem resposta e a otimização contínua de alocação de recursos entre diferentes

estratégias e públicos. É nessa camada que os sistemas de social listening eleitoral descritos no Capítulo 14 têm sua aplicação mais direta, e é aqui que a distinção entre monitoramento legítimo do ambiente informacional e coleta de dados pessoais sem base legal adequada é mais frequentemente cruzada sem que os operadores percebam que a cruzaram.

A análise das cinco camadas revela um padrão que percorre toda esta obra: em cada camada, a fronteira entre uso lícito e uso irregular da IA eleitoral não é uma linha clara e estável, mas uma zona de indeterminação que a tecnologia progressivamente desloca na direção de maior capacidade de manipulação e que a regulação progressivamente busca recuar na direção de maior proteção da autonomia do eleitor. Essa tensão estrutural não será resolvida por nenhuma resolução, lei ou ferramenta tecnológica específica. Ela é o campo em que o direito eleitoral e a democracia precisarão operar indefinidamente, e compreendê-la com precisão é a condição para operá-la com eficácia.

Tabela 04. As Cinco Camadas da IA em Campanhas Eleitorais

Camada	Função	Tecnologia Central	Risco Jurídico Principal
1. Coleta	Obter dados sobre eleitores	APIs, web scraping, data brokers	LGPD art. 7º e 11º — consentimento
2. Modelagem	Construir perfis psicográficos	Machine Learning, OCEAN model	LGPD art. 11º — dados sensíveis
3. Segmentação	Identificar alvos prioritários	Algoritmos preditivos, clustering	Lei das Eleições — abuso de poder
4. Produção	Criar conteúdo personalizado	IA Generativa, LLMs, síntese de mídia	Res. TSE 23.732/2024 — rotulagem
5. Distribuição	Entregar conteúdo ao eleitor	Ad tech, bots, disparos em massa	Lei das Eleições — propaganda irregular

Fonte: Elaborado pelo autor com base em Issenberg (2012), Zuboff (2019) e Brasil, TSE (2024)

LGPD — Art. 5º, II: Dado Sensível

Para os fins desta Lei, considera-se dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural.

Brasil. Lei nº 13.709/2018 (LGPD), art. 5º, II.

A qualificação das opiniões políticas como dados sensíveis tem implicação direta e grave para o microtargeting eleitoral: qualquer sistema que utilize dados sobre preferências políticas de eleitores — inclusive dados inferidos a partir de comportamento digital, como curtidas em publicações de determinado candidato — está tratando dados sensíveis, o que exige, na ausência de outra hipótese legal aplicável, o consentimento específico e destacado do titular (art. 11, I, LGPD).

O CASO EMBLEMATICO: ELEIÇÕES BRASILEIRAS 2018 — A ANATOMIA DE UMA CAMPANHA ALGORÍTMICA

As eleições presidenciais brasileiras de 2018 configuram o caso de estudo mais relevante e amplamente documentado do uso intensivo de tecnologias digitais em campanhas eleitorais no Brasil, incluindo a incorporação, ainda que em estágio inicial, de componentes associados à Inteligência Artificial.

A análise desse episódio, com base em investigações conduzidas pelo Tribunal Superior Eleitoral, pelo Ministério Público Eleitoral, pela Polícia Federal e por pesquisadores independentes, permite observar, em termos concretos, a operacionalização das diferentes camadas do modelo analítico anteriormente proposto. Mais do que um evento isolado, o pleito de 2018 representa um ponto de inflexão na forma como campanhas políticas se estruturam em ambientes digitais altamente conectados.

WhatsApp como infraestrutura de campanha

O WhatsApp, plataforma de mensageria pertencente ao grupo Meta, com mais de 120 milhões de usuários ativos no Brasil à época, consolidou-se como a principal infraestrutura de circulação de conteúdo político informal durante o processo eleitoral. Sua centralidade não decorre apenas de seu alcance quantitativo, mas de um conjunto específico de características técnicas e operacionais que o tornam particularmente adequado à disseminação descentralizada de mensagens. Entre essas características, destacam-se a comunicação em ambiente privado, a criptografia de ponta a ponta, a ausência de mecanismos robustos de verificação de autenticidade de conteúdo e as limitações estruturais à rastreabilidade da origem das mensagens.

A adoção do aplicativo como canal prioritário de campanha revela uma estratégia orientada não apenas pela maximização de alcance, mas pela exploração de zonas de menor visibilidade institucional. Diferentemente das plataformas abertas, nas quais o conteúdo circula em ambientes passíveis de monitoramento público e jornalístico, os aplicativos de mensageria operam em redes fechadas, estruturadas em torno de relações de confiança interpessoal. Essa configuração confere maior credibilidade às mensagens recebidas, ao mesmo tempo em que reduz significativamente a capacidade de fiscalização por parte das autoridades eleitorais e da sociedade civil.

Investigações posteriores identificaram a existência de esquemas organizados de disparo em massa, frequentemente qualificados pelo Tribunal Superior Eleitoral como formas de impulsionamento irregular. Esses esquemas envolviam a contratação de empresas de marketing digital especializadas na automação do envio de mensagens para grandes bases de números de telefone, previamente segmentadas com base em critérios geográficos, comportamentais ou inferências sobre preferências políticas. A infraestrutura utilizada incluía sistemas capazes de simular interações humanas, contornar limitações técnicas da plataforma e operar múltiplas contas de forma coordenada.

Do ponto de vista operacional, esses disparos permitiam alcançar centenas de milhares, e em alguns casos milhões, de eleitores em intervalos extremamente reduzidos de tempo. A escala e a velocidade da disseminação eram incompatíveis com qualquer forma de envio manual, evidenciando o papel central da automação nesse processo. Além disso, a lógica de encaminhamento em cadeia, característica do aplicativo, potencializava exponencialmente o alcance inicial das mensagens, à medida que conteúdos recebidos eram retransmitidos por usuários a seus próprios contatos e grupos.

Sob a perspectiva jurídica, tais práticas suscitam questionamentos relevantes quanto à sua compatibilidade com o ordenamento eleitoral vigente. A utilização de bases de dados obtidas sem consentimento, a ausência de transparência quanto ao financiamento das operações e a dificuldade de identificação dos responsáveis diretos pela disseminação configuram elementos que tensionam princípios fundamentais como a lisura do pleito, a igualdade de oportunidades entre candidatos e a transparência das campanhas. A qualificação dessas condutas como abuso de poder econômico ou uso indevido dos meios de comunicação social tem sido objeto de debates intensos na jurisprudência eleitoral.

A análise das práticas de disparo em massa no WhatsApp não se esgota, contudo, na descrição de sua arquitetura técnica e de sua lógica operacional. A dimensão jurídica

dessas condutas foi objeto de apreciação direta pela Justiça Eleitoral, que passou a enfrentar, em casos concretos, os desafios decorrentes da utilização de infraestruturas digitais para a disseminação automatizada de conteúdo político. Nesse contexto, a atuação do Tribunal Superior Eleitoral revela-se particularmente relevante, na medida em que inaugura a construção jurisprudencial sobre os limites e as consequências do uso dessas tecnologias no processo eleitoral brasileiro.

Um dos marcos mais significativos dessa evolução interpretativa encontra-se no julgamento da Ação de Investigação Judicial Eleitoral referente ao pleito de 2018, no qual se discutiu precisamente a utilização de estruturas empresariais para viabilizar disparos em massa de mensagens com conteúdo eleitoral, sem a devida transparência quanto ao financiamento e à contabilização dos gastos de campanha. A decisão proferida nesse caso não apenas reconheceu a irregularidade das práticas adotadas, mas também estabeleceu parâmetros relevantes para a responsabilização dos agentes envolvidos e para a conformação do debate regulatório subsequente.

É nesse sentido que se insere o quadro jurisprudencial a seguir, que sintetiza os principais elementos da decisão e seus desdobramentos no ordenamento eleitoral brasileiro:

⚖ JURISPRUDÊNCIA TSE — AIJE nº 0601771-86.2018.6.00.0000

EMENTA:

Ação de Investigação Judicial Eleitoral. Disparo em massa pelo WhatsApp. Uso de empresas de tecnologia para envio automatizado de conteúdo eleitoral a listas de eleitores segmentadas. Financiamento por empresários não declarado como despesa de campanha.

RESULTADO:

O TSE reconheceu a irregularidade das práticas de disparo em massa não declaradas como despesa de campanha e a responsabilidade solidária entre os contratantes e as empresas executoras. A investigação resultou em multas e impactou o debate regulatório que culminou na Resolução nº 23.732/2024.

Desinformação em escala industrial

Paralelamente às práticas de disparo em massa, as eleições de 2018 foram marcadas pela circulação de conteúdo desinformativo em escala sem precedentes na história eleitoral brasileira. Evidências empíricas produzidas pelo Grupo de Pesquisa em Políticas Públicas de Acesso à Informação da USP indicam que, entre os cem conteúdos mais compartilhados no WhatsApp durante o período eleitoral, uma proporção significativa

continha informações falsas, enganosas ou gravemente distorcidas, revelando a centralidade da desinformação como instrumento de disputa política.

A análise qualitativa desse material permite identificar padrões recorrentes que extrapolam a lógica da disseminação espontânea. Observa-se, por exemplo, a replicação de narrativas idênticas ou estruturalmente semelhantes em diferentes grupos e regiões do país, frequentemente com pequenas variações de linguagem, enquadramento ou apelo emocional. Esse fenômeno sugere a existência de processos de adaptação estratégica de conteúdo, nos quais uma mesma mensagem-base é ajustada para públicos distintos, mantendo sua essência argumentativa enquanto maximiza sua ressonância local ou segmentada.

Além disso, a temporalidade da circulação desses conteúdos revela indícios adicionais de coordenação. Em diversos casos, mensagens com conteúdo semelhante surgem de forma quase simultânea em redes distintas de grupos, o que indica a existência de pontos centrais de produção e distribuição, responsáveis por alimentar múltiplos canais de disseminação. Tal padrão é inconsistente com dinâmicas orgânicas de compartilhamento, que tendem a apresentar maior variabilidade temporal e estrutural, e é mais compatível com operações coordenadas de comunicação política.

Do ponto de vista analítico, esses elementos permitem caracterizar a desinformação não apenas como um subproduto do ambiente digital, mas como uma estratégia deliberada, integrada à lógica mais ampla de campanhas orientadas por dados. A combinação entre segmentação de público, automação de envio e adaptação de conteúdo potencializa a eficácia dessas operações, permitindo que mensagens específicas atinjam públicos específicos em momentos estrategicamente definidos.

Essa constatação reforça a necessidade de compreender a desinformação como parte de uma infraestrutura comunicacional complexa, na qual produção, distribuição e recepção estão articuladas por sistemas técnicos e decisões estratégicas. Nesse contexto, o desafio regulatório não se limita à identificação de conteúdos falsos, mas envolve a compreensão das redes, dos mecanismos e dos incentivos que sustentam sua circulação em larga escala.

A INVISIBILIDADE COMO ESTRATÉGIA: POR QUE O ELEITOR NÃO VÊ O QUE ACONTECE

Uma das características mais perturbadoras da campanha eleitoral algorítmica reside em sua invisibilidade estrutural perante o eleitor. Diferentemente da propaganda eleitoral tradicional, cuja natureza pública permite o escrutínio coletivo e a comparação entre mensagens, a comunicação política mediada por sistemas digitais é concebida para operar de forma individualizada. Nesse modelo, diferentes eleitores são expostos a conteúdos distintos, frequentemente incompatíveis entre si, sem qualquer possibilidade de conhecimento recíproco. A experiência informacional torna-se, assim, fragmentada e assimétrica, comprometendo a formação de um espaço público minimamente compartilhado.

Essa invisibilidade manifesta-se em duas dimensões analiticamente distintas, embora interdependentes. A primeira é de natureza técnica e diz respeito à opacidade dos sistemas algorítmicos responsáveis pela curadoria e distribuição de conteúdo. Plataformas digitais utilizam modelos preditivos que definem, com base em múltiplas variáveis comportamentais, quais mensagens serão exibidas a cada usuário, sem que os critérios de seleção sejam acessíveis ou inteligíveis ao indivíduo afetado. Como resultado, o eleitor não dispõe de meios para compreender por que foi exposto a determinado conteúdo, nem para avaliar quais informações foram deliberadamente excluídas de seu horizonte informacional.

A segunda dimensão é de natureza política e está relacionada à intencionalidade estratégica das campanhas. A possibilidade de direcionamento diferenciado de mensagens permite a construção de narrativas múltiplas e, em alguns casos, contraditórias, ajustadas a segmentos específicos do eleitorado. Um mesmo candidato pode, por exemplo, enfatizar agendas distintas conforme o perfil do público-alvo, ou ainda direcionar conteúdos destinados a reduzir a participação eleitoral de grupos identificados como favoráveis ao adversário. Tais práticas, quando observadas em ambientes públicos, seriam prontamente reconhecidas como incompatíveis com padrões mínimos de lealdade eleitoral, mas, em contextos de comunicação privada e segmentada, tornam-se significativamente mais difíceis de detectar e contestar.

Essa combinação entre opacidade técnica e intencionalidade estratégica produz um ambiente no qual a assimetria informacional deixa de ser contingencial e passa a constituir um elemento estrutural da disputa política. A ausência de visibilidade compartilhada dificulta não apenas o controle institucional, mas também o próprio debate público, na medida em que diferentes eleitores passam a operar com conjuntos distintos de informações e enquadramentos interpretativos.

É precisamente em resposta a esse cenário que se insere a exigência de rotulagem de conteúdo gerado ou significativamente modificado por Inteligência Artificial, prevista na Resolução TSE nº 23.732/2024, editada pelo Tribunal Superior Eleitoral. Ao impor a identificação explícita desse tipo de conteúdo, o ordenamento busca mitigar os efeitos da opacidade algorítmica, introduzindo um elemento mínimo de transparência na comunicação digital. Embora tal medida não elimine a assimetria inerente aos sistemas de personalização, ela representa um esforço relevante no sentido de restabelecer condições mais equilibradas de informação, permitindo ao eleitor reconhecer a natureza mediada e potencialmente manipulável das mensagens que recebe.

Sob essa perspectiva, a rotulagem não deve ser compreendida apenas como uma exigência formal, mas como um instrumento normativo voltado à preservação da integridade do processo democrático em um ambiente comunicacional profundamente transformado pela tecnologia.

HABERMAS, BOBBIO E A AMEAÇA AO ESPAÇO PÚBLICO DEMOCRÁTICO

A compreensão teórica da ameaça que a campanha algorítmica representa à democracia exige o recurso a categorias clássicas da teoria política, capazes de iluminar não apenas os efeitos empíricos dessas transformações, mas sobretudo suas implicações normativas. Entre os autores que oferecem chaves interpretativas particularmente fecundas, destacam-se Jürgen Habermas e Norberto Bobbio, cujas formulações permitem compreender, em níveis distintos, a tensão entre as novas tecnologias de comunicação política e os fundamentos da legitimidade democrática.

No âmbito da teoria habermasiana, a democracia é concebida como um processo comunicativo no qual a formação da vontade política depende da existência de um espaço público estruturado por condições de deliberação racional. Em obras como *Teoria da Ação Comunicativa* e *Direito e Democracia*, Jürgen Habermas sustenta que a legitimidade das decisões políticas decorre da possibilidade de que os cidadãos participem, direta ou indiretamente, de processos discursivos orientados pelo melhor argumento, em condições de simetria comunicativa. O espaço público, nesse sentido, não é apenas um ambiente de circulação de informações, mas um arranjo normativo no qual prevalece a racionalidade comunicativa, entendida como a capacidade de justificar pretensões de validade por meio de razões suscetíveis de aceitação intersubjetiva.

A campanha algorítmica, contudo, introduz uma ruptura estrutural nesse modelo. Em vez de se orientar pela troca pública de argumentos, ela opera segundo uma lógica de racionalidade estratégica, na qual a comunicação é instrumentalizada com o objetivo de produzir efeitos comportamentais específicos. A centralidade da modelagem psicográfica e da segmentação de audiência desloca o foco da persuasão racional para a exploração de disposições subjetivas individuais, muitas vezes em nível pré-reflexivo. A mensagem política deixa de ser concebida como argumento a ser avaliado em um espaço público compartilhado e passa a ser calibrada como estímulo direcionado, ajustado às vulnerabilidades cognitivas e emocionais de cada eleitor.

Essa transformação pode ser compreendida, na linguagem habermasiana, como uma forma de colonização do mundo da vida pela racionalidade instrumental. Os sistemas técnicos que mediam a comunicação política passam a impor lógicas operacionais que não se orientam pela busca de entendimento, mas pela maximização de resultados estratégicos, como engajamento, conversão ou mobilização eleitoral. O resultado é a progressiva erosão das condições que tornam possível a deliberação democrática, na medida em que o espaço público é fragmentado em múltiplas esferas privadas de comunicação, nas quais não há visibilidade recíproca nem possibilidade efetiva de contestação discursiva.

O próprio princípio do discurso, formulado por Jürgen Habermas, explicita o alcance dessa tensão ao estabelecer que apenas são válidas as normas de ação às quais todos os possíveis atingidos poderiam dar seu assentimento na qualidade de participantes de discursos racionais. Em um ambiente informacional no qual diferentes indivíduos são expostos a conteúdos distintos, potencialmente contraditórios e invisíveis entre si, torna-se difícil sustentar que as condições para tal assentimento estejam presentes. A fragmentação informacional compromete a universalização das pretensões de validade e, com isso, enfraquece o próprio fundamento discursivo da legitimidade democrática.

A contribuição de Norberto Bobbio permite aprofundar essa análise a partir de uma perspectiva mais institucional e procedimental. Em sua concepção mínima de democracia, Bobbio identifica como elemento essencial a realização de eleições livres, periódicas e competitivas. A liberdade do processo eleitoral, nesse contexto, não se limita à ausência de coerção física ou jurídica, mas inclui a garantia de que os eleitores possam formar suas preferências de maneira autônoma, sem interferências indevidas que comprometam sua capacidade de escolha.

A incorporação de sistemas algorítmicos na comunicação política tensiona diretamente esse requisito. Quando a formação de preferências passa a ser influenciada por mecanismos opacos, operados por atores com acesso desigual a recursos tecnológicos e informacionais, a assimetria deixa de ser apenas contingente e passa a estruturar o próprio processo eleitoral. A manipulação não se apresenta de forma explícita, como na propaganda coercitiva tradicional, mas opera de maneira difusa, personalizada e muitas vezes imperceptível, o que dificulta sua identificação e contestação.

Nesse sentido, torna-se possível afirmar que determinadas práticas de campanha algorítmica preservam a forma da democracia, expressa na realização do ato de votar, ao mesmo tempo em que fragilizam sua substância, entendida como a formação livre e informada da vontade política. A distinção entre forma e substância, central na obra de Norberto Bobbio, revela-se particularmente útil para compreender esse fenômeno. A existência de eleições não é, por si só, garantia de democracia, se as condições que antecedem o voto estiverem comprometidas por processos sistemáticos de manipulação informacional.

A articulação entre as perspectivas de Jürgen Habermas e Norberto Bobbio permite, assim, identificar uma dupla crise. De um lado, uma crise da deliberação, marcada pela fragmentação do espaço público e pela substituição da racionalidade comunicativa por estratégias de influência individualizada. De outro, uma crise da legitimidade procedimental, na medida em que a liberdade de formação das preferências eleitorais é tensionada por mecanismos tecnológicos de persuasão assimétrica.

Essa dupla dimensão evidencia que o problema central não reside apenas na existência de novas tecnologias, mas na forma como elas reconfiguram as condições estruturais da vida democrática. A tecnologia, enquanto instrumento, não é portadora de uma direção normativa intrínseca. No entanto, sua incorporação em contextos institucionais específicos pode produzir efeitos que ampliam ou restringem valores democráticos fundamentais, como autonomia, igualdade e transparência.

Diante desse quadro, a questão que se impõe não é se a tecnologia deve ou não ser utilizada na política, mas sob quais condições normativas, institucionais e culturais sua utilização pode ser compatibilizada com os princípios da democracia. A resposta a esse desafio não será encontrada exclusivamente no desenvolvimento tecnológico, mas nas escolhas regulatórias, nas práticas institucionais e na capacidade crítica da sociedade civil.

Em última instância, trata-se de definir se a democracia será capaz de adaptar seus mecanismos de proteção e legitimação a um ambiente comunicacional profundamente transformado, ou se será progressivamente moldada pelas lógicas operacionais das tecnologias que hoje mediam a esfera pública.

■ RESUMO DO CAPÍTULO:

Este capítulo demonstrou que as eleições contemporâneas não podem mais ser compreendidas apenas a partir de suas dimensões institucionais e comunicacionais tradicionais. A incorporação de tecnologias digitais avançadas introduziu uma camada estrutural adicional, caracterizada por sua invisibilidade ao eleitor comum e por sua centralidade na definição dos resultados eleitorais. A campanha algorítmica foi apresentada como um modelo operacional composto por cinco camadas interdependentes: coleta de dados, modelagem psicográfica, segmentação estratégica, produção automatizada de conteúdo e distribuição direcionada. Esse modelo representa uma ruptura qualitativa em relação às formas históricas de comunicação política, na medida em que altera simultaneamente a escala, a precisão e a transparência das estratégias de campanha. A escala decorre da capacidade de produzir e distribuir mensagens individualizadas para milhões de eleitores de forma simultânea. A precisão resulta do uso de modelos preditivos capazes de inferir traços psicológicos e comportamentais a partir de dados digitais. A invisibilidade, por sua vez, constitui o elemento mais crítico, pois impede que o eleitor tenha consciência das condições sob as quais sua percepção política está sendo influenciada. Do ponto de vista teórico, o capítulo evidenciou que esse novo paradigma tensiona fundamentos clássicos da democracia. A perspectiva de Jürgen Habermas indica uma crise da deliberação pública, marcada pela fragmentação do espaço comunicativo e pela substituição da racionalidade argumentativa por estratégias de influência individualizada. Já a abordagem de Norberto Bobbio evidencia uma tensão entre a forma e a substância da democracia, na medida em que a preservação do ato formal do voto pode coexistir com a erosão das condições de formação autônoma das preferências eleitorais. Diante desse cenário, a questão central que orienta esta obra pode ser reformulada nos seguintes termos: embora o eleitor permaneça, do ponto de vista formal, como o agente decisório do processo democrático, a crescente mediação tecnológica sugere que algoritmos, dados e atores com capacidade de operá-los exercem influência material cada vez mais significativa sobre os resultados eleitorais. A investigação dessa hipótese constitui o eixo analítico dos capítulos subsequentes.

QUESTÕES PARA REFLEXÃO E DEBATE

1. É possível sustentar a legitimidade de um processo eleitoral no qual os eleitores são expostos a conjuntos distintos e potencialmente contraditórios de informações?
2. A modelagem psicográfica pode ser considerada uma forma legítima de persuasão política ou configura, em determinados contextos, uma forma de manipulação incompatível com a democracia?
3. Como a distinção proposta por Norberto Bobbio entre forma e substância democrática ajuda a compreender os impactos da campanha algorítmica?
4. A opacidade dos algoritmos utilizados na distribuição de conteúdo político pode ser compatibilizada com os princípios de transparência e accountability eleitoral?
5. Quais são os limites éticos e jurídicos do uso de Inteligência Artificial na produção e disseminação de conteúdo político?

6. A regulação estatal, como a estabelecida pelo Tribunal Superior Eleitoral, é suficiente para mitigar os riscos da campanha algorítmica, ou são necessárias novas formas de controle?

7. Em que medida o eleitor contemporâneo possui condições reais de autonomia decisória em um ambiente informacional altamente mediado por sistemas algorítmicos?

8. A tecnologia deve ser compreendida como instrumento neutro ou como elemento que incorpora e reproduz relações de poder no processo democrático?

9. Quais estratégias institucionais e sociais podem ser desenvolvidas para preservar a integridade da democracia em um contexto de crescente digitalização da política?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas?

1. Você não vê a campanha que está sendo feita para você.

Enquanto assiste ao debate na televisão, outra disputa acontece no seu smartphone. Sistemas automatizados analisam suas reações em tempo real, publicidade segmentada é recalibrada minuto a minuto e conteúdos específicos são enviados para grupos de WhatsApp de acordo com o seu perfil. Você recebe uma versão da campanha. Seu vizinho, com perfil diferente, recebe outra. Nenhum dos dois sabe o que o outro está vendo.

2. Suas curtidas dizem mais sobre você do que você imagina.

A pesquisa de Kosinski, Stillwell e Graepel demonstrou que curtidas em redes sociais permitem inferir traços de personalidade com precisão comparável à de familiares próximos. Para profissionais de campanha, isso significa que o comportamento digital do eleitor, mesmo o mais aparentemente trivial, é matéria-prima para modelagem psicográfica. Para o eleitor, significa que cada interação digital é um dado fornecido a sistemas que ele não controla e cujo funcionamento não conhece.

3. A campanha tem cinco camadas. O eleitor só vê uma.

A superfície visível da campanha, os debates, os comícios, os anúncios, corresponde apenas à camada de distribuição. As quatro camadas anteriores (coleta de dados, modelagem psicográfica, segmentação e produção de conteúdo) operam de forma invisível ao eleitor comum e são, em muitos aspectos, mais determinantes do resultado eleitoral do que aquilo que aparece nas telas. Quem assessora campanhas precisa dominar todas as cinco. Quem fiscaliza eleições precisa saber que elas existem.

4. O WhatsApp não é apenas um canal. É uma zona de baixíssima fiscalização.

As eleições de 2018 demonstraram que o WhatsApp funciona como infraestrutura de campanha precisamente porque opera em ambiente privado, com criptografia de ponta a ponta e rastreabilidade limitada. Mensagens com desinformação circularam em escala industrial, amplificadas por disparos automáticos, sem que houvesse mecanismo efetivo de monitoramento em tempo real. Para campanhas que trabalham no limite da legalidade, esse ambiente é estrategicamente atraente. Para a Justiça Eleitoral, é o ponto cego mais difícil de cobrir.

5. Conteúdo gerado por IA sem rotulagem é infração eleitoral.

A Resolução TSE nº 23.732/2024 exige identificação explícita de todo conteúdo gerado ou significativamente modificado por inteligência artificial. Isso inclui textos, imagens, áudios e vídeos. A omissão configura infração eleitoral. Na prática, campanhas que usam IA generativa para produzir conteúdo em escala precisam de processos internos de controle que garantam a rotulagem adequada

antes da publicação ou do disparo. A ausência desse controle expõe candidatos a responsabilização mesmo quando o conteúdo foi produzido por terceiros.

6. Mensagens diferentes para públicos diferentes podem configurar abuso de poder.

A segmentação algorítmica permite enviar mensagens distintas, e até contraditórias, para diferentes grupos de eleitores. O que é estratégia de marketing em outros contextos pode, no ambiente eleitoral, configurar uso indevido dos meios de comunicação ou abuso de poder econômico, especialmente quando combinado com dados obtidos sem consentimento ou com financiamento não declarado. O TSE tem construído jurisprudência sobre esse tema a partir de casos concretos das eleições de 2018 e 2022, e as sanções podem chegar à cassação de mandato.

7. Habermas e Bobbio em linguagem de campanha: o eleitor precisa de informação real para votar de verdade.

A teoria democrática não é abstração acadêmica. Habermas diz que a legitimidade política depende de argumentos trocados em espaço público compartilhado. Bobbio diz que eleição livre exige formação autônoma de preferências. Traduzindo para a prática: uma campanha que fragmenta o espaço informacional, expõe eleitores a versões incompatíveis da realidade e explora vulnerabilidades cognitivas pode vencer formalmente uma eleição enquanto corrói materialmente a democracia. Essa distinção entre forma e substância é o núcleo do debate regulatório que define o campo em que este livro se insere.

CAPÍTULO 2

Como os Algoritmos Passaram a Decidir o Debate Público

Bolhas de Filtro, Câmaras de Eco e a Arquitetura da Polarização

ALGORITMOS E DEMOCRACIA

O campo de estudos que investiga a relação entre algoritmos e democracia consolidou-se nas últimas duas décadas em resposta à expansão das plataformas digitais e à crescente evidência empírica de seus efeitos sobre a polarização política, a circulação de desinformação e a reconfiguração das dinâmicas de formação da opinião pública. A emergência desse campo não representa apenas a incorporação de novos objetos de estudo, mas a necessidade de revisão de categorias clássicas da teoria democrática à luz de transformações tecnológicas que alteram profundamente as condições de produção, circulação e recepção da informação política.

Três tradições teóricas distintas, embora progressivamente convergentes, estruturam esse debate contemporâneo: A primeira tradição é a dos estudos de comunicação e esfera pública. A formulação clássica de Jürgen Habermas estabelece a esfera pública como um espaço de deliberação racional, no qual a legitimidade democrática decorre da força do melhor argumento em condições de simetria comunicativa. No contexto digital, essa concepção é tensionada por transformações estruturais introduzidas pelas plataformas. Autores como Peter Dahlgren destacam a ampliação das oportunidades de participação política proporcionadas pelos meios digitais, enquanto Zizi Papacharissi argumenta que essas mesmas plataformas produzem uma “esfera pública afetiva”, marcada pela centralidade das emoções e pela fragmentação discursiva. O debate contemporâneo nessa tradição desloca-se, portanto, da pergunta sobre acesso à esfera pública para a análise da qualidade das interações que nela ocorrem.

A segunda tradição é a da economia comportamental e da psicologia social, que fornece os instrumentos analíticos para compreender por que os sistemas algorítmicos são tão eficazes em influenciar comportamentos. A obra de Daniel Kahneman demonstra que grande parte das decisões humanas é mediada por processos cognitivos automáticos, baseados em heurísticas e suscetíveis a vieses sistemáticos. A partir dessa constatação, Richard Thaler e Cass Sunstein desenvolvem o conceito de arquitetura de escolha,

evidenciando como o ambiente decisório pode ser estruturado para induzir comportamentos previsíveis. Já Robert Cialdini identifica princípios de persuasão que são amplamente explorados em contextos digitais, como prova social, escassez e autoridade. No contexto das plataformas, esses mecanismos deixam de ser aplicados de forma pontual e passam a ser operacionalizados em larga escala por sistemas automatizados.

A terceira tradição é a dos estudos críticos de tecnologia, que desloca a análise do nível comportamental para o nível estrutural. Shoshana Zuboff demonstra que a lógica de funcionamento das plataformas digitais está ancorada na extração e monetização de dados comportamentais, configurando um modelo econômico que incentiva a maximização do engajamento. Safiya Umoja Noble evidencia que sistemas algorítmicos não apenas organizam informação, mas reproduzem hierarquias sociais e vieses estruturais. Virginia Eubanks, por sua vez, revela como sistemas automatizados podem reforçar desigualdades na distribuição de recursos e oportunidades. Essa tradição é fundamental para compreender que os algoritmos não operam em um vazio normativo, mas estão inseridos em estruturas de poder e interesse econômico.

A convergência dessas três tradições permite avançar para uma compreensão integrada do fenômeno. Os algoritmos podem ser simultaneamente compreendidos como mediadores da esfera pública, exploradores de vulnerabilidades cognitivas e instrumentos de modelos econômicos orientados por incentivos específicos. Essa abordagem multidimensional revela que os efeitos políticos dos sistemas algorítmicos não são acidentais, mas resultam da interação entre arquitetura tecnológica, comportamento humano e estruturas de poder.

Nesse sentido, a maximização do engajamento, objetivo central das plataformas digitais, produz consequências normativamente relevantes. Conteúdos que despertam emoções intensas, especialmente indignação e medo, tendem a ser priorizados pelos algoritmos, o que contribui para a amplificação de discursos polarizadores e para a degradação da qualidade do debate público. A lógica de recomendação algorítmica, ao privilegiar relevância individual em detrimento de visibilidade coletiva, reforça a fragmentação informacional e dificulta a construção de referenciais comuns de realidade.

Essa dinâmica coloca em evidência uma tensão central: os sistemas que organizam o fluxo de informação na sociedade contemporânea não foram projetados com objetivos democráticos, mas econômicos. No entanto, seus efeitos são profundamente políticos. A

ausência de alinhamento entre esses dois níveis constitui um dos principais desafios teóricos e normativos do campo.

É nesse ponto que esta obra se posiciona. Ao integrar as contribuições dessas três tradições, propõe-se a analisar a campanha algorítmica não apenas como inovação tecnológica, mas como fenômeno político que reconfigura as condições estruturais da democracia contemporânea. A hipótese que orienta essa análise é a de que a mediação algorítmica da comunicação política não apenas influencia resultados eleitorais, mas altera, de forma mais profunda, os próprios critérios de legitimidade democrática.

Para compreender como a IA amplificou os riscos à integridade eleitoral, é necessário primeiro compreender o ecossistema informacional contemporâneo e dois fenômenos que o caracterizam: a pós-verdade e a infodemia:

Tabela 05. Fenômenos do Ecossistema Informacional Contemporâneo

Fenômeno	Definição	Mecanismo	Impacto Eleitoral
Pós-verdade	Circunstâncias nas quais fatos objetivos influenciam menos a opinião pública do que apelos à emoção e crenças pessoais. Palavra do Ano do Oxford Dictionary em 2016.	Conteúdo emocionalmente ativador supera conteúdo factualmente preciso em alcance e engajamento.	Eleitores votam com base em como se sentem sobre um candidato, não sobre o que ele propõe. IA generativa é muito boa em gerar conteúdo emocionalmente ativador.
Infodemia	Proliferação excessiva e rápida de informações, corretas e incorretas, dificultando que as pessoas encontrem fontes confiáveis. Conceito da OMS (COVID-19).	Espalhamento epidêmico de conteúdo por redes de contato, com aceleração exponencial em plataformas de alto alcance.	Eleitores incapazes de distinguir informação confiável de desinformação no calor da campanha.
Bolha de filtro	Ambiente informacional único construído pelos algoritmos para cada usuário, calibrado às suas preferências, progressivamente distante de perspectivas divergentes (Pariser, 2011).	Filtragem algorítmica passiva: o usuário raramente percebe a personalização.	Erosão das bases epistemológicas compartilhadas necessárias para a deliberação democrática.

Câmara de eco	Ambiente informacional fechado no qual crenças são amplificadas e reforçadas pela ausência de perspectivas externas. Produzida pela escolha ativa do usuário.	Homogeneidade construída pelo usuário: segue apenas fontes que confirmam sua visão.	Radicalização progressiva, groupthink, hostilidade ao diferente.
Polarização digital	Intensificação da divisão política acima do que seria esperado por diferenças genuínas de preferências, alimentada por mecanismos digitais de amplificação de conflito.	Indignação e conteúdo de alta intensidade emocional recebem mais engajamento, criando incentivo algorítmico para conteúdo extremado.	Incapacidade de diálogo político entre grupos, legitimação de candidatos extremistas.
Ecologia da desinformação	Ecossistema de produtores, distribuidores e consumidores de desinformação: fábricas de conteúdo, redes de amplificação (bots + cyborgs), plataformas e receptores.	Cada elemento tem papel específico e responde a incentivos distintos.	Intervenções precisam ser desenhadas para cada elo da cadeia, não apenas para um.

Fonte: Elaborado pelo autor com base em Oxford Dictionary (2016), OMS (2020), Pariser (2011).

A LÓGICA DO ENGAJAMENTO: COMO OS ALGORITMOS DECIDEM O QUE VOCÊ VÊ

Para compreender o impacto dos algoritmos sobre o debate político, é necessário partir da lógica econômica que orienta seu funcionamento. As principais plataformas digitais contemporâneas, como Meta, Google, ByteDance e X, operam segundo um modelo de negócios baseado na monetização da atenção. Nesse modelo, o tempo de permanência do usuário constitui o principal ativo econômico, na medida em que está diretamente relacionado ao volume de anúncios que podem ser exibidos e, consequentemente, à receita gerada.

Essa estrutura econômica não apenas condiciona, mas determina os objetivos dos sistemas algorítmicos que organizam o fluxo de informação nessas plataformas. Os algoritmos de recomendação são concebidos como sistemas de otimização cujo objetivo central é maximizar métricas de engajamento, como tempo de visualização, número de interações e frequência de retorno do usuário. Para atingir esse objetivo, tais sistemas utilizam técnicas avançadas de aprendizado de máquina, incluindo aprendizado por reforço,

por meio das quais ajustam continuamente suas estratégias de recomendação com base no comportamento observado dos usuários. Cada interação, como cliques, curtidas, compartilhamentos ou tempo de visualização, funciona como um sinal que retroalimenta o sistema, permitindo o refinamento progressivo das previsões sobre o que manterá o usuário engajado.

O resultado desse processo é a construção de ambientes informacionais altamente personalizados, nos quais o conteúdo exibido não é resultado de uma curadoria editorial tradicional, mas de decisões automatizadas orientadas por probabilidade de engajamento. O feed de cada usuário torna-se, assim, uma representação dinâmica de suas preferências comportamentais, construída a partir de dados históricos e continuamente atualizada. Essa personalização confere ao conteúdo uma aparência de relevância imediata e proximidade subjetiva, na medida em que reflete padrões previamente identificados no comportamento do próprio usuário.

O problema democrático emerge precisamente da natureza dos conteúdos que tendem a maximizar essas métricas de engajamento. A literatura empírica indica de forma consistente que conteúdos emocionalmente intensos apresentam desempenho superior em termos de disseminação e interação. Emoções como indignação, medo e surpresa ativam respostas cognitivas rápidas e incentivam comportamentos de compartilhamento, ampliando o alcance dessas mensagens. Em contraste, conteúdos informativos mais equilibrados, complexos ou nuançados tendem a gerar menor engajamento imediato, sendo, portanto, menos favorecidos pelos sistemas de recomendação.

O estudo conduzido por Soroush Vosoughi, Deb Roy e Sinan Aral, publicado na revista Science em 2018, fornece evidência robusta desse fenômeno ao demonstrar que notícias falsas se propagam significativamente mais rápido e alcançam um número maior de usuários do que notícias verdadeiras. A explicação para esse padrão não reside apenas na arquitetura das plataformas, mas na interação entre essa arquitetura e as predisposições cognitivas dos usuários. Conteúdos falsos frequentemente apresentam maior novidade, carga emocional e potencial de surpresa, características que aumentam sua probabilidade de engajamento e, consequentemente, sua priorização algorítmica.

Evidência Empírica Fundamental

Soroush Vosoughi, Deb Roy e Sinan Aral, em estudo publicado na Science (2018), analisaram 126.000 rumores circulados no Twitter entre 2006 e 2017. Resultado: notícias falsas espalharam-se, em média, 6 vezes mais rápido que notícias verdadeiras, chegaram a 10 vezes mais pessoas e penetraram muito mais profundamente nas redes sociais. O fator explicativo não foram os bots. Quando controlados estatisticamente, os humanos foram o principal vetor de propagação de desinformação. A razão: conteúdo falso é, na média, mais novo, surpreendente e emocionalmente ativador, características que o humano, e o algoritmo otimizado para engajamento, tende a privilegiar.

Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. Science, v. 359, n. 6380, p. 1146-1151, 2018.

Essa dinâmica revela uma tensão estrutural entre a lógica econômica das plataformas e os requisitos normativos da democracia. Enquanto o funcionamento dos sistemas algorítmicos é orientado pela maximização do engajamento, a qualidade do debate público depende da circulação de informações confiáveis, da exposição a perspectivas diversas e da possibilidade de reflexão crítica. A ausência de alinhamento entre esses objetivos produz um ambiente no qual conteúdos que degradam o debate democrático podem ser sistematicamente amplificados, não por intenção política direta, mas como consequência de um modelo econômico que privilegia a atenção como recurso escasso.

Nesse contexto, os algoritmos deixam de atuar como meros intermediários na circulação de informação e passam a exercer função estruturante na esfera pública digital. Ao definir quais conteúdos são visíveis, em que ordem e com que frequência, esses sistemas influenciam diretamente a agenda pública, a formação de preferências e a dinâmica do debate político.

A análise desse fenômeno pode ser aprofundada pela identificação das características que tornam determinados conteúdos mais propensos à priorização algorítmica. Tais características não operam de forma isolada, mas articulam efeitos cognitivos e políticos que ajudam a explicar a predominância de certos tipos de mensagem no ambiente informacional digital.

A literatura especializada permite, nesse sentido, sistematizar um conjunto de atributos recorrentes associados ao conteúdo de alto engajamento, evidenciando como disposições psicológicas individuais se conectam a dinâmicas coletivas de formação da opinião pública. O quadro a seguir sintetiza essas relações, destacando os vínculos entre

características do conteúdo, efeitos cognitivos, desdobramentos políticos e riscos para a democracia.

Tabela 06. Características do Conteúdo de Alto Engajamento

Característica	Efeito Cognitivo	Efeito Político	Risco Democrático
Confirmação de crenças	Viés de confirmação	Radicalização de posições	Câmaras de eco
Alta carga emocional	Ativação do sistema límbico	Decisões afetivas vs. racionais	Pós-verdade
Novidade e surpresa	Atenção involuntária	Amplificação de desinformação	Infodemia
Antagonismo grupal	Identidade social	Polarização afetiva	Tribalisismo político
Urgência e ameaça	Resposta ao estresse	Mobilização do medo	Autoritarismo eleitoral

Fonte: Elaborado pelo autor com base em Kahneman (2011), Vosoughi, Roy e Aral (2018) e Sunstein (2017).

A leitura do quadro revela um padrão consistente: em todos os casos, a característica que maximiza o engajamento individual é a mesma que produz o maior risco coletivo para a democracia. Não se trata de coincidência, mas de uma tensão estrutural inscrita na própria arquitetura das plataformas.

Estudo de Caso Aprofundado: Cambridge Analytica

O caso Cambridge Analytica é o estudo de caso fundador do campo da regulação de IA eleitoral — não porque foi o primeiro uso de dados em campanha, mas porque foi o primeiro a expor, em escala global e com documentação detalhada, os mecanismos pelos quais dados de cidadãos coletados sem consentimento podem ser instrumentalizados para manipulação psicológica de massa em contexto eleitoral.

Ele é, simultaneamente, um caso de violação de dados, um caso de manipulação eleitoral, um caso de responsabilização de plataformas — e, para o Brasil, um caso de origem de transformações regulatórias importantes.

Cambridge Analytica — Os fatos essenciais

A Cambridge Analytica (CA), empresa britânica fundada por Alexander Nix e financiada pelo bilionário Robert Mercer, utilizou um aplicativo de pesquisa ('thisisyourdigitallife') desenvolvido pelo Dr. Aleksandr Kogan para coletar dados de 87 milhões de usuários do Facebook sem o consentimento deles. O app foi instalado por apenas 270.000 usuários, mas explorava uma API do Facebook que permitia acesso aos dados de todos os amigos de cada instalador — sem o conhecimento ou consentimento desses amigos. Os dados coletados foram usados para construir perfis psicográficos baseados no modelo OCEAN (Abertura, Conscienciosidade, Extroversão, Amabilidade, Neuroticismo) de dezenas de milhões de americanos. Esses perfis alimentaram o sistema de microtargeting eleitoral da campanha de Donald Trump em 2016 — identificando subgrupos de eleitores particularmente suscetíveis a determinadas mensagens e, crucialmente, identificando eleitores democratas que poderiam ser desestimulados a votar por meio de narrativas negativas direcionadas (*voter suppression digital*).

Dimensão	O que aconteceu	Consequência jurídica	Lição para o Brasil
Coleta de dados	87 milhões de perfis coletados via app com autorização de 270.000 — exploração de API do Facebook que permitia acesso a dados de amigos não instaladores	Facebook: multa US$ 5 bilhões (FTC, 2019). ICO britânico: £500.000. Nenhuma condenação criminal individual.	A LGPD exige consentimento específico e destacado para dados sensíveis (opinião política). Coleta indireta via amigos sem consentimento seria ilegal no Brasil desde 2020.
Perfil psicográfico	Modelo OCEAN inferido a partir de curtidas — relacionando comportamento digital a traços de personalidade e vulnerabilidades psicológicas específicas	Dado político inferido (opinião política) = dado sensível pela LGPD, art. 5º, II. Sem base legal para tratamento.	A inferência de opinião política a partir de dados não sensíveis exige consentimento específico — questão jurídica ainda não resolvida definitivamente pela ANPD.
Microtargeting psicográfico	Mensagens calibradas para explorar vulnerabilidades específicas de cada perfil OCEAN — especialmente neuroticismo alto (suscetibilidade a medo) para narrativas de ameaça	Não há tipo penal específico nos EUA ou Brasil. Na teoria, pode configurar manipulação eleitoral ilícita se demonstrado dolo.	Res. TSE 23.732/2024 proíbe sistemas que 'utilizem técnicas subliminares ou deliberadamente manipulativas' — mas threshold probatório é elevado.
Voter suppression digital	Identificação de eleitores democratas potencialmente persuasíveis para 'descampaign' — envio de narrativas negativas sobre Clinton para desestimulá-los a votar	Voter suppression é crime federal nos EUA (Voting Rights Act). No Brasil, desestímulo ao voto por meios fraudulentos pode configurar crime eleitoral.	Art. 222-A da Lei 9.504/97 proíbe disparo em massa para fins ilícitos — inclui desestímulo fraudulento ao comparecimento.

■■ **Desdobramentos do caso Cambridge Analytica para o contexto brasileiro**

O caso Cambridge Analytica influenciou diretamente três desenvolvimentos no Brasil: (1) LGPD aprovada em agosto de 2018 — o escândalo, revelado em março de 2018, acelerou a aprovação da lei que havia sido debatida por anos sem urgência; (2) O TSE instaurou investigação sobre uso de serviços similares em campanhas brasileiras de 2018 — investigação inconclusiva por dificuldade de acesso a dados, mas que gerou alertas regulatórios; (3) A Resolução TSE 23.732/2024 incorpora explicitamente, em seus fundamentos, a vedação a técnicas de manipulação psicológica similares às utilizadas pela CA — como resposta normativa direta à lição do caso.

BOLHAS DE FILTRO: A TEORIA E A EVIDÊNCIA

O conceito de bolha de filtro, introduzido por Eli Pariser, descreve o ambiente informacional individualizado que resulta da operação de sistemas algorítmicos de personalização. Nesses sistemas, cada usuário é exposto predominantemente a conteúdos selecionados com base em sua probabilidade de engajamento, o que implica não apenas a priorização de informações consideradas relevantes, mas também a exclusão sistemática de conteúdos que não se alinham aos padrões comportamentais previamente identificados. O efeito agregado é a construção de ecossistemas informacionais assimétricos, nos quais diferentes indivíduos passam a habitar realidades informativas parcialmente desconectadas entre si.

A formulação original de Pariser teve o mérito de tornar visível a dimensão oculta da curadoria algorítmica. No entanto, foi posteriormente submetida a escrutínio empírico que revelou um quadro consideravelmente mais complexo. O estudo de Bakshy, Messing e Adamic, baseado em dados de usuários do Facebook, demonstrou que, embora a personalização algorítmica contribua para a redução da exposição a conteúdos ideologicamente divergentes, esse efeito é superado pelas próprias escolhas dos usuários, que tendem a selecionar ativamente fontes alinhadas às suas preferências preexistentes.

Esse resultado é teoricamente significativo: desloca o problema de uma explicação exclusivamente tecnológica para uma perspectiva que integra dimensões algorítmicas e comportamentais, recusando tanto o determinismo técnico quanto a ilusão de uma agência individual plenamente autônoma.

A partir dessa constatação, torna-se mais preciso compreender a bolha de filtro como um fenômeno sociotécnico, resultante da interação entre sistemas de recomendação e predisposições cognitivas dos usuários. Os algoritmos não criam, *ex nihilo*, a segmentação informacional; eles a intensificam e a estabilizam ao longo do tempo, ao retroalimentar padrões de consumo já existentes.

Cada escolha individual, ao ser capturada como dado, reforça as inferências do sistema e contribui para a redução progressiva da diversidade informacional no ambiente do usuário. Forma-se, assim, um ciclo de reforço no qual preferências individuais e curadoria algorítmica se amplificam mutuamente, tornando cada vez mais improvável a exposição espontânea a perspectivas discordantes.

Fragmentação do Espaço Público

No plano coletivo, esse processo produz efeitos que ultrapassam a experiência individual e incidem sobre a própria estrutura do debate público. A fragmentação informacional decorrente da personalização algorítmica dificulta a constituição de um repertório comum de fatos e argumentos, condição que a teoria democrática deliberativa historicamente considerou indispensável para a legitimidade da deliberação coletiva. Em lugar de um espaço público relativamente integrado, observa-se a emergência de múltiplas esferas comunicativas paralelas, com baixa interconexão e circulação restrita de conteúdos entre grupos ideologicamente distintos.

No contexto brasileiro, a relevância empírica desse fenômeno foi documentada por Araújo e Oliveira, que identificaram, durante as eleições de 2018, a formação de comunidades informacionais altamente polarizadas na plataforma X. A análise das redes de interação revelou baixa sobreposição entre grupos associados a diferentes candidatos, indicando a existência de clusters relativamente isolados, nos quais a circulação de informação ocorria predominantemente de forma endógena. Esse padrão é consistente com a hipótese de segmentação informacional e sugere que a exposição a perspectivas divergentes foi significativamente limitada ao longo do processo eleitoral, com consequências que se estendem para além do resultado das urnas.

Implicações para a Teoria Democrática

As implicações normativas desse cenário são de longo alcance. A teoria democrática deliberativa, especialmente na tradição habermasiana, pressupõe condições mínimas de comunicação entre cidadãos com posições divergentes, de modo que argumentos possam ser confrontados e avaliados em um espaço público compartilhado. A fragmentação promovida pelas bolhas de filtro compromete essas condições ao reduzir as oportunidades de interação entre diferentes visões de mundo, favorecendo a cristalização de crenças e o endurecimento progressivo das posições políticas.

Há, contudo, uma dimensão mais sutil e igualmente preocupante. A redução da exposição a conteúdos divergentes não apenas restringe o acesso à informação; ela reconfigura a forma como os indivíduos percebem a própria legitimidade do dissenso. Em ambientes informacionais altamente homogêneos, opiniões discordantes tendem a ser interpretadas não como posições a serem refutadas argumentativamente, mas como anomalias ou ameaças à coesão do grupo. Esse mecanismo alimenta a polarização afetiva e contribui para a deslegitimação do adversário político, dificultando a construção dos consensos mínimos que o funcionamento das instituições democráticas requer.

A bolha de filtro, portanto, não deve ser compreendida apenas como um problema de acesso à informação. Trata-se de um fenômeno que opera na interseção entre tecnologia, comportamento e estrutura social, e que reconfigura as condições estruturais de formação da opinião pública. Ao substituir a mediação argumentativa pela gestão silenciosa da segmentação, ela contribui para a emergência de um ambiente comunicacional no qual a divergência deixa de ser enfrentada e passa a ser evitada. O resultado é uma esfera pública fragmentada, na qual a coexistência de múltiplas realidades informacionais desafia não apenas a qualidade do debate democrático, mas os próprios fundamentos sobre os quais esse debate se sustenta.

■ **Evidência empírica — Velocidade de propagação de desinformação**
Soroush Vosoughi, Deb Roy e Sinan Aral, em estudo publicado na Science (2018), analisaram 126.000 rumores circulados no Twitter entre 2006 e 2017 por 3 milhões de usuários. Resultado: notícias falsas se espalharam, em média, 6 vezes mais rápido que notícias verdadeiras, chegaram a 10 vezes mais pessoas e penetraram muito mais profundamente nas redes sociais. O fator explicativo não foram os bots, quando controlados estatisticamente, os humanos foram o principal vetor de propagação de desinformação. A razão: conteúdo falso é, na média, mais novo, surpreendente e emocionalmente ativador, características que o humano, e o algoritmo otimizado para engajamento, tende a privilegiar.

Câmaras de Eco e Radicalização Algorítmica

Relacionado, mas conceitualmente distinto da bolha de filtro, o fenômeno das câmaras de eco descreve o processo pelo qual grupos de indivíduos com crenças semelhantes, ao interagirem predominantemente entre si, tendem a radicalizar suas posições pela ausência de contato com perspectivas críticas ou divergentes. A distinção conceitual importa: enquanto a bolha de filtro é essencialmente um fenômeno de curadoria,

no qual o algoritmo seleciona o que o usuário vê, a câmara de eco é um fenômeno de homofilia, no qual são os próprios padrões de interação social que determinam a composição do ambiente informacional. Os algoritmos não criaram as câmaras de eco; elas existiam muito antes da internet, inscritas nas tendências naturais de associação humana. O que os sistemas de personalização fizeram foi criar as condições estruturais para sua proliferação em escala sem precedentes.

A evidência empírica mais influente sobre esse fenômeno vem dos experimentos de Cass Sunstein, que demonstrou que grupos compostos por indivíduos com posições políticas semelhantes, quando deliberam entre si sem exposição a perspectivas externas, tendem a adotar posições mais extremas ao final da discussão do que sustentavam ao início. Sunstein denominou esse mecanismo polarização em grupo, e sua lógica é relativamente simples: em ambientes homogêneos, os argumentos disponíveis apontam sistematicamente na mesma direção, os sinais de aprovação social reforçam as posições dominantes e a ausência de contraposição elimina os incentivos à moderação.

Um ponto crucial para a análise doutoral: a pesquisa empírica sobre a relação entre algoritmos e polarização produziu resultados mais nuançados do que o senso comum e o debate público sugerem. Levi Boxell, Matthew Gentzkow e Jesse Shapiro demonstraram, em artigo na *PNAS* (2017), que a polarização política nos Estados Unidos cresceu mais rapidamente entre grupos demográficos com *menor* presença nas redes sociais — notadamente americanos acima de 65 anos. Se o algoritmo fosse o principal motor da polarização, o padrão seria invertido.

Isso não significa que algoritmos não contribuam para a polarização — a evidência de que contribuem é substancial. Significa que a polarização é um fenômeno multidimensional com causas múltiplas: segregação residencial e geográfica, fragmentação da mídia tradicional (especialmente a televisão a cabo nos EUA), declínio de instituições integradoras (partidos de massa, sindicatos, igrejas de bairro), além das redes sociais. Atribuir a polarização exclusivamente aos algoritmos é, simultaneamente, exagerar o papel das plataformas e subestimar as causas estruturais mais profundas.

Tabela 07. Câmara de Eco versus Bolha de Filtro

Dimensão	Câmara de Eco	Bolha de Filtro
Origem	Escolha ativa do usuário: segue, bloqueia, ignora, seleciona	Seleção passiva pelo algoritmo da plataforma
Percepção pelo usuário	O usuário pode perceber que está numa câmara (escolheu estar)	O usuário raramente percebe a filtragem algorítmica

Principal tecnologia	Ferramentas de seguir/bloquear/silenciar/criar grupos	Algoritmos de recomendação, feeds personalizados
Efeito político primário	Radicalização progressiva, groupthink, hostilidade ao diferente	Desinformação passiva, distorção da percepção do outro
Responsabilidade	Compartilhada: usuário e design da plataforma	Primariamente da plataforma, o usuário não controla o algoritmo
Intervenção possível	Alfabetização midiática, regulação de design	Transparência algorítmica, auditoria pública, diversidade de conteúdo

Fonte: Elaborado pelo autor com base em Pariser (2011), Sunstein (2017)

Aplicado ao contexto das redes sociais, esse princípio adquire implicações de longo alcance: os feeds personalizados não se limitam a refletir as posições políticas dos usuários; eles as radicalizam progressivamente, ao reproduzir, de forma ampliada e contínua, as condições experimentais descritas por Sunstein.

O Funil da Radicalização

A dimensão mais perturbadora desse processo não é a homofilia passiva, mas a radicalização ativa induzida pelos próprios mecanismos de recomendação. Zeynep Tufekci descreveu esse fenômeno com precisão ao analisar o funcionamento do algoritmo do YouTube, identificando o que denominou "funil do YouTube": usuários que consomem conteúdo político moderado são sistematicamente recomendados conteúdo progressivamente mais extremado, porque o algoritmo de recomendação aprende que conteúdo de maior intensidade emocional gera maior engajamento.

A lógica subjacente é estritamente econômica: o sistema não foi projetado para radicalizar usuários, mas para maximizar o tempo de permanência na plataforma. A radicalização é, nesse sentido, um efeito colateral estrutural de uma arquitetura otimizada para o engajamento.

O resultado é uma espiral que pode conduzir um usuário de conteúdo político convencional a material francamente extremista em poucas semanas de uso intensivo. Cada etapa da trajetória parece individualmente pequena e razoável; o salto só se torna visível quando se observa o ponto de partida e o ponto de chegada. Esse caráter gradual e imperceptível é precisamente o que torna o fenômeno tão difícil de ser detectado e interrompido pelo próprio usuário.

Implicações Democráticas

As implicações políticas desse processo são de primeira ordem. Câmaras de eco e radicalização algorítmica não operam apenas sobre indivíduos isolados; elas reestruturam coletivamente o campo de posições disponíveis no debate público, deslocando o centro de gravidade do espectro político em direção aos extremos. Quando parcelas significativas da população são expostas a esse mecanismo de forma sistemática e prolongada, os efeitos sobre a qualidade da deliberação democrática tornam-se estruturais: diminui a tolerância ao dissenso, estreita-se o vocabulário político compartilhado e aumenta a distância afetiva entre grupos com visões de mundo distintas.

Nesse sentido, o problema não é apenas que as câmaras de eco distorcem a percepção individual da realidade política. É que, ao operar simultaneamente sobre milhões de usuários, elas produzem uma reconfiguração coletiva do espaço público, na qual a moderação passa a ser percebida como fraqueza, o compromisso como traição e o adversário como inimigo. A democracia, enquanto sistema que pressupõe a capacidade de conviver com a diferença e de construir acordos provisórios sobre bases contestadas, é diretamente afetada por essa dinâmica.

> **FIQUE LIGADO (A)!**
> Para o profissional de campanha: o mesmo mecanismo de radicalização algorítmica que aprofunda a polarização política pode ser estrategicamente utilizado para mobilizar a base eleitoral do candidato. Conteúdo de alta intensidade emocional, indignação com adversários, orgulho de conquistas, ameaças à identidade do grupo, tende a gerar maior engajamento orgânico e, consequentemente, maior alcance gratuito nas plataformas. Isso, porém, tem custos: aprofunda a polarização e pode provocar reações legais se o conteúdo cruzar a linha da desinformação ou do discurso de ódio.

ANÁLISE DE SENTIMENTO: FERRAMENTA ESTRATÉGICA E INSTRUMENTO DE FISCALIZAÇÃO

A análise de sentimento é o ramo do Processamento de Linguagem Natural dedicado à identificação e classificação de opiniões, emoções e atitudes em textos. No contexto eleitoral, constitui ferramenta de inteligência estratégica para campanhas, e também instrumento de monitoramento para a Justiça Eleitoral e para pesquisadores.

Tabela 08. Abordagens Técnicas de Análise de Sentimento

Abordagem	Como funciona	Vantagens	Limitações	Aplicação Eleitoral
Análise lexical	Utiliza dicionários de palavras classificadas por valência emocional.	Rápida e computacionalmente barata.	Ignora contexto, ironia, sarcasmo e negação.	Monitoramento em tempo real de volume de menções.
Machine Learning supervisionado	Modelos treinados em corpus anotado por humanos.	Alta precisão em domínios específicos.	Requer corpus grande; generaliza mal para novos domínios.	Classificação de conteúdo eleitoral por tema e tom.
Modelos de linguagem (BERT, GPT)	Modelos pré-treinados ajustados para a tarefa. Capturam contexto semântico profundo.	Estado da arte; detecta ironia e negação.	Custo computacional alto; modelos em português menos desenvolvidos.	Detecção automatizada de desinformação e discurso de ódio eleitoral.
Análise de aspecto	Identifica sentimento por dimensão temática específica no mesmo texto.	Muito útil para discurso eleitoral multidimensional.	Complexidade técnica maior.	Avaliar posição do eleitorado sobre temas específicos da campanha.

Fonte: Elaborado pelo autor com base em literatura de NLP.

Bots, Cyborgs e Automação de Conteúdo Político

A automação de comportamento em redes sociais constitui um espectro de fenômenos que vai de contas totalmente automatizadas (bots) até perfis humanos assistidos por software (cyborgs). Compreender as distinções nesse espectro é essencial tanto para a análise acadêmica quanto para a aplicação do direito eleitoral porque cada categoria apresenta características técnicas, detectabilidade e responsabilidade jurídica distintas.

A análise normativa das práticas de microtargeting, embora essencial para delimitar fronteiras jurídicas e éticas, não esgota a compreensão dos mecanismos concretos por meio dos quais a influência digital é operacionalizada no ambiente eleitoral. Em particular, a dinâmica de amplificação e disseminação de conteúdo político em larga escala depende, de forma crescente, de infraestruturas sociotécnicas baseadas em automação, coordenação e simulação de comportamento humano.

Nesse contexto, torna-se necessário examinar os diferentes tipos de agentes

digitais e híbridos que atuam na circulação de informações políticas, variando em grau de automação, sofisticação tecnológica e capacidade de evasão de mecanismos de detecção. Esses agentes não apenas ampliam o alcance das mensagens, mas também desempenham papel central na construção artificial de relevância, consenso e legitimidade no espaço público digital.

O quadro a seguir sistematiza uma tipologia desses atores, distinguindo-os segundo sua definição operacional, capacidade de atuação, grau de detectabilidade e os principais riscos que representam para a integridade do processo eleitoral.

Tabela 09. Tipologia de Agentes Automatizados e Híbridos na Comunicação Eleitoral Digital

Tipo	Definição	Capacidade operacional	Detectabilidade	Risco eleitoral principal
Bot simples	Conta totalmente automatizada, operada por scripts sem intervenção humana	Alto volume de postagem, padrões repetitivos, baixa adaptabilidade	Alta, devido a padrões anômalos e comportamento mecânico	Inflação artificial de trending topics e falsa percepção de popularidade
Bot sofisticado (ML-based)	Automação baseada em modelos de linguagem com capacidade de adaptação	Alto volume com variação linguística, respostas contextuais, simulação de comportamento humano	Média a baixa, dependendo da sofisticação do modelo	Disseminação convincente de desinformação e simulação de consenso social
Cyborg	Conta humana que utiliza ferramentas de automação para amplificação	Combinação de autenticidade humana com eficiência automatizada	Baixa, pois aparenta comportamento humano	Astroturfing credível, infiltração em grupos e coordenação inautêntica
Sockpuppet	Identidade falsa operada manualmente por um indivíduo real	Alta adaptabilidade, baixo volume, atuação estratégica e personalizada	Muito baixa, pois se comporta como usuário legítimo	Manipulação de opinião, criação de dissidência artificial e influência sobre lideranças

Troll farm	Operação coordenada de múltiplos operadores humanos com perfis falsos	Escala industrial com qualidade humana, atuação organizada e persistente	Muito baixa, especialmente quando profissionalizada	Campanhas de influência coordenada, polarização e construção de narrativas falsas

Fonte: Elaborado pelo autor, com base em Samuel Woolley e Philip N. Howard (2016), Oxford Internet Institute; Yochai Benkler, Robert Faris e Hal Roberts (2018); NetLab UFRJ; e literatura contemporânea sobre desinformação digital, bots e campanhas coordenadas.

■ Astroturfing — A Grama Artificial da Política Digital

O termo astroturfing derivado de AstroTurf, marca de grama sintética americana descreve a criação de aparência de movimento político espontâneo (grassroots) que é, na realidade, coordenado e financiado por atores identificáveis. Em sua versão digital, o astroturfing utiliza redes de bots, cyborgs e sockpuppets para criar a ilusão de que determinada posição, candidato ou narrativa tem amplo apoio popular espontâneo quando o apoio é artificialmente construído. O efeito sobre eleitores indecisos é devastador: ao perceberem (falsamente) que 'todo mundo' apoia X, ativam o viés de prova social e tendem a seguir o (inexistente) consenso.

A Arquitetura da Desinformação: Como Fake News Vencem o Debate

A compreensão do papel dos algoritmos no debate eleitoral seria incompleta sem uma análise da dinâmica de propagação da desinformação. Trata-se de um fenômeno que, como demonstrado empiricamente, é estruturalmente favorecido pela lógica de engajamento das plataformas digitais: não porque as plataformas tenham sido projetadas para amplificar mentiras, mas porque a mentira, com frequência, é mais eficaz do que a verdade em despertar as respostas emocionais que o algoritmo aprende a recompensar.

Tipologia e Distinções Conceituais

O ponto de partida analítico mais rigoroso para o estudo da desinformação é a tipologia proposta por Wardle e Derakhshan, em trabalho de referência publicado para o Conselho da Europa. Os autores distinguem três categorias fundamentais: a *mis-information*, definida como conteúdo falso disseminado sem intenção deliberada de causar dano; a *dis-information*, caracterizada pela falsidade intencional e pelo propósito explícito de prejudicar; e a *mal-information*, que envolve conteúdo verdadeiro utilizado de forma estratégica para causar dano.

A aparente simplicidade dessa classificação esconde sua relevância analítica e jurídica: cada categoria implica um diagnóstico distinto sobre o papel dos atores envolvidos e exige, por consequência, respostas institucionais diferenciadas.

No plano do direito eleitoral brasileiro, essa distinção adquire implicações concretas. A disseminação de *dis-information* com fins eleitorais pode configurar, dependendo das circunstâncias fáticas, o crime de calúnia eleitoral, previsto no artigo 324 do Código Eleitoral, ou o uso indevido dos meios de comunicação. A *mis-information*, por sua vez, por carecer do elemento volitivo, situa-se em um terreno juridicamente mais ambíguo, ainda que seus efeitos sobre o debate público possam ser igualmente deletérios. Já a *mal-information* coloca desafios ainda mais complexos, pois opera no limite entre o exercício legítimo da liberdade de expressão e a manipulação deliberada do contexto em que informações verdadeiras são apresentadas.

A Anatomia de uma Campanha de Desinformação

Compreender a desinformação eleitoral exige ir além da análise do conteúdo falso em si e examinar a infraestrutura que o produz, amplifica e dissemina. Uma campanha sofisticada de desinformação opera, tipicamente, em quatro etapas articuladas.

A primeira é a produção de conteúdo falso ou deliberadamente distorcido, realizada por atores que podem ser identificados ou agir sob anonimato. A segunda é a amplificação artificial do conteúdo por redes de perfis automatizados, os chamados *bots*, e por contas coordenadas que simulam engajamento orgânico, criando a ilusão de que determinado conteúdo já circula amplamente e desperta interesse genuíno. A terceira etapa envolve a cobertura por veículos de mídia de baixa credibilidade editorial, que conferem ao conteúdo uma aparência de legitimidade jornalística suficiente para superar o ceticismo de parcela dos usuários. A quarta e última etapa é a disseminação orgânica por usuários reais que acreditam genuinamente no conteúdo que compartilham, tornando-se, sem o saber, agentes involuntários da campanha.

Cada uma dessas etapas tem lógica própria e pode, em tese, ser interrompida de forma independente. Na prática, contudo, elas se retroalimentam com tal velocidade que, quando a desinformação atinge a fase de disseminação orgânica, qualquer intervenção corretiva já enfrenta o problema estrutural de que as desmentidas raramente alcançam a mesma escala e velocidade do conteúdo original. Para melhor entendimento, confira a tabela a seguir:

Tabela 10. Anatomia de uma Campanha de Desinformação Eleitoral Sofisticada

Fase	Atividades	Tecnologias de IA Utilizadas	Ator Responsável	Ponto de Intervenção Regulatória
Planejamento	Definição de objetivos; segmentação do eleitorado-alvo; identificação de vulnerabilidades psicológicas e informacionais.	Análise preditiva; modelagem psicográfica; monitoramento de redes sociais.	Estrategistas políticos; empresas de marketing digital.	Exigência de declaração de prestadores de serviço de IA na prestação de contas.
Produção	Criação de conteúdo falso ou distorcido: textos, imagens, áudios, vídeos.	IA generativa (LLMs, Midjourney, ElevenLabs, Sora).	Fábricas de conteúdo; freelancers especializados.	Vedação de deepfakes; obrigação de rotulagem (Res. TSE 23.732/2024).
Amplificação inicial	Distribuição artificial para criar aparência de popularidade e acionar algoritmos das plataformas.	Redes de bots; sistemas de disparo em massa; automação de redes sociais.	Empresas de automação; operadores de bot farms.	Proibição de disparo em massa (art. 222-A, Lei 9.504/1997).
Legitimação	Republicação por veículos de baixa credibilidade para conferir aparência jornalística.	Geração automatizada de sites de notícias falsas; SEO malicioso.	Sites de notícias falsas; blogs partidários não declarados.	Obrigação de identificação de veículos pagos; prestação de contas de impulsionamento.
Disseminação orgânica	Distribuição por eleitores reais de boa-fé que acreditam no conteúdo.	Plataformas de redes sociais; WhatsApp; Telegram.	Eleitores comuns sem consciência de participar de operação coordenada.	Literacia midiática; fact-checking; programas de contranarratIva.

Fonte: Elaborado pelo autor com base em Howard (2020), Benkler, Faris e Roberts (2018) e documentação do TSE.

O Papel da Inteligência Artificial Generativa

A introdução da inteligência artificial generativa representa uma ruptura qualitativa nesse processo. A produção de conteúdo desinformativo sofisticado, que anteriormente exigia habilidades específicas de redação, edição de imagem ou produção audiovisual,

tornou-se acessível a qualquer usuário com acesso às ferramentas disponíveis. Textos convincentes, imagens sintéticas, áudios clonados e vídeos manipulados podem hoje ser gerados em minutos, a custo praticamente nulo, por atores sem qualquer formação técnica especializada.

Essa democratização das ferramentas de produção de desinformação tem consequências que vão além do aumento quantitativo do volume de conteúdo falso em circulação. Ela altera a própria epistemologia do debate público: quando qualquer imagem pode ser fabricada, qualquer áudio pode ser clonado e qualquer declaração pode ser atribuída a qualquer pessoa com aparência de autenticidade, o ônus da prova sobre o que é real se desloca de forma radical.

O resultado não é apenas a proliferação da mentira, mas a erosão da confiança na evidência, o que representa, do ponto de vista democrático, um dano ainda mais profundo e duradouro do que qualquer campanha de desinformação específica.

FAKE ENGAGEMENT E A MANIPULAÇÃO DOS ALGORITMOS

O fake engagement, a inflação artificial de métricas de engajamento em plataformas digitais, é um problema que opera em dois níveis simultâneos: distorce a percepção pública de popularidade de candidatos e posições políticas, e distorce os próprios algoritmos das plataformas, que interpretam engajamento artificial como sinal de relevância e distribuem o conteúdo organicamente para audiências reais.

Tabela 11. Formas de Fake Engagement Eleitoral

Prática	Mecanismo	Custo Estimado (2024)	Risco Jurídico
Compra de seguidores e curtidas	Serviços que vendem seguidores falsos (bots) ou curtidas artificiais.	US$ 10-50 por mil seguidores no Instagram	Infração às regras das plataformas; abuso de poder econômico se financiado irregularmente
Grupos de engajamento coordenado	Redes privadas de usuários reais que se comprometem a curtir e compartilhar mutuamente.	Sem custo financeiro direto	Comportamento inautêntico coordenado, vedado pelas plataformas
Inflação de trending topics	Coordenação de publicações simultâneas com hashtags para	Baixo (coordenação	Propaganda irregular; abuso de poder econômico

	simular tendências orgânicas.	manual ou automatizada)	
Fake views em vídeo	Geração de visualizações artificiais para sinalizar relevância ao algoritmo.	US$ 5-20 por mil views no YouTube	Violação dos termos de serviço das plataformas; propaganda irregular
Comentários automatizados	Geração em massa de comentários genéricos por bots para simular engajamento qualitativo.	Muito baixo (automação por LLMs)	Conteúdo gerado por IA sem identificação viola Res. TSE 23.732/2024

Fonte: Elaborado pelo autor com base em Apostila Curso IA e Eleições módulo 2 (2026) e literatura sobre desinformação digital.

⚡ ATENÇÃO PRÁTICA — O Efeito de Segunda Ordem — Quando o Algoritmo Amplifica o Falso

Este é o ponto mais crítico do problema do fake engagement: as plataformas digitais utilizam métricas de engajamento como sinal de qualidade e relevância para seus algoritmos de distribuição. Um post com 10.000 curtidas artificiais recebe distribuição orgânica para audiências reais muito maior que um post com 100 curtidas genuínas. Isso significa que um investimento relativamente pequeno em fake engagement pode resultar em exposição orgânica massiva: o algoritmo da plataforma, enganado pelas métricas falsas, subsidia involuntariamente a desinformação com distribuição gratuita.

Castells e a Comunicação de Massa Individualizada

A análise do impacto dos algoritmos sobre o debate político encontra na obra de Manuel Castells um enquadramento teórico de elevado poder explicativo. Em sua trilogia sobre a sociedade em rede e, de forma mais sistemática, em *Comunicação e Poder*, o autor identifica a emergência de um novo paradigma comunicacional que denomina autocomunicação de massa. Trata-se de uma forma de comunicação que combina, simultaneamente, características de alcance massivo com níveis inéditos de individualização na produção, distribuição e recepção das mensagens.

Diferentemente dos modelos tradicionais de comunicação de massa, estruturados em torno de emissores institucionais e audiências relativamente passivas, a autocomunicação de massa é marcada pela descentralização dos polos emissores e pela ampliação das capacidades expressivas dos indivíduos. Nesse modelo, qualquer usuário conectado à rede pode produzir conteúdo, distribuí-lo em escala potencialmente global e direcioná-lo a públicos específicos, sem depender de mediação institucional.

Nesse sentido, Manuel Castells afirma que:

> A autocomunicação de massa expande enormemente o alcance dos meios de comunicação. Ela inclui a produção de mensagens de muitos para muitos, o que permite que grupos e indivíduos contornem os meios de comunicação corporativos para criar suas próprias redes de comunicação horizontal (CASTELLS, 2009, p. 89).

Essa formulação é central para compreender a reconfiguração contemporânea da esfera pública. Ao mesmo tempo em que rompe com a lógica vertical dos meios tradicionais, a autocomunicação de massa introduz uma nova arquitetura comunicacional baseada em redes distribuídas, nas quais a distinção entre emissor e receptor torna-se progressivamente difusa.

Essa característica apresenta uma ambivalência normativa que é particularmente relevante para a análise democrática. De um lado, configura uma promessa de ampliação da participação política, ao permitir que grupos historicamente excluídos dos meios de comunicação tradicionais produzam e disseminem suas próprias narrativas. A redução das barreiras de entrada no sistema comunicacional amplia o pluralismo de vozes e potencializa formas horizontais de mobilização social.

De outro lado, essa mesma estrutura implica a ausência de mecanismos tradicionais de filtragem editorial, o que altera profundamente os critérios de validação da informação. A descentralização da produção comunicacional elimina intermediários responsáveis pela verificação, contextualização e hierarquização do conteúdo, criando um ambiente no qual diferentes tipos de informação, independentemente de sua veracidade ou qualidade, competem em condições semelhantes de visibilidade.

A introdução de sistemas algorítmicos nesse ambiente intensifica essa ambivalência. Se, por um lado, a autocomunicação de massa amplia a capacidade de produção de conteúdo, por outro, os algoritmos passam a desempenhar o papel de curadores invisíveis daquilo que efetivamente circula e ganha relevância. A combinação entre produção descentralizada e distribuição algorítmica cria um sistema híbrido, no qual a liberdade de emissão é elevada, mas a visibilidade permanece condicionada por lógicas opacas de priorização.

No contexto eleitoral brasileiro, essa dinâmica encontra expressão particularmente evidente no WhatsApp. A plataforma materializa de forma exemplar a lógica da autocomunicação de massa ao possibilitar que qualquer indivíduo produza e distribua conteúdo político em redes privadas, com alcance potencialmente exponencial e custo

marginal próximo de zero. Ao mesmo tempo, sua arquitetura baseada em comunicação criptografada e em redes fechadas dificulta a atuação de mecanismos institucionais de controle e fiscalização.

Essa combinação entre descentralização comunicacional e opacidade estrutural redefine os termos do debate democrático. A comunicação política deixa de ser mediada predominantemente por instituições públicas e privadas identificáveis e passa a ocorrer em redes distribuídas, nas quais a origem, a circulação e o impacto das mensagens são frequentemente difíceis de rastrear. O resultado é a emergência de um ambiente comunicacional no qual a ampliação da liberdade de expressão coexiste com o aumento dos riscos associados à desinformação, à manipulação e à fragmentação da esfera pública.

Nesse sentido, a contribuição de Manuel Castells não apenas descreve uma transformação tecnológica, mas oferece uma chave interpretativa fundamental para compreender as tensões contemporâneas entre comunicação digital e democracia. A autocomunicação de massa, ao mesmo tempo em que expande as possibilidades de participação, reconfigura os mecanismos de formação da opinião pública, colocando novos desafios para a teoria e para a prática democrática.

Implicações para o Marco Regulatório: O que os algoritmos devem ao eleitorado

A análise desenvolvida neste capítulo projeta consequências diretas sobre o debate contemporâneo acerca da regulação de sistemas algorítmicos no contexto eleitoral. Se os algoritmos de recomendação operados por plataformas digitais produzem efeitos estruturais sobre o ambiente informacional, amplificando conteúdos desinformativos, intensificando dinâmicas de polarização e fragmentando a esfera pública em comunidades informacionais isoladas, então não se trata apenas de um problema tecnológico ou comunicacional, mas de uma questão normativa central para a teoria democrática.

A formulação do problema regulatório pode ser expressa nos seguintes termos: em que medida sistemas privados, orientados por objetivos econômicos, podem exercer influência significativa sobre a formação da opinião pública sem estarem sujeitos a obrigações compatíveis com os princípios da democracia? Em outras palavras, que deveres podem ser legitimamente impostos às plataformas digitais enquanto infraestruturas centrais da comunicação política contemporânea?

Uma primeira resposta emerge no eixo da transparência. A assimetria informacional entre plataformas e sociedade constitui um dos principais obstáculos à compreensão e ao controle dos efeitos dos algoritmos. Os sistemas de recomendação operam segundo

lógicas proprietárias, frequentemente protegidas por sigilo comercial, o que limita o escrutínio público e dificulta a produção de evidência empírica independente. Nesse contexto, a imposição de deveres de transparência não se configura como mera exigência administrativa, mas como condição necessária para a própria inteligibilidade do fenômeno regulado.

A Resolução TSE nº 23.732/2024, editada pelo Tribunal Superior Eleitoral, representa um avanço relevante nesse sentido ao estabelecer obrigações específicas para plataformas digitais durante o período eleitoral. Entre essas obrigações, destaca-se a exigência de disponibilização de dados sobre o funcionamento dos sistemas de recomendação a pesquisadores independentes credenciados, criando um mecanismo institucional de mediação entre o sigilo empresarial e o interesse público na fiscalização do processo eleitoral.

Essa abordagem reconhece que a regulação eficaz de sistemas algorítmicos depende da produção de conhecimento técnico qualificado, que não pode ser monopolizado pelas próprias plataformas. Ao abrir espaço para auditorias externas e para a análise independente, o modelo brasileiro busca reduzir a opacidade estrutural desses sistemas, ainda que de forma limitada e condicionada.

No plano internacional, o Digital Services Act, em vigor na União Europeia desde 2024, amplia significativamente o escopo dessas obrigações ao instituir um regime regulatório mais robusto para as chamadas Very Large Online Platforms. O diploma europeu introduz mecanismos de auditoria algorítmica obrigatória, avaliação de riscos sistêmicos e deveres de mitigação de impactos negativos sobre direitos fundamentais e processos democráticos. Além disso, estabelece restrições ao uso de dados sensíveis, incluindo dados relacionados a opiniões políticas, para fins de recomendação e publicidade direcionada.

A comparação entre os modelos evidencia diferentes estratégias regulatórias, mas também uma convergência quanto ao diagnóstico. A atuação dos algoritmos no ambiente informacional contemporâneo não pode ser tratada como atividade neutra ou puramente técnica. Trata-se de uma função com efeitos públicos relevantes, que demanda formas proporcionais de responsabilização.

Entretanto, a construção desse marco regulatório enfrenta tensões normativas significativas. A imposição de obrigações às plataformas deve ser calibrada de modo a evitar restrições indevidas à liberdade de expressão, especialmente em contextos nos quais a definição do que constitui conteúdo problemático pode ser controversa. Há, portanto, o

risco de que medidas destinadas a proteger a integridade do debate democrático produzam efeitos colaterais de limitação do pluralismo, caso não sejam cuidadosamente desenhadas.

Além disso, a própria natureza dinâmica e adaptativa dos sistemas algorítmicos impõe desafios adicionais à regulação. Diferentemente de meios tradicionais, cujos conteúdos podem ser analisados de forma relativamente estática, os algoritmos operam em tempo real, ajustando continuamente suas estratégias com base em dados comportamentais. Isso exige modelos regulatórios capazes de lidar não apenas com conteúdos específicos, mas com processos e estruturas de decisão automatizada.

Diante desse cenário, emerge a necessidade de uma abordagem regulatória baseada em princípios, mais do que em regras estritamente prescritivas. Entre esses princípios, destacam-se a transparência, a accountability, a não discriminação, a proteção de dados pessoais e a integridade do processo eleitoral. Tais diretrizes podem orientar a construção de mecanismos institucionais flexíveis, capazes de se adaptar à evolução tecnológica sem perder de vista os valores fundamentais da democracia.

Em última instância, a questão que orienta este debate pode ser formulada em termos mais amplos: se os algoritmos passaram a desempenhar papel estruturante na formação da vontade política, então devem ser submetidos a um regime de responsabilidades compatível com essa função. O desafio não consiste em limitar a tecnologia em si, mas em garantir que sua utilização ocorra dentro de parâmetros que preservem as condições de autonomia, igualdade e liberdade que fundamentam o processo democrático.

⚖ JURISPRUDÊNCIA TSE — TSE — Resolução nº 23.732/2024, art. 27

EMENTA:

As plataformas de redes sociais que operam no Brasil ficam obrigadas a disponibilizar, mediante requisição fundamentada do Tribunal Superior Eleitoral ou dos Tribunais Regionais Eleitorais, informações sobre o funcionamento de seus sistemas de recomendação de conteúdo em período eleitoral, incluindo os critérios de priorização e os mecanismos de moderação de conteúdo eleitoral.

RESULTADO:

As plataformas que descumprirem a obrigação no prazo estabelecido ficam sujeitas à multa diária de R$ 150.000,00 (cento e cinquenta mil reais), sem prejuízo da responsabilização solidária pelo conteúdo irregular mantido após notificação.

■ RESUMO DO CAPÍTULO:

Este capítulo demonstrou que os algoritmos de recomendação das plataformas digitais não constituem instrumentos neutros de distribuição de informação, mas sistemas econômicos orientados pela

maximização do engajamento que produzem efeitos políticos sistemáticos. Entre esses efeitos, destacam-se a amplificação de conteúdo emocionalmente provocativo, a formação de bolhas de filtro, a consolidação de câmaras de eco e o favorecimento estrutural da desinformação em detrimento da informação factualmente verificável. Do ponto de vista teórico, a articulação entre as contribuições de Jürgen Habermas, Eli Pariser, Cass Sunstein e Manuel Castells permite construir um enquadramento analítico robusto para compreender as transformações contemporâneas da esfera pública. Em conjunto, esses autores evidenciam que a mediação algorítmica da comunicação não apenas altera a circulação da informação, mas compromete, em nível estrutural, as condições necessárias para a deliberação democrática, ao fragmentar o espaço público, intensificar a polarização e deslocar a centralidade da argumentação racional para dinâmicas de engajamento emocional. No plano empírico e estratégico, o capítulo evidenciou que os mesmos mecanismos algorítmicos que degradam a qualidade do debate público podem ser mobilizados de forma deliberada por campanhas eleitorais. A capacidade de segmentação, personalização e amplificação seletiva de mensagens transforma os algoritmos em instrumentos centrais de disputa política, com impactos diretos sobre a formação da opinião pública e sobre os resultados eleitorais. Essa constatação introduz uma tensão normativa relevante. O uso estratégico de sistemas algorítmicos pode situar-se em uma zona intermediária entre práticas legítimas de comunicação política e formas sofisticadas de manipulação eleitoral. A delimitação dessa fronteira, tanto do ponto de vista jurídico quanto ético, constitui um dos desafios centrais enfrentados pelo direito eleitoral contemporâneo. As próximas partes desta obra aprofundam essa problemática ao examinar, de forma sistemática, os mecanismos operacionais da campanha algorítmica, os atores envolvidos, os limites impostos pelo ordenamento jurídico e as possíveis estratégias institucionais para a preservação da integridade do processo democrático em um ambiente informacional mediado por tecnologia.

QUESTÕES PARA REFLEXÃO E DEBATE

1. Em que medida os algoritmos de recomendação comprometem as condições de deliberação racional previstas por Jürgen Habermas, e é possível adaptar sua teoria da esfera pública ao ambiente digital contemporâneo?
2. A bolha de filtro, conforme formulada por Eli Pariser, deve ser compreendida como um produto da arquitetura algorítmica ou das escolhas individuais dos usuários? Quais são as implicações dessa distinção para a regulação?
3. A maximização do engajamento pelas plataformas digitais é compatível com a integridade do processo democrático, ou produz uma distorção estrutural inevitável do debate público?
4. É possível estabelecer critérios jurídicos claros para distinguir o uso legítimo de estratégias algorítmicas em campanhas eleitorais de práticas configuradoras de manipulação ilícita?
5. Quais deveres normativos devem ser atribuídos às plataformas digitais diante de seu papel na formação da opinião pública, especialmente à luz da Resolução TSE nº 23.732/2024 e do Digital Services Act?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. O algoritmo não foi programado para prejudicar a democracia. Mas prejudica.

As plataformas digitais não foram projetadas com objetivos políticos. Foram projetadas para maximizar o tempo que você passa nelas, porque tempo de tela equivale a receita publicitária. O problema é que o conteúdo que mais prende a atenção tende a ser o que desperta emoções fortes: indignação, medo,

surpresa, escândalo. Conteúdo equilibrado e factualmente preciso raramente compete nesse ambiente. O resultado é um ambiente informacional estruturalmente enviesado para o extremo, não por intenção, mas por design econômico.

2. Notícia falsa viaja mais rápido do que notícia verdadeira. Isso não é opinião, é dado.

O estudo de Vosoughi, Roy e Aral, publicado na revista Science em 2018, analisou mais de 126 mil histórias no Twitter e concluiu que notícias falsas se propagam significativamente mais rápido e alcançam muito mais pessoas do que notícias verdadeiras. A explicação é simples: conteúdo falso tende a ser mais novo, mais surpreendente e mais emocionalmente carregado, características que aumentam o engajamento e, portanto, a priorização algorítmica. Para profissionais de campanha, isso significa que desinformação sobre seu candidato, uma vez lançada, tem vantagem estrutural de velocidade que a desmentida raramente consegue reverter.

3. Bolha de filtro e câmara de eco não são a mesma coisa. A distinção importa na prática.

A bolha de filtro é o que o algoritmo faz: seleciona conteúdos com base no seu histórico e reduz sua exposição ao que diverge. A câmara de eco é o que você faz: escolhe seguir, curtir e interagir com quem pensa como você. Os dois se reforçam mutuamente. Para campanhas, isso significa que eleitores já engajados tendem a ficar mais radicalizados ao longo do tempo, enquanto eleitores indecisos estão num ambiente informacional mais diverso e, portanto, mais aberto à persuasão. Concentrar recursos nos indecisos e não apenas na base é uma implicação direta dessa distinção.

4. O funil do YouTube é uma estratégia que pode ser usada contra você.

O mecanismo descrito por Zeynep Tufekci, em que usuários são progressivamente recomendados conteúdo mais extremo, funciona nos dois sentidos. Adversários podem produzir conteúdo moderado sobre seu candidato como porta de entrada, e deixar o algoritmo fazer o trabalho de conduzir o usuário a versões progressivamente mais negativas. Monitorar o que circula sobre sua candidatura nas plataformas de vídeo, especialmente o que aparece nas recomendações automáticas, é uma forma de identificar esse tipo de estratégia antes que ela se consolide.

5. Mis-information, dis-information e mal-information têm tratamentos jurídicos diferentes.

Essa distinção conceitual de Wardle e Derakhshan tem consequências diretas no direito eleitoral brasileiro. Conteúdo falso produzido com intenção de prejudicar um candidato (dis-information) pode configurar calúnia eleitoral, prevista no artigo 324 do Código Eleitoral. Conteúdo verdadeiro usado fora de contexto para causar dano (mal-information) situa-se no limite entre liberdade de expressão e manipulação, um terreno ainda sendo construído pela jurisprudência do TSE. Equipes jurídicas de campanha precisam entender essa tipologia para classificar corretamente o que monitoram e escolher a via processual adequada.

6. Uma campanha de desinformação tem quatro etapas. Interromper qualquer uma delas muda o resultado.

Produção, amplificação artificial por bots, cobertura por veículos de baixa credibilidade e disseminação orgânica por usuários reais: cada etapa tem uma lógica própria e pode ser interrompida. Denúncias ao TSE e às plataformas são mais eficazes nas primeiras etapas, quando o conteúdo ainda não atingiu escala. Quando a desinformação já está em fase de disseminação orgânica, a desmentida enfrenta o problema de que os usuários que a compartilham acreditam genuinamente nela e tendem a resistir à correção. Agilidade de resposta, portanto, não é apenas eficiência operacional: é determinante para a efetividade da ação.

7. As plataformas têm obrigações legais de transparência durante as eleições. Use isso.

A Resolução TSE nº 23.732/2024 obriga as plataformas a disponibilizar informações sobre seus sistemas de recomendação mediante requisição da Justiça Eleitoral, com multa diária de R$ 150 mil pelo descumprimento. Isso significa que, diante de indícios de manipulação algorítmica ou amplificação artificial de conteúdo adversário, há um mecanismo institucional disponível para exigir transparência. Campanhas que monitoram o ambiente digital e identificam padrões suspeitos de comportamento coordenado têm base para acionar esse mecanismo, seja diretamente ou por meio do Ministério Público Eleitoral.

CAPÍTULO 3

O Brasil Está Preparado para Isso?

Capacidade Institucional, Marco Regulatório e Vulnerabilidades Estruturais

"A democracia não perece de
um golpe só. Ela morre de mil
cortes pequenos, cada um,
individualmente, tolerável;
juntos, letais."

Levitsky & Ziblatt, 2018

RESILIÊNCIA DEMOCRÁTICA E INTEGRIDADE ELEITORAL

O conceito de integridade eleitoral, sistematizado por Pippa Norris, Richard Frank e Ferrán Martínez i Coma no âmbito do Electoral Integrity Project, refere-se ao conjunto de condições institucionais, normativas e procedimentais que asseguram que os processos eleitorais expressem, de forma autêntica, as preferências do eleitorado. Em sua formulação mais abrangente, a integridade eleitoral transcende a dimensão estritamente procedimental da apuração dos votos e incorpora elementos substantivos relacionados à qualidade do ambiente democrático em que essas preferências são formadas.

Essa ampliação conceitual não é trivial: ela implica reconhecer que uma eleição pode ser formalmente impecável e, ainda assim, produzir resultados que não expressam genuinamente a vontade do eleitorado, caso as condições informacionais que antecederam o voto tenham sido sistematicamente distorcidas.

Essa perspectiva é particularmente relevante no contexto das democracias digitais. A integridade de uma eleição não pode ser reduzida à ausência de fraude no momento do voto ou à correção dos procedimentos formais de totalização. Ela depende, de maneira crescente, das condições informacionais que antecedem o ato de votar: a qualidade, diversidade e confiabilidade das informações às quais os eleitores têm acesso, bem como a equidade na disputa comunicacional entre os diferentes atores políticos. O controle do fluxo informacional tornou-se, nas disputas contemporâneas, tão ou mais relevante do que os mecanismos tradicionais de mobilização política, o que obriga a teoria e o direito eleitoral a incorporarem dimensões que, até recentemente, permaneciam fora de seu escopo.

LINHA DO TEMPO: A EVOLUÇÃO DA IA NA POLÍTICA ELEITORAL (1952-2026)

A história da IA aplicada à política é mais longa do que o debate contemporâneo sugere. Ela não começou com o ChatGPT em 2022, nem com o Cambridge Analytica em 2016. Compreender essa trajetória é indispensável para entender por que os desafios atuais são diferentes em escala, velocidade e sofisticação de tudo que veio antes.

Tabela 12. Marcos Históricos da IA na Comunicação Política

Ano	Marco	Descrição	Implicação para o Brasil
1952	Primeiros computadores em análise eleitoral	CBS utilizou o UNIVAC I para prever eleição presidencial americana. Primeira demonstração do poder computacional na análise eleitoral.	Marco histórico: a tecnologia no processo eleitoral não é novidade do século XXI.
2004	Microtargeting moderno: Bush vs. Kerry	Campanha de Bush utilizou pela primeira vez em escala nacional modelos preditivos sofisticados para identificar eleitores indecisos persuasíveis.	Base do que se tornaria o microtargeting moderno adotado por campanhas brasileiras.
2008	Obama e a primeira campanha 2.0	Primeiro uso sistemático de dados de comportamento digital para segmentação eleitoral e mobilização via redes sociais.	Modelo adotado progressivamente por campanhas brasileiras a partir de 2010.
2012	Project Narwhal e Big Data	Campanha de reeleição de Obama integra dados de 180 milhões de americanos. 66.000 simulações computacionais por noite.	Estabelece o paradigma data-driven que influencia todas as campanhas presidenciais subsequentes, inclusive brasileiras.
2016	Cambridge Analytica, Trump, Brexit	Dados de 87 milhões de usuários do Facebook coletados sem consentimento. Primeiro uso documentado de perfis OCEAN para manipulação eleitoral em escala.	Impulsionou aprovação da LGPD no Brasil em 2018. Alerta regulatório para o TSE.
2018	Brasil: WhatsApp e disparo em massa	Primeiro caso documentado de uso do WhatsApp para disparo em massa de desinformação	Introduz novo vetor de manipulação: plataformas de mensageria criptografadas fora do alcance da fiscalização tradicional.

		eleitoral em escala industrial no Brasil.	
2022 2023	IA generativa: ChatGPT muda o jogo	Lançamento do ChatGPT (nov. 2022) democratiza o acesso a modelos de linguagem. Início da era da desinformação industrial automatizada.	Qualquer pessoa pode gerar conteúdo falso convincente em escala. Desafio regulatório urgente para as eleições de 2024 e 2026.
2024	Primeiro ciclo eleitoral global com IA generativa	Mais de 60 países realizaram eleições em 2024 com IA generativa amplamente disponível. Primeiros deepfakes eleitorais de alto impacto documentados.	Brasil enfrenta eleições municipais com marco regulatório novo (Res. TSE 23.732/2024). Ensaio geral para 2026.
2026	Eleições presidenciais e legislativas brasileiras	Primeiro ciclo eleitoral presidencial brasileiro sob plena vigência da Resolução TSE nº 23.732/2024 e do marco regulatório de IA.	Teste definitivo da robustez do sistema regulatório brasileiro diante dos desafios da IA eleitoral.

Fonte: Elaborado pelo autor com base em Oxford Internet Institute, MIT Election Lab, TSE.

O SISTEMA ELEITORAL BRASILEIRO: ESTRUTURA E COMPETÊNCIAS

O Brasil possui um dos sistemas eleitorais mais sofisticados e mais auditados do mundo. Quatro traços estruturais o distinguem da maioria das democracias: a Justiça Eleitoral, ramo especializado e independente do Judiciário; o financiamento público de campanhas com prestação de contas obrigatória; as urnas eletrônicas utilizadas em 100% das zonas eleitorais desde 2000; e a representação proporcional de lista aberta que cria o cenário de fragmentação partidária brasileiro.

Tabela 13. Estrutura Institucional da Justiça Eleitoral Brasileira

Instância	Composição	Competências Centrais	Relevância para IA Eleitoral
TSE — Tribunal Superior Eleitoral	7 ministros: 3 do STF, 2 do STJ e 2 advogados nomeados pelo Presidente com aprovação do Senado.	Organizar eleições nacionais; editar resoluções normativas; julgar recursos; registrar candidatos a Presidente.	Principal autoridade regulatória sobre IA eleitoral no Brasil. Editou a Resolução nº 23.732/2024.
TREs — Tribunais	Um TRE por estado (27 TREs total). Composição mista:	Organizar eleições estaduais e municipais; julgar recursos em	Primeiro nível de fiscalização de propaganda eleitoral irregular com IA em âmbito regional.

Regionais Eleitorais	magistrados do TJEJ e advogados.	segunda instância; aplicar sanções eleitorais.	
Juízes Eleitorais	Aprox. 2.800 zonas eleitorais. Juízes de carreira com designação temporária.	Alistamento eleitoral; fiscalização da propaganda eleitoral local; recebimento de denúncias.	Porta de entrada para denúncias de deepfakes e propaganda irregular com IA no nível local.
Ministério Público Eleitoral	Membros do MPF. Atua em todas as instâncias.	Propor AIJEs, representações por propaganda irregular e ações penais eleitorais.	Órgão fundamental de fiscalização e responsabilização por uso irregular de IA em campanhas.
Partidos Políticos	Registrados perante o TSE. Sujeitos à Lei 9.096/1995.	Apresentar candidatos; fiscalizar adversários; gerir fundo partidário e acesso ao HGPE.	Responsabilidade solidária por propaganda irregular veiculada em canais oficiais, incluindo conteúdo gerado por IA.

Fonte: Elaborado pelo autor com base no Código Eleitoral

A Urna Eletrônica: Segurança, Auditoria e a Narrativa da Fraude

A urna eletrônica brasileira é um dos temas mais malcompreendidos e mais instrumentalizados politicamente do sistema eleitoral, e essa combinação de incompreensão técnica e instrumentalização política não é acidental: ela é produto de uma estratégia documentada de erosão da confiança institucional que explora deliberadamente a opacidade técnica do sistema para um eleitorado sem formação em segurança computacional. Seus mecanismos de segurança e auditoria são objetos de intensa desinformação que seguem padrões identificáveis de propaganda política e exigem, portanto, tratamento rigoroso, baseado em evidências verificadas e suficientemente técnico para resistir ao escrutínio de quem genuinamente quer compreender o sistema, e não apenas questionar seus resultados.

O ponto de partida analítico mais importante para qualquer discussão sobre a urna eletrônica é a distinção entre crítica técnica legítima e narrativa de fraude instrumentalizada. Crítica técnica legítima é aquela que identifica vulnerabilidades específicas, descreve mecanismos concretos de exploração, propõe melhorias verificáveis e aceita as respostas técnicas quando elas são apresentadas. Ela existe, é saudável e o TSE a encoraja formalmente por meio do Teste Público de Segurança, no qual pesquisadores e entidades credenciadas têm acesso ao código-fonte e aos equipamentos para identificar e reportar

vulnerabilidades em ambiente controlado. A narrativa de fraude instrumentalizada, por sua vez, não tem interesse nas respostas técnicas porque seu objetivo não é melhorar o sistema: é minar a legitimidade do resultado eleitoral de forma preventiva, independentemente de qualquer evidência específica. Como documentaram Levitsky e Ziblatt (2018), essa estratégia de delegação preventiva de legitimidade é um dos mecanismos mais eficazes de erosão democrática precisamente porque é difícil de refutar: qualquer prova apresentada em favor da integridade do sistema pode ser descartada como parte da conspiração (LEVITSKY; ZIBLATT, 2018, p. 89-92).

O sistema de votação eletrônica brasileiro foi desenvolvido pelo TSE a partir de 1996, tornando-se obrigatório em todo o território nacional em 2000, e passou desde então por ciclos sucessivos de auditoria, atualização e melhoria de segurança que o tornaram um dos sistemas de votação eletrônica mais auditados do mundo. Sua arquitetura de segurança é construída sobre quatro princípios que precisam ser compreendidos em conjunto para que a avaliação de sua robustez seja adequada.

O primeiro princípio é o isolamento: a urna eletrônica não está conectada à internet nem a qualquer rede durante o processo de votação, o que elimina a classe inteira de ataques remotos que vulnerabilidades em sistemas conectados permitem. Essa característica é frequentemente apresentada na narrativa de fraude como limitação do sistema, quando é precisamente o contrário: o isolamento é a medida de segurança mais eficaz disponível contra interferência externa, e os sistemas de votação conectados adotados por outras democracias são objeto de preocupações de segurança significativamente maiores.

O segundo princípio é a verificabilidade por múltiplos atores com interesses potencialmente conflitantes: partidos políticos, candidatos, representantes da Ordem dos Advogados do Brasil, do Ministério Público e de entidades fiscalizadoras credenciadas têm acesso a diferentes etapas do processo de verificação, desde a lacração dos sistemas antes do pleito até a auditoria dos resultados depois. Essa multiplicidade de atores com interesses distintos é o mecanismo mais robusto disponível contra manipulação interna: uma fraude que beneficiasse determinado candidato precisaria envolver a cumplicidade ou a cegueira simultânea de representantes dos candidatos adversários em cada uma das etapas verificáveis.

O terceiro princípio é a auditabilidade por Registro Digital do Voto e Boletim de Urna: cada urna eletrônica produz ao final da votação um Boletim de Urna que registra os totais de cada candidato e que é disponibilizado publicamente antes mesmo de ser transmitido para os sistemas de totalização. Qualquer pessoa ou organização pode coletar esses

boletins e realizar sua própria totalização independente, verificando se o resultado oficial coincide com a soma dos boletins individuais. Essa possibilidade de auditoria cidadã independente foi exercida em múltiplos ciclos eleitorais por organizações não governamentais e por pesquisadores acadêmicos, com resultados consistentes com os totalizados pelo TSE.

O quarto princípio é o Teste Público de Segurança, mecanismo formal pelo qual o TSE disponibiliza anualmente o código-fonte dos sistemas eleitorais e os próprios equipamentos para que pesquisadores, partidos, universidades e outras entidades credenciadas tentem encontrar vulnerabilidades em ambiente controlado. Os resultados desses testes são publicados, as vulnerabilidades identificadas são corrigidas antes do pleito e os relatórios das comissões de avaliação são disponibilizados ao público. É um mecanismo de transparência que a maioria dos sistemas de votação do mundo não possui e que transforma a segurança do sistema de afirmação institucional em resultado verificável de processo adversarial.

A narrativa de fraude que circulou nas eleições brasileiras de 2022, analisada em detalhe no Capítulo 10 desta obra como caso paradigmático de desinformação estrutural, não produziu em nenhum momento uma alegação técnica específica que pudesse ser verificada e refutada pelos mecanismos de auditoria disponíveis. As alegações foram sistematicamente vagas, mutáveis e imunes a evidências contraditórias: quando uma alegação específica era refutada por dados, era substituída por outra de diferente natureza sem que a refutação da anterior produzisse qualquer atualização de crença nos proponentes. Esse padrão epistemológico, documentado por Pennycook e Rand (2019) como característico de crenças motivadas em vez de crenças evidência-responsivas, é o que distingue a narrativa de fraude de 2022 de crítica técnica legítima: ela não estava interessada em evidências porque não era sobre o sistema de votação (PENNYCOOK; RAND, 2019, p. 2521-2523).

A instrumentalização da urna eletrônica como vetor de desinformação eleitoral é, por essas razões, um caso de estudo privilegiado para os temas centrais desta obra. Ela demonstra empiricamente como tecnologias de comunicação digital, especialmente o WhatsApp e as redes sociais, podem transformar narrativas sem fundamento técnico em crenças amplamente difundidas por segmentos significativos do eleitorado. Ela demonstra como a opacidade técnica de sistemas complexos cria oportunidade para desinformação que explora a assimetria de informação entre produtores da narrativa e eleitorado receptor.

E ela demonstra por que literacia midiática não é apenas uma questão de verificação de fontes jornalísticas, mas de capacidade de avaliar criticamente afirmações técnicas

sobre sistemas computacionais, o que é uma competência específica que a maioria dos programas de literacia midiática ainda não desenvolveu adequadamente.

O tratamento rigoroso e baseado em evidências da urna eletrônica que esta seção procurou fornecer não é defesa incondicional do sistema: é o ponto de partida para qualquer discussão séria sobre seus méritos e suas limitações. Melhorias são sempre possíveis e desejáveis, e o mecanismo do Teste Público de Segurança existe precisamente para identificá-las. Mas melhorias genuínas dependem de crítica técnica específica e verificável, não de narrativas de fraude que, como os capítulos desta obra demonstraram, têm por objetivo não melhorar o sistema mas minar a confiança no resultado que ele produz.

Tabela 14. Componentes Técnicos de Segurança da Urna Eletrônica Brasileira

Componente	Como funciona	Garantia de Segurança Oferecida
Hardware dedicado e air-gapped	Equipamento proprietário sem conexão à internet, WiFi ou Bluetooth durante a votação. Fisicamente desconectado de qualquer rede externa no dia da eleição.	Impossibilidade de ataque remoto durante a votação. Atacante precisaria de acesso físico à urna, detectável pelos fiscais presentes.
Sistema operacional customizado	Linux customizado e auditado, compilado especificamente para cada eleição a partir de código-fonte verificado.	Superfície de ataque mínima. Código-fonte disponível para auditorias por qualquer partido ou especialista credenciado.
Assinatura digital do software	Cada componente de software é assinado com chave criptográfica. Hash verificado automaticamente no início de cada votação.	Impossibilidade de substituir o software por versão adulterada. Tentativa de modificação é detectada automaticamente.
Boletim de Urna (BU) impresso	Ao encerrar a votação, a urna imprime automaticamente o total de votos de cada candidato. Cópias são afixadas publicamente na seção eleitoral.	Verificação independente dos resultados digitais. Mais de 500.000 BUs são publicados online pelo TSE para verificação por qualquer cidadão.
Teste Público de Segurança (TPS)	Antes de cada eleição, o TSE convida especialistas de partidos, academia e sociedade civil a tentar encontrar vulnerabilidades com acesso completo ao código-fonte.	Transparência ativa: qualquer vulnerabilidade identificada é publicada e corrigida. Nenhuma edição do TPS resultou em comprometimento comprovado da integridade do voto.

Fonte: Elaborado pelo autor com base em TSE (2022)

■ **A narrativa de fraude nas urnas eletrônicas — Por que persiste apesar das evidências**

Apesar de todos os mecanismos de segurança e transparência documentados, e apesar de múltiplas auditorias independentes confirmarem a integridade do sistema, incluindo verificações realizadas pelas próprias Forças Armadas em 2022, a narrativa de que as urnas eletrônicas brasileiras são fraudáveis circulou com intensidade crescente entre 2018 e 2022. Isso não é paradoxo: é uma demonstração da eficácia da desinformação estrutural que ataca não um fato específico, mas a credibilidade do sistema como um todo, tornando qualquer evidência de segurança suspeita de fazer parte de um encobrimento. Para o profissional do direito eleitoral, isso ensina uma lição fundamental: argumentos técnicos por si sós não derrotam narrativas políticas. É necessário combinar a prova técnica com estratégias de comunicação que construam confiança institucional não apenas refutem alegações pontuais.

LEGISLAÇÃO ELEITORAL: DO CÓDIGO ELEITORAL À RESOLUÇÃO 23.732/2024

O ordenamento jurídico eleitoral brasileiro é composto por um conjunto de diplomas que se articulam em camadas: a Constituição Federal como base, o Código Eleitoral como lei orgânica, a Lei das Eleições como instrumento operacional e as resoluções do TSE como regulamentação específica. Para o contexto digital e de IA, as alterações introduzidas pelas Leis 12.034/2009, 13.165/2015 e 14.197/2021 são as mais relevantes.

Tabela 15. Marco Legal Eleitoral Aplicável à IA

Diploma Legal	Conteúdo Central	Disposições Relevantes para IA
Constituição Federal (1988)	Fundamentos do sistema eleitoral: direito de voto (art. 14), Justiça Eleitoral (art. 121), propaganda eleitoral.	Art. 5º, IV e IX (liberdade de expressão); art. 5º, X (privacidade, honra e imagem); art. 5º, XII (sigilo de comunicações); art. 14 (soberania popular).
Código Eleitoral (Lei 4.737/1965)	Estrutura da Justiça Eleitoral, registro de candidatos, crimes eleitorais (arts. 289-354), processo eleitoral.	Arts. 323-326: crimes contra a honra eleitoral (calúnia, difamação, injúria), aplicáveis a deepfakes e desinformação. Art. 23, IX: competência normativa do TSE.
Lei das Eleições (Lei 9.504/1997)	Regras operacionais das eleições: propaganda, financiamento, votação, apuração, recursos.	Arts. 57-B a 57-J: propaganda na internet; art. 73: condutas vedadas; art. 98: direito de resposta e remoção; art. 222-A: disparo em massa.
Marco Civil da Internet (Lei 12.965/2014)	Direitos e deveres de usuários e provedores; responsabilidade civil de provedores (art. 19).	Art. 19: responsabilidade de provedores (debate com jurisprudência eleitoral sobre responsabilidade por omissão após notificação).

LGPD (Lei 13.709/2018)	Proteção de dados pessoais; bases legais para tratamento; direitos dos titulares; sanções.	Art. 5º, II: opinião política = dado sensível; art. 11: tratamento de dados sensíveis exige consentimento específico; art. 52: sanções da ANPD.
Lei 14.197/2021	Crimes contra o Estado Democrático de Direito (arts. 359-L a 359-S).	Aplicável a campanhas sistemáticas de desinformação que atinjam a credibilidade das instituições eleitorais.
Resolução TSE 23.732/2024	Regulamentação do uso de IA nas eleições: primeiro marco normativo específico do mundo.	Definição de IA, conteúdo sintético e agentes virtuais; obrigações de transparência; vedações absolutas; sanções por violação.

Fonte: Elaborado pelo autor com base na Apostila Curso IA e Eleições módulo 1 (2026) e legislação eleitoral brasileira.

Erosão Democrática e Infraestrutura Digital

A literatura sobre resiliência democrática frente a ameaças tecnológicas expandiu-se de forma significativa a partir de eventos críticos como as eleições norte-americanas de 2016 e o escândalo envolvendo a Cambridge Analytica. Esses episódios evidenciaram a vulnerabilidade das democracias contemporâneas à manipulação informacional em escala e catalisaram um campo de pesquisa voltado à análise dos impactos políticos da infraestrutura digital.

Nesse contexto, Steven Levitsky e Daniel Ziblatt oferecem um diagnóstico que se tornou referência central no debate: as democracias contemporâneas tendem a deteriorar-se não por rupturas abruptas, mas por processos graduais de erosão institucional, nos quais normas informais, práticas democráticas e mecanismos de controle são progressivamente enfraquecidos. No ambiente digital, essa erosão assume uma forma particular: ela ocorre, em grande medida, pela degradação do espaço informacional, no qual a desinformação e a polarização dificultam a construção de consensos mínimos e minam a própria legitimidade das instituições. A ruptura não vem de fora do sistema; ela se instala por dentro, alimentada por dinâmicas comunicacionais que corroem, de forma lenta e muitas vezes imperceptível, os fundamentos do acordo democrático.

De forma complementar, David Runciman identifica o que denomina sequestro tecnológico da democracia: um processo pelo qual a dependência crescente de infraestruturas digitais privadas transfere poder significativo para atores que não estão submetidos ao controle democrático. Plataformas tecnológicas e sistemas algorítmicos passam a influenciar dimensões centrais do processo político sem que haja mecanismos proporcionais de accountability. Essa assimetria entre poder e responsabilidade constitui

um dos problemas estruturais mais urgentes das democracias contemporâneas, e sua resolução não pode ser alcançada apenas por instrumentos jurídicos tradicionais.

O Caso Brasileiro: Vulnerabilidades e Respostas Institucionais

No contexto brasileiro, essas questões assumem contornos específicos e empiricamente documentados. As pesquisas de Jairo Nicolau, Afonso de Albuquerque e Vanessa Quinan, bem como as investigações conduzidas pelo NetLab UFRJ, oferecem evidência consistente sobre a centralidade da desinformação digital nas disputas eleitorais do país. Essas análises identificam a formação de ecossistemas informacionais altamente fragmentados, nos quais conteúdos de baixa qualidade informativa circulam com elevada intensidade, frequentemente impulsionados por dinâmicas de engajamento e coordenação em rede que ampliam artificialmente seu alcance e sua aparência de credibilidade.

A literatura convergente sobre o caso brasileiro aponta para um diagnóstico preocupante: o país apresenta elevada vulnerabilidade estrutural à desinformação digital, decorrente de uma combinação de fatores que se reforçam mutuamente. A ampla penetração de aplicativos de mensageria privada, em especial o WhatsApp, transformou esses ambientes em canais primários de circulação de informação política, com características que os tornam particularmente problemáticos: a comunicação ocorre em espaços opacos, fora do alcance dos mecanismos de moderação das plataformas abertas, e se beneficia do capital de confiança associado às redes de relacionamento pessoal. A isso se soma um nível relativamente baixo de literacia midiática em segmentos significativos da população, o que reduz a capacidade de avaliação crítica do conteúdo recebido.

Ao mesmo tempo, observa-se uma resposta institucional em processo de amadurecimento. Iniciativas regulatórias, decisões judiciais e mecanismos de cooperação entre instituições públicas e plataformas digitais indicam um movimento de adaptação do sistema eleitoral brasileiro às novas dinâmicas tecnológicas. Esse processo, contudo, enfrenta limitações estruturais relevantes: a dificuldade de monitorar ambientes comunicacionais privados, a velocidade com que novas estratégias de desinformação emergem e se transformam e a assimetria de recursos entre os órgãos de controle e os atores que operam as infraestruturas digitais. A resposta institucional, por mais que avance, tende a chegar com atraso em relação à sofisticação crescente das práticas que busca coibir.

Integridade Eleitoral como Problema Sistêmico

A questão da integridade eleitoral, no contexto contemporâneo, não pode ser tratada como um problema técnico de fiscalização procedimental. Ela deve ser compreendida como um problema sistêmico, que envolve a interação entre tecnologia, comportamento social e instituições políticas, e que exige respostas igualmente sistêmicas.

A preservação da integridade das eleições depende não apenas da robustez dos procedimentos formais, mas da capacidade do sistema democrático de garantir condições equitativas e transparentes para a formação da vontade política em um ambiente informacional crescentemente mediado por plataformas digitais. Isso implica reconhecer que a regulação do espaço informacional é, ela própria, uma condição de possibilidade da democracia, e não uma intervenção externa sobre ela.

Sem informação de qualidade, sem equidade comunicacional e sem mecanismos efetivos de accountability sobre os sistemas que organizam o fluxo informacional, o voto permanece formalmente livre, mas substantivamente condicionado por forças que o eleitor, em grande medida, não vê e não controla.

O Paradoxo Brasileiro: Segurança Técnica e Vulnerabilidade Informacional

O caso brasileiro ilustra esse problema sistêmico com particular clareza. O país possui um dos sistemas eleitorais mais tecnologicamente avançados do mundo, paradoxo que precisa ser compreendido em sua complexidade. A adoção da urna eletrônica em 1996, pioneira no contexto das democracias de grande porte, eliminou virtualmente as formas clássicas de fraude na contagem de votos e na apuração dos resultados.

O Tribunal Superior Eleitoral administra um processo de mais de 150 milhões de eleitores com eficiência logística comparável à de países com populações muito menores. Esse mesmo sistema, contudo, enfrenta vulnerabilidades estruturais no ambiente digital que os mecanismos de segurança da urna eletrônica não foram projetados para enfrentar. A fraude eleitoral migrou do momento da votação, onde a barreira tecnológica é alta, para o momento da formação de preferências, onde a barreira regulatória é significativamente menor. É nesse intervalo, entre o momento em que o eleitor começa a formar sua opinião e o momento em que deposita seu voto na urna, que a comunicação política algorítmica opera com maior liberdade e menor fiscalização. A solidez do voto não garante, por si só, a integridade da escolha que o precedeu.

O quadro a seguir sistematiza esse diagnóstico, organizando as principais dimensões do sistema eleitoral brasileiro a partir da tensão entre suas forças institucionais e suas vulnerabilidades estruturais no ambiente digital. A justaposição entre as duas colunas revela um padrão consistente: o Brasil construiu salvaguardas robustas para os riscos que enfrentava no século XX, mas ainda busca respostas proporcionais aos desafios que o século XXI introduziu.

Tabela 16. Forças e Vulnerabilidades do Sistema Eleitoral Brasileiro na Era da IA

Dimensão	Força	Vulnerabilidade
Votação	Urna eletrônica com múltiplas camadas de segurança e auditoria	Narrativas falsas sobre suposta fraude na urna como instrumento de deslegitimação
Regulação	TSE com poder normativo e jurisdicional amplo	Assimetria entre velocidade tecnológica e ritmo regulatório
Fiscalização	CIEDDE e parcerias com plataformas e fact-checkers	Comunicação privada (WhatsApp) fora do alcance efetivo de monitoramento
Sanção	Sanções eleitorais com consequências até cassação de mandato	Dificuldade de identificação de responsáveis em operações anônimas
Cidadania	Alta taxa de participação eleitoral (voto obrigatório)	Baixa literacia midiática média, especialmente em classes C e D
Tecnologia	Infraestrutura tecnológica do TSE reconhecida internacionalmente	Plataformas digitais com sede no exterior e alcance global

Fonte: Elaborado pelo autor, 2026.

A Resposta Institucional: TSE, CIEDDE e o Marco Regulatório

A resposta institucional brasileira à ameaça da desinformação e da manipulação algorítmica nas eleições tem sido progressivamente mais sofisticada. O Tribunal Superior Eleitoral desenvolveu, ao longo das últimas gestões, uma estratégia multidimensional que combina regulação normativa, cooperação com plataformas digitais, parceria com agências de *fact-checking* e desenvolvimento de capacidade técnica interna. Essa estratégia não emergiu de forma linear: ela foi construída de maneira incremental, em resposta direta aos desafios que cada ciclo eleitoral revelou, e reflete um aprendizado institucional acumulado sobre a natureza das ameaças que o ambiente digital representa para a integridade do processo eleitoral.

O CIEDDE como Expressão Institucional

O Centro Integrado de Enfrentamento à Desinformação e Defesa da Democracia Eleitoral, o CIEDDE, é a expressão mais acabada dessa estratégia. Criado em 2022, o centro opera de forma permanente, com intensificação das atividades nos períodos eleitorais, e integra equipes de análise técnica, jurídica e de comunicação em uma estrutura concebida para responder com agilidade a um ambiente informacional em constante transformação.

Sua atuação compreende quatro frentes articuladas. A primeira é o monitoramento ativo do ambiente informacional digital, com identificação sistemática de conteúdos potencialmente desinformativos em circulação. A segunda é a análise e classificação desses conteúdos, que envolve a avaliação de sua natureza, origem e potencial de dano ao processo eleitoral. A terceira é o acionamento das plataformas para remoção de conteúdo irregular, com base nos marcos normativos vigentes e nos acordos de cooperação estabelecidos. A quarta é a produção de conteúdo informativo destinado à contraposição ativa de narrativas falsas, reconhecendo que a remoção isolada de conteúdo é insuficiente quando as narrativas já circulam amplamente.

Essa arquitetura institucional representa um avanço significativo em relação aos modelos reativos que predominavam até recentemente, nos quais a resposta ao conteúdo desinformativo dependia de denúncias externas e tramitava por vias exclusivamente judiciais. A criação de uma estrutura permanente, com capacidade técnica própria e canais diretos de comunicação com as plataformas, reduz o tempo de resposta e amplia a capacidade de atuação preventiva.

Alcance e Limites da Resposta Institucional

A sofisticação crescente da resposta institucional, contudo, não elimina os limites estruturais que a condicionam. O primeiro e mais persistente desses limites é a assimetria de velocidade: o ritmo dos processos normativos e judiciais é, por natureza, mais lento do que a velocidade com que novas estratégias de desinformação emergem, se disseminam e se transformam. Quando uma regulação é elaborada, debatida e implementada, o fenômeno que a motivou frequentemente já assumiu uma forma distinta.

O segundo limite é a opacidade dos ambientes de comunicação privada. O WhatsApp, canal primário de circulação de desinformação política no Brasil, opera em um espaço estruturalmente resistente ao monitoramento institucional: as mensagens circulam em grupos fechados, sem visibilidade pública, e a arquitetura de criptografia da plataforma impõe restrições técnicas e jurídicas ao acesso por parte de autoridades. Esse conjunto de características torna a fiscalização efetiva desse ambiente um desafio que nenhuma das respostas institucionais disponíveis até o momento foi capaz de resolver de forma satisfatória.

O terceiro limite é a extraterritorialidade das plataformas. Empresas com sede no exterior e alcance global estão sujeitas a marcos regulatórios distintos e nem sempre respondem com a celeridade que as autoridades eleitorais brasileiras demandam. A cooperação voluntária, embora tenha avançado, permanece dependente da disposição das plataformas em acatar as solicitações do TSE, o que introduz uma variável de incerteza que o sistema institucional não controla inteiramente.

Esses limites não diminuem a relevância do que foi construído. Eles indicam, antes, que a resposta institucional à desinformação eleitoral é, por sua própria natureza, um projeto inacabado, que precisa ser continuamente revisado à medida que o ambiente tecnológico evolui. A questão central não é se o sistema institucional brasileiro está à altura dos desafios que enfrenta, mas se ele está aprendendo com a velocidade necessária para não ficar permanentemente um passo atrás.

O CIEDDE — Competências e Estrutura

O Centro Integrado de Enfrentamento à Desinformação e Defesa da Democracia Eleitoral (CIEDDE) é estrutura permanente do TSE, com as seguintes competências: (i) monitoramento do ambiente informacional eleitoral nas principais plataformas digitais; (ii) análise técnica e jurídica de conteúdo potencialmente desinformativo; (iii) articulação com plataformas digitais para remoção de conteúdo eleitoral irregular; (iv) coordenação com agências de checagem credenciadas; (v) produção de alertas e comunicados sobre desinformação eleitoral; (vi) cooperação com organismos internacionais de integridade eleitoral.

Brasil. TSE. Portaria TSE nº 544/2022. Brasília, DF: TSE, 2022.

Fluxo Operacional do CIEDDE

1. **Monitoramento automatizado:** Ferramentas de social listening monitoram em tempo real o fluxo de conteúdo nas principais plataformas, identificando

narrativas com crescimento anômalo de velocidade de propagação — sinal potencial de amplificação coordenada (bots ou disparo em massa).

2. **Triagem e priorização:** A equipe técnica avalia os alertas gerados pelo monitoramento automatizado e prioriza com base em: potencial de dano eleitoral, velocidade de propagação, plausibilidade da desinformação e proximidade temporal de eventos eleitorais críticos.
3. **Verificação e checagem:** Agências de checagem parceiras são acionadas para verificação de fatos. O resultado da checagem alimenta tanto as decisões de moderação das plataformas quanto as comunicações públicas do TSE.
4. **Resposta coordenada:** Dependendo da avaliação, a resposta pode incluir: notificação às plataformas para moderação voluntária; decisão judicial de remoção pelo corregedor eleitoral; comunicação pública de desmentido pelo TSE; e, nos casos mais graves, acionamento da Polícia Federal para investigação de crimes eleitorais.
5. **Documentação para responsabilização:** Todas as decisões e o conteúdo removido são documentados em repositório que pode ser utilizado em processos de responsabilização pós-eleitoral — cassação de mandato, impugnação de registro, ação penal eleitoral.

■ Avaliação crítica do CIEDDE — Conquistas e desafios persistentes

O CIEDDE representa avanço significativo em relação ao modelo anterior de resposta isolada, caso a caso. A integração de atores reduz o tempo de resposta e o risco de decisões contraditórias. Contudo, desafios persistem: (1) O modelo funciona bem para desinformação em plataformas abertas (Facebook, Twitter/X) — mas tem alcance limitado para o WhatsApp, cujo conteúdo criptografado não é acessível ao monitoramento externo; (2) A capacidade de moderação das plataformas é exercida por equipes externas ao controle do TSE — o que cria dependência de cooperação voluntária sem enforcement adequado; (3) O CIEDDE opera em modo intensivo apenas durante o período eleitoral — a desinformação que prepara o terreno nos meses anteriores à janela formal recebe menos atenção.

Cooperação com Plataformas, Agências de Checagem e Anatel

A arquitetura de combate à desinformação eleitoral no Brasil não repousa sobre um único ator com poder centralizado de intervenção. Ela depende, estruturalmente, de uma rede de cooperação entre atores de natureza, motivação e capacidade técnica distintas: o TSE como autoridade eleitoral com poder normativo e jurisdicional; as plataformas digitais como intermediários que controlam, de fato, o fluxo de conteúdo que alcança o eleitor; as agências de checagem como verificadores independentes com credibilidade editorial; e a

Anatel como regulador das infraestruturas de comunicação sobre as quais toda essa cadeia opera.

Essa arquitetura distribuída tem uma vantagem e uma fragilidade que decorrem da mesma característica. A vantagem é a complementaridade: nenhum dos atores, isoladamente, tem capacidade de endereçar o problema em toda a sua extensão, e a combinação de competências distintas permite uma cobertura mais ampla do que qualquer modelo centralizado conseguiria. A fragilidade é a dependência: a eficácia do sistema como um todo é limitada pelo elo mais fraco, e a cooperação entre atores com interesses nem sempre convergentes introduz uma variável de incerteza que nenhum acordo formal consegue eliminar inteiramente.

A questão central que essa arquitetura coloca não é técnica, mas política e econômica. As plataformas digitais são empresas privadas com sede no exterior, sujeitas a múltiplas jurisdições e com modelos de negócios que, como os capítulos anteriores demonstraram, têm incentivos estruturais que nem sempre se alinham com os objetivos de integridade eleitoral. As agências de checagem dependem de financiamento e de acesso a dados que as plataformas controlam. A Anatel tem competência sobre a infraestrutura, mas não sobre o conteúdo. E o TSE tem autoridade eleitoral, mas não tem poder regulatório sobre as plataformas fora do período eleitoral.

A eficácia dessa rede depende, portanto, de duas condições que precisam ser examinadas separadamente: a qualidade dos acordos que formalizam a cooperação entre seus membros, e a convergência suficiente de interesses que torna esses acordos funcionais na prática. A primeira condição é relativamente acessível à análise jurídica. A segunda é a que mais frequentemente determina se o sistema funciona quando é mais necessário.

O quadro a seguir sistematiza os principais parceiros do TSE nessa rede de cooperação, detalhando os acordos formalizados com cada ator e as responsabilidades específicas que assumem no ecossistema de integridade informacional eleitoral. A leitura revela tanto o alcance do que foi construído quanto os pontos de fragilidade que permanecem: a cooperação voluntária das plataformas cobre as superfícies visíveis do problema, mas deixa descoberto precisamente o ambiente onde a desinformação circula com maior intensidade e menor rastreabilidade.

REDE DE COOPERAÇÃO ELEITORAL CONTRA DESINFORMAÇÃO

Parceiros do TSE e seus acordos no ecossistema de integridade informacional

Ator	Acordos e responsabilidades no ecossistema eleitoral
PLATAFORMAS DIGITAIS	
Meta *Facebook / Instagram / WhatsApp*	— Canal prioritário de reporte de desinformação eleitoral ao TSE. — Remoção de conteúdo em desacordo com políticas e ordens judiciais. — Fornecimento de dados agregados de tendências de conteúdo ao CIEDDE (sem acesso a conteúdo individual criptografado). — Limitação ao encaminhamento em massa no WhatsApp. — Limitação crítica: conteúdo em grupos privados permanece fora do alcance de moderação proativa.
Google *YouTube / Search*	— Remoção de vídeos com desinformação eleitoral no YouTube mediante solicitação do TSE. — Implementação de avisos contextuais em pesquisas sobre temas eleitorais sensíveis. — Fornecimento de dados de tendências de busca ao TSE. — Proibição de publicidade eleitoral paga por atores não certificados.
TikTok *Crescimento acelerado entre jovens eleitores*	— Acordo com TSE desde 2022 para remoção de conteúdo eleitoral enganoso. — Implementação de tags de verificação em conteúdo eleitoral. — Restrições a anúncios políticos pagos. — Desafio específico: moderação de vídeo curto em escala é tecnicamente mais complexa do que texto.
VERIFICAÇÃO E CONTROLE	
Agências de checagem *Lupa, Aos Fatos, AFP Checamos, UOL Confere, E-Farsas, Estadão Verifica*	— Compõem a rede de verificadores parceiros do CIEDDE. — Responsáveis pela verificação de fatos que alimenta decisões de moderação e comunicações públicas do TSE. — Independência editorial preservada: o TSE não pode determinar o resultado das checagens.
Anatel *Regulador de infraestrutura de comunicações*	— Atuação no combate a disparos em massa via SMS e ligações automatizadas. — Monitoramento de atividade anômala de chips — padrão que identifica registros em massa para disparos. — Cooperação com Polícia Federal para investigação de serviços de disparo ilegal.
Polícia Federal *Investigação de crimes eleitorais digitais*	— Investigação de financiamento ilegal de campanhas de desinformação. — Crimes contra a honra de candidatos via internet. — Operações coordenadas de influência. — Integra o CIEDDE durante períodos eleitorais e recebe relatórios de conteúdo criminal identificado pelas plataformas.

Fonte: Elaborado pelo autor com base em acordos públicos do TSE com plataformas digitais (2022–2024) e Portaria TSE nº 544/2022.

A Resolução TSE nº 23.732/2024: Avanços e Limitações

A Resolução TSE nº 23.732/2024 representa o marco regulatório mais relevante sobre o uso de inteligência artificial no processo eleitoral brasileiro. Sua edição, fruto de processo de consulta pública e de análise comparada de regulações internacionais, estabelece um regime dual: obrigações de transparência para os usos permitidos de IA e vedações expressas para os usos mais problemáticos.

Entre os avanços mais significativos, destacam-se a definição normativa de IA e de conteúdo sintético, que cria base conceitual para a aplicação das demais normas; a obrigação de rotulagem de conteúdo gerado ou significativamente modificado por IA, que visa restaurar a transparência retirada pela comunicação política automatizada; a vedação expressa de deepfakes e simulações fraudulentas de candidatos; e a extensão das obrigações ao período de pré-campanha, fechando uma lacuna explorada nos ciclos eleitorais anteriores.

As limitações da resolução são igualmente relevantes para uma análise crítica honesta. A efetividade da obrigação de rotulagem depende da capacidade das plataformas de detectar e classificar conteúdo gerado por IA, capacidade que ainda é limitada, especialmente para deepfakes de alta qualidade. Esse problema é, em larga medida, uma manifestação específica do limite estrutural já identificado: a regulação tende a chegar depois da tecnologia que busca regular.

Mais complexo, porém, é o problema da responsabilização. Atribuir a candidatos responsabilidade por conteúdo produzido por terceiros sem seu conhecimento direto levanta questões de garantia constitucional que a jurisprudência do TSE ainda está consolidando, e para as quais a resolução não oferece resposta inteiramente satisfatória. Definir o limiar entre negligência, omissão e dolo em campanhas que operam em ambientes digitais descentralizados é um desafio que extrapola a capacidade normativa de qualquer resolução isolada e demanda construção jurisprudencial progressiva.

Fact-Checking: Metodologia, Alcance e Limitações

O fact-checking, entendido como a verificação profissional e sistemática de afirmações factuais circulantes no espaço público, é hoje uma das respostas mais institucionalizadas à desinformação e, ao mesmo tempo, uma das mais frequentemente mal compreendidas em seu alcance real.

No Brasil, um ecossistema relativamente robusto de agências de checagem se desenvolveu a partir de 2015, quando o país se tornou signatário da rede internacional IFCN, a International Fact-Checking Network, que estabelece padrões metodológicos e critérios de credenciamento para agências em todo o mundo. Compreender tanto a metodologia quanto os limites estruturais do fact-checking é condição para avaliar com

precisão seu papel como instrumento de combate à desinformação eleitoral, evitando tanto a superestimação de suas capacidades quanto a subestimação de sua contribuição.

O processo metodológico padrão das agências credenciadas pelo IFCN organiza-se em cinco etapas que, em conjunto, constituem o que se pode denominar uma epistemologia jornalística aplicada à verificação de fatos.

A primeira etapa é a de seleção e priorização. Nem toda afirmação falsa é verificável no sentido técnico do termo: opiniões, previsões e interpretações subjetivas estão fora do escopo do fact-checking, que se concentra em afirmações factuais concretas, passíveis de confronto com evidências.

A priorização entre o universo de afirmações verificáveis é orientada por critérios de impacto potencial, alcance de circulação e relevância eleitoral. Em períodos eleitorais, essa triagem é decisiva: o volume de conteúdo desinformativo em circulação supera em ordens de magnitude a capacidade de verificação das agências, o que torna a escolha do que verificar um ato com consequências políticas próprias.

A segunda etapa é a coleta de evidências, que constitui o núcleo metodológico do processo. A checagem profissional recorre a fontes primárias oficiais, como dados do IBGE, do TSE e de ministérios, a publicações científicas revisadas por pares, ao contato direto com especialistas independentes e à análise documental.

Para conteúdo audiovisual, a metodologia inclui técnicas de análise de autenticidade que permitem identificar manipulações, inclusive em nível de metadados. A qualidade da evidência coletada é o que distingue o fact-checking jornalístico profissional do mero contradiscurso: a checagem não refuta uma afirmação com outra afirmação, mas com evidência documentada e metodologicamente rastreável.

A terceira etapa é a classificação. O conteúdo verificado é enquadrado em uma escala tipicamente de quatro a seis níveis, que varia conforme a agência, mas converge em torno de categorias como Verdadeiro, Verdadeiro com contexto, Impreciso, Falso, Sem evidência e Sátira.

A classificação é publicada junto com o raciocínio completo que a sustenta, o que constitui o elemento de transparência metodológica exigido pelo IFCN como critério central de credenciamento. Essa exigência de transparência distingue o fact-checking da moderação de conteúdo por plataformas: a checagem é um ato público e contestável, não uma decisão opaca tomada por algoritmo.

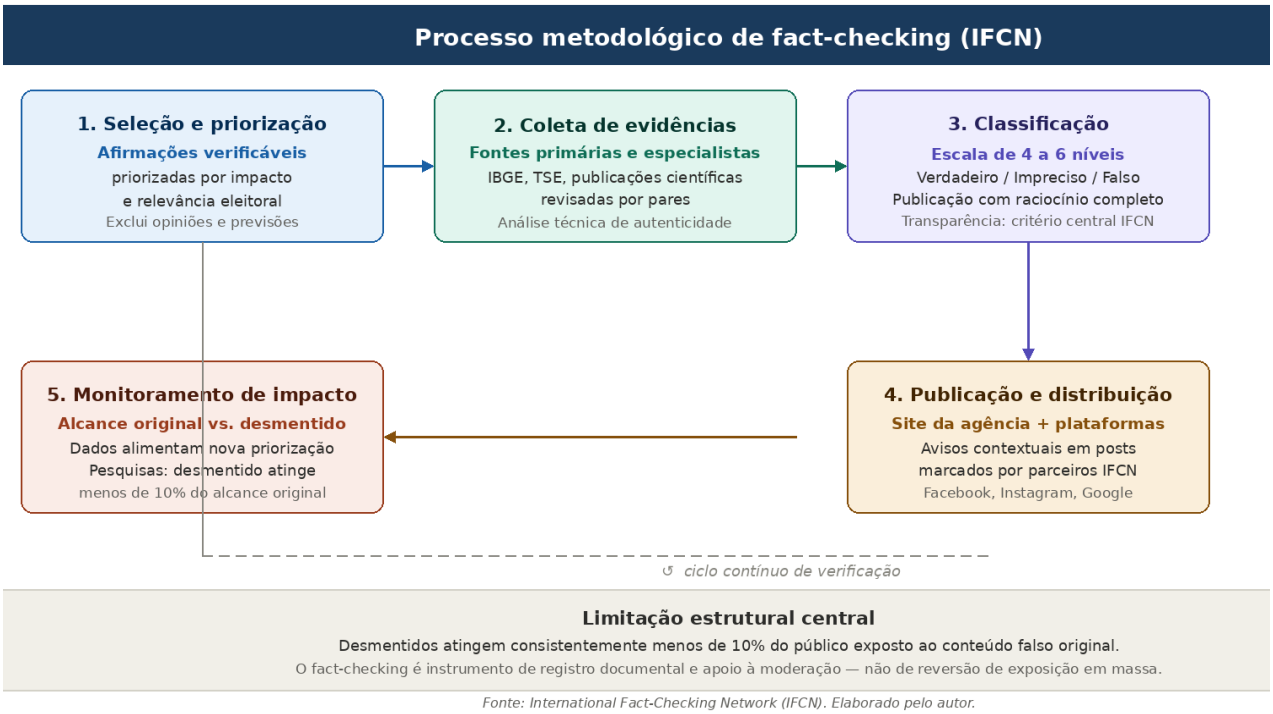
A quarta etapa é a publicação e distribuição do resultado. O conteúdo verificado é publicado no site da agência e, nos casos em que existem parcerias formalizadas com plataformas, pode ser utilizado para a implementação de avisos contextuais nos posts originais. O Facebook e o Instagram, por exemplo, exibem alertas visíveis em publicações marcadas por parceiros de checagem credenciados, reduzindo o alcance orgânico do conteúdo e informando os usuários sobre a existência de verificação contrária. Essa integração entre checagem editorial e moderação algorítmica é um dos desenvolvimentos mais relevantes do ecossistema de combate à desinformação da última década.

A quinta etapa é o monitoramento de impacto, e é aqui que a maior limitação estrutural do modelo se revela com clareza. Pesquisas sistemáticas sobre o alcance comparativo entre conteúdo desinformativo original e seu respectivo desmentido mostram de forma consistente que os desmentidos atingem uma fração do público que foi exposto ao conteúdo falso.

A assimetria não é marginal: em muitos casos documentados, a checagem alcança menos de 10% do público original. O mecanismo que explica esse resultado é o mesmo que o capítulo anterior descreveu: o conteúdo falso tende a ser emocionalmente mais intenso, mais novo e mais surpreendente, o que o torna estruturalmente mais viral do que o desmentido, por mais preciso e bem documentado que seja.

Essa assimetria de alcance não invalida o fact-checking como instrumento, mas redefine sua função. A checagem profissional não é, na maioria dos casos, uma ferramenta de alcance de massa capaz de reverter a exposição de grandes públicos a conteúdo falso.

Ela é, antes, um instrumento de construção de registro documental público, de apoio às decisões de moderação das plataformas, de fornecimento de evidências para processos judiciais e administrativos e de referência para jornalistas e autoridades que precisam de suporte técnico para suas próprias avaliações. Compreender essa função real, em lugar da função idealizada, é condição para desenhar estratégias de combate à desinformação eleitoral que sejam calibradas para o que o fact-checking efetivamente pode entregar.



Fonte: International Fact-Checking Network (IFCN). Elaborado pelo autor.

■ O problema da assimetria de alcance e o 'backfire effect'

Duas limitações estruturais do fact-checking merecem atenção especial. A primeira é a assimetria de alcance: o conteúdo falso viral atinge milhões de pessoas; o desmentido atinge dezenas de milhares, na melhor das hipóteses. Essa assimetria é sistemática e documentada, o desmentido não desfaz o dano. A segunda é o chamado 'backfire effect', documentado por Nyhan & Reifler (2010): em algumas condições, expor pessoas a um desmentido de informação que confirma suas crenças políticas resulta em aumento da adesão à crença falsa, o desmentido ativa defensividade cognitiva. Pesquisas mais recentes questionam a generalidade desse efeito, mas o alerta metodológico permanece relevante: fact-checking pode não funcionar, e em alguns casos pode ser contraproducente, dependendo do perfil do receptor e da forma como o desmentido é comunicado.

A Assimetria de Literacia Digital

O Brasil é um país de extremos digitais. Nas grandes cidades e entre os segmentos de maior escolaridade, há usuários com capacidade de verificar fontes, identificar desinformação e navegar criticamente pelo ambiente digital. Nos estratos de menor escolaridade e nas regiões de menor desenvolvimento socioeconômico, o WhatsApp é frequentemente a principal fonte de informação política, e a capacidade de avaliação crítica do conteúdo recebido é significativamente limitada.

Essa assimetria tem implicações eleitorais diretas: os segmentos mais vulneráveis à desinformação são, com frequência, os que mais a consomem, e que podem ter seu comportamento eleitoral mais significativamente alterado por campanhas bem executadas.

A regulação eleitoral não alcança essa dimensão do problema. Apenas programas de educação midiática de longo prazo, integrados ao sistema educacional formal e desenhados com atenção às especificidades regionais e socioeconômicas do país, têm potencial de impacto estrutural. Trata-se, portanto, de uma vulnerabilidade que se resolve na escala de gerações, não de ciclos eleitorais.

■ Evidência de eficácia — Prebunking: inoculação cognitiva contra desinformação

Uma das abordagens de literacia midiática com maior evidência experimental de eficácia é o 'prebunking', inoculação cognitiva contra desinformação antes da exposição. Analogamente a uma vacina (que expõe o organismo a uma versão enfraquecida do patógeno para criar imunidade), o prebunking expõe o indivíduo a exemplos de técnicas de manipulação antes que ele as encontre em contexto real, criando resistência cognitiva. O projeto Inoculation Theory, desenvolvido por pesquisadores de Cambridge com apoio do Google e verificado em experimentos com mais de 30.000 participantes, demonstrou que o prebunking — especialmente quando implementado como jogo interativo (como o jogo 'Go Viral!' sobre desinformação de COVID) — reduz significativamente a suscetibilidade a desinformação por semanas após a intervenção, independente do viés político do participante.

A Opacidade das Plataformas

Os limites da cooperação voluntária entre o TSE e as plataformas digitais já foram identificados. O que importa examinar aqui é a resposta regulatória internacional que emerge como alternativa a esse modelo e suas implicações para o caso brasileiro.

A aprovação do Digital Services Act na União Europeia estabelece um paradigma distinto: em vez de depender da disposição das plataformas em cooperar, impõe obrigações legalmente vinculantes de transparência algorítmica e garante acesso de reguladores e pesquisadores a dados e sistemas. Trata-se de uma mudança qualitativa na relação entre Estado e plataformas, que desloca o poder de definir os termos da cooperação das empresas para as autoridades públicas.

O Projeto de Lei nº 2.338/2023, em tramitação no Congresso Nacional, busca parcialmente replicar esse modelo no Brasil. Sua aprovação e implementação efetiva, contudo, permanecem incertas ao tempo da redação desta obra. Enquanto esse marco não se consolida, o sistema eleitoral brasileiro continua dependente de uma arquitetura de cooperação cujos limites são definidos, em última instância, pelas próprias plataformas que deveriam ser fiscalizadas.

As Janelas de Vulnerabilidade

A assimetria entre evolução tecnológica e produção normativa, já identificada como limite estrutural da resposta institucional, assume no contexto eleitoral uma dimensão específica que merece atenção própria: o problema das janelas de vulnerabilidade.

A Resolução TSE nº 23.732/2024 foi editada para regular tecnologias disponíveis até o início de 2024. Nos meses que se seguiram à sua publicação, novos modelos de inteligência artificial generativa foram lançados com capacidades significativamente superiores às que a norma foi projetada para enfrentar. Esse intervalo, entre a edição de uma norma e sua próxima atualização, não é um vazio regulatório qualquer: é precisamente o período em que atores maliciosos podem explorar, com relativa impunidade, as tecnologias que a regulação ainda não alcança.

No contexto eleitoral, esse timing tem consequências agravadas. As eleições ocorrem em datas fixas e previsíveis. Um ator que identifica uma janela de vulnerabilidade pode planejar sua exploração para o momento de maior impacto potencial, que é o período de campanha imediatamente anterior ao pleito. A resposta regulatória, quando vier, chegará depois dos votos. Esse é o desafio que nenhuma arquitetura normativa estática consegue resolver por si só, e que aponta para a necessidade de mecanismos de atualização normativa contínua, capazes de acompanhar o ritmo da evolução tecnológica sem depender exclusivamente dos ciclos legislativos convencionais.

Comparação Internacional: O Brasil em Perspectiva

Uma avaliação honesta da preparação do Brasil para os desafios da IA eleitoral exige perspectiva comparada. Nenhuma democracia encontrou, até o momento, respostas plenamente satisfatórias para esses desafios, o que não significa que todas estejam igualmente preparadas.

O quadro a seguir organiza esse panorama comparativo, reunindo os principais marcos regulatórios adotados por democracias que enfrentam desafios semelhantes. A comparação revela que o Brasil não está sozinho nas suas dificuldades, mas também que algumas escolhas regulatórias de outros países apontam caminhos ainda não percorridos.

Tabela 17. Comparação Internacional de Marcos Regulatórios sobre IA Eleitoral

País/Bloco	Marco Regulatório	Ponto Forte	Limitação
Brasil	Resolução TSE nº 23.732/2024	Cobertura ampla, aplicação imediata	Dependência de cooperação voluntária das plataformas
União Europeia	AI Act + DSA (2024)	Obrigações vinculantes, auditoria algorítmica	Implementação gradual, jurisdição limitada ao território da UE
Estados Unidos	Diretrizes da FEC (2024)	Foco em publicidade política	Fragmentado, sem lei federal abrangente sobre IA eleitoral
Reino Unido	Online Safety Act (2023)	Obrigações de moderação de conteúdo	Não aborda especificamente IA eleitoral
Índia	Código de Conduta Eleitoral (2024)	Rápida implementação em eleições massivas	Autorregulação, sem mecanismos de enforcement robusto
Finlândia	Educação midiática integrada ao currículo	Alta resiliência populacional à desinformação	Abordagem de longo prazo, sem impacto imediato

Fonte: Elaborado pelo autor com base em Norris (2014), AI Act (2024) e TSE (2024).

A análise comparada revela que o Brasil figura entre os países com resposta institucional mais proativa ao problema da IA eleitoral, especialmente considerando o porte do eleitorado e os recursos institucionais disponíveis. A Resolução TSE nº 23.732/2024 é mais abrangente do que a maioria das regulações setoriais eleitorais existentes no mundo, e o CIEDDE tem sido objeto de interesse de organismos internacionais como modelo de estrutura permanente de combate à desinformação.

Esse reconhecimento, contudo, não autoriza complacência. As limitações estruturais identificadas ao longo deste capítulo, a assimetria de literacia digital, a dependência de cooperação voluntária das plataformas e a velocidade da evolução tecnológica, são vulnerabilidades reais que a resposta institucional ainda não endereçou de forma satisfatória. O Brasil construiu uma das arquiteturas de integridade eleitoral mais sofisticadas do mundo para os riscos que conhece. O desafio é garantir que essa arquitetura evolua na mesma velocidade dos riscos que ainda estão por vir.

Levitsky, Ziblatt e o Risco de Erosão Democrática

A conclusão deste capítulo exige um olhar sobre o horizonte mais amplo. Levitsky e Ziblatt, em sua análise dos processos de erosão democrática contemporânea, identificam

um padrão que se tornou referência obrigatória no debate sobre fragilidade institucional: democracias não morrem subitamente, por golpes militares clássicos, mas gradualmente, por uma acumulação de pequenas erosões, cada uma individualmente tolerável, mas cujo conjunto acaba por comprometer de forma irreversível as condições que tornam o regime legítimo.

O uso sistemático da inteligência artificial para manipular o ambiente informacional eleitoral se insere precisamente nesse padrão. Desinformação em escala, microtargeting sem transparência, deepfakes de candidatos, amplificação artificial de narrativas polarizadoras: nenhum desses mecanismos, isoladamente, cancela uma eleição. Nenhuma eleição em que operam deixa de ser, formalmente, uma eleição democrática. Mas cada uma delas corrói, um pouco mais, as condições substantivas que tornam a democracia legítima: a informação adequada dos eleitores, a deliberação pública autêntica e a competição em condições de razoável igualdade. A erosão não aparece nos resultados oficiais. Ela se acumula silenciosamente, eleição após eleição, até que o regime que permanece de pé já não é, em sentido pleno, o que costumava ser.

A pergunta que esta primeira parte do livro coloca ao leitor não é retórica. O Brasil está preparado? A resposta honesta é: parcialmente. Mais do que muitas democracias de porte comparável; menos do que seria necessário para enfrentar os desafios que as eleições de 2026 e os ciclos seguintes apresentarão. O que separa "parcialmente" de "suficientemente" é, em grande medida, o conhecimento, o compromisso e a ação dos atores que este livro se propõe a alcançar. Não há arquitetura institucional capaz de substituir, por si só, a vigilância de uma sociedade que compreende o que está em jogo.

■ RESUMO DO CAPÍTULO:

Este capítulo completou o diagnóstico iniciado nos capítulos anteriores, avaliando a capacidade institucional brasileira de resposta à ameaça da inteligência artificial eleitoral. A análise revelou um quadro de contrastes: o Brasil possui forças institucionais relevantes, especialmente o TSE, o CIEDDE e um marco normativo progressivamente mais robusto, mas enfrenta vulnerabilidades estruturais que nenhuma regulação, por mais sofisticada, é capaz de resolver isoladamente. A assimetria de literacia digital, a dependência de cooperação voluntária das plataformas e a velocidade da evolução tecnológica configuram um conjunto de fragilidades que exigem respostas de natureza distinta e de horizonte mais longo do que o ciclo eleitoral permite. A análise comparada situou o Brasil em posição que merece ser lida com equilíbrio: mais avançado institucionalmente do que muitos países de porte comparável, mas ainda aquém do que a magnitude do problema exige. Esse reconhecimento não é motivo de conforto; é, antes, um ponto de partida para o que ainda precisa ser construído. A perspectiva de Levitsky e Ziblatt ofereceu o enquadramento teórico necessário para compreender a natureza do risco: a erosão democrática não se anuncia por rupturas visíveis, mas se instala por acumulação silenciosa de pequenas vulnerabilidades, cada uma tolerável em si mesma, mas cujo conjunto pode tornar-se irreversível. É nesse registro que a ameaça

da IA eleitoral deve ser lida. Com esse diagnóstico estabelecido, a Parte II desta obra volta o olhar para os instrumentos concretos da ameaça: as ferramentas tecnológicas que candidatos, campanhas e atores maliciosos têm à sua disposição, e os limites legais que a regulação brasileira estabelece para seu uso.

QUESTÕES PARA REFLEXÃO E DEBATE

1. A distinção proposta por Wardle e Derakhshan entre *mis-information*, *dis-information* e *mal-information* é operacionalizável no direito eleitoral brasileiro, ou os tipos legais existentes são insuficientes para capturar a complexidade do fenômeno?

2. O conceito de integridade eleitoral, na formulação de Pippa Norris, pode ser adequadamente aplicado a ambientes em que a formação da preferência eleitoral ocorre predominantemente em plataformas privadas sujeitas a lógicas de engajamento? Quais revisões conceituais essa aplicação exige?

3. A criação do CIEDDE representa um avanço institucional genuíno ou um modelo de resposta que, por depender da cooperação voluntária das plataformas, reproduz estruturalmente a assimetria que pretende corrigir?

4. O paradoxo brasileiro, alta segurança técnica no momento do voto e elevada vulnerabilidade no momento da formação da preferência, revela uma limitação do próprio conceito de integridade eleitoral tal como foi historicamente construído? Como o direito eleitoral deve responder a essa defasagem?

5. A perspectiva de Levitsky e Ziblatt sobre erosão democrática gradual implica que o sistema jurídico-eleitoral deve adotar um princípio de precaução diante de tecnologias cujos efeitos sobre a deliberação ainda não são inteiramente mensuráveis? Quais seriam os limites e os riscos desse princípio?

6. A assimetria de literacia digital entre diferentes estratos da população brasileira coloca em questão a premissa de igualdade do voto em sentido substantivo? É possível, e desejável, que o direito eleitoral incorpore essa dimensão como critério de validade do processo eleitoral?

7. Diante da velocidade da evolução tecnológica e do problema das janelas de vulnerabilidade identificado neste capítulo, qual modelo de produção normativa, legislativa, regulatória ou jurisprudencial, está mais bem equipado para responder aos desafios da IA eleitoral em tempo útil?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas?

1. A urna é segura. O caminho até ela não é.

Se você trabalha com campanha eleitoral, já sabe que ninguém frauda a urna eletrônica brasileira. O problema está antes disso. A batalha eleitoral decisiva hoje não acontece na seção de votação; acontece no feed do eleitor às 22h, no grupo de WhatsApp da família, no vídeo recomendado automaticamente antes de dormir. Uma campanha que protege apenas o voto, e ignora o ambiente em que a opinião é formada, está protegendo o resultado de uma escolha que já foi feita sob influência que ela não monitorou.

2. Desinformação sobre você pode circular antes que você saiba que existe.

Em 2022, candidatos descobriram que vídeos falsos com suas imagens já tinham sido compartilhados milhares de vezes antes de chegarem ao seu conhecimento. Quando a equipe jurídica acionou o TSE e o conteúdo foi removido, a narrativa já havia se instalado. A lição prática: monitoramento do ambiente digital não é luxo de campanha grande. É infraestrutura mínima para qualquer candidatura que dispute em ambiente com penetração significativa de redes sociais e WhatsApp.

3. O WhatsApp é o campo de batalha que ninguém consegue fiscalizar completamente.

O CIEDDE monitora plataformas abertas com eficiência crescente. O WhatsApp é diferente: as mensagens circulam em grupos fechados, criptografados, invisíveis para qualquer sistema de monitoramento externo. Isso significa que campanhas de desinformação que operam exclusivamente pelo WhatsApp têm, na prática, menor risco de detecção e remoção do que aquelas que operam no Facebook ou no YouTube. Profissionais de marketing político precisam entender que esse é o ambiente de menor fiscalização e maior velocidade de circulação simultânea.

4. Rotular IA não é só obrigação legal. É gestão de reputação.

A Resolução TSE nº 23.732/2024 exige que conteúdo gerado ou significativamente modificado por IA seja rotulado. Além da obrigação jurídica, há uma dimensão estratégica: campanhas que usam IA sem transparência e são expostas por isso enfrentam um custo reputacional que tende a superar qualquer ganho que o conteúdo não rotulado poderia ter gerado. A transparência sobre o uso de IA, quando bem comunicada, pode inclusive ser posicionada como diferencial de honestidade em um ambiente de desconfiança generalizada.

5. Seu eleitor mais vulnerável à desinformação pode ser exatamente o eleitor que você mais precisa alcançar.

A assimetria de literacia digital no Brasil não é geograficamente neutra. Os segmentos com menor capacidade de verificar informações são, com frequência, os mesmos que campanhas disputam com maior intensidade em regiões de menor desenvolvimento socioeconômico. Isso cria uma responsabilidade ética concreta: campanhas que exploram essa vulnerabilidade para disseminar narrativas distorcidas não estão apenas violando a lei eleitoral — estão aprofundando uma desigualdade estrutural que corrói a qualidade da democracia que as elegeu.

6. A janela entre as eleições de 2024 e 2026 é uma janela de risco.

A Resolução TSE nº 23.732/2024 regula as tecnologias de IA disponíveis até o início de 2024. Desde então, novos modelos foram lançados com capacidades superiores. O período entre agora e as eleições de 2026 é precisamente o intervalo em que tecnologias ainda não cobertas pela regulação vigente podem ser utilizadas com menor risco jurídico. Profissionais que assessoram campanhas precisam conhecer não apenas o que a regulação proíbe hoje, mas o que ela ainda não alcança — tanto para evitar exposição quanto para identificar riscos do adversário.

CAPÍTULO 4

A máquina de persuadir

Microtargeting, Psicologia Comportamental e os Limites Éticos e Jurídicos da Personalização Eleitoral

DO MARKETING DE MASSA AO MARKETING DE PRECISÃO ELEITORAL

A convergência entre inteligência artificial e política eleitoral representa um dos fenômenos mais disruptivos do pensamento democrático contemporâneo. Ao longo das últimas duas décadas, o que era um campo restrito a consultores políticos e agências de publicidade tornou-se uma infraestrutura técnica de escala industrial, capaz de identificar, segmentar, persuadir e mobilizar eleitores com precisão sem precedente histórico.

A história do marketing político é, em larga medida, a história de uma tensão irresolvida entre dois imperativos concorrentes. O primeiro é o imperativo comunicativo, que exige que a mensagem alcance o maior número possível de pessoas. O segundo é o imperativo persuasivo, que exige que a mensagem seja suficientemente relevante para cada receptor individual a ponto de modificar seu comportamento eleitoral. Durante décadas, esses dois objetivos se excluíam mutuamente: quanto mais massiva a comunicação, menos personalizada; quanto mais personalizada, menos escalável.

A tecnologia digital, e especialmente a inteligência artificial aplicada ao comportamento do consumidor, transformou essa tensão ao tornar tecnicamente possível o que antes era logicamente contraditório: uma comunicação ao mesmo tempo massiva e profundamente individualizada.

O campo de estudos sobre persuasão política tem raízes que remontam ao trabalho pioneiro de Lazarsfeld, Berelson e Gaudet, publicado em 1944 sob o título *The People's Choice*. A pesquisa identificou o papel dos líderes de opinião na mediação entre as mensagens de campanha e os eleitores comuns, e produziu uma descoberta que contrariava as expectativas da época: a comunicação política de massa raramente converte eleitores diretamente. Ela opera, predominantemente, por reforço das predisposições existentes, e seu impacto mais visível recai sobre a mobilização da base, não sobre a conversão de adversários.

Essa constatação, que parecia impor um limite estrutural ao alcance do marketing político, foi progressivamente reinterpretada pela literatura posterior como ponto de partida

para uma estratégia mais sofisticada. Se a conversão direta é rara e custosa, a solução é concentrar recursos na identificação e mobilização dos eleitores já predispostos mas insuficientemente motivados a votar. É precisamente essa lógica que o microtargeting viabilizou em escala industrial, ao permitir que campanhas identifiquem, com precisão estatística, quem são esses eleitores, onde estão e quais mensagens têm maior probabilidade de ativá-los.

A literatura contemporânea sobre o tema organiza-se em torno de contribuições complementares. No plano descritivo e estratégico, a obra de Issenberg, *A Máquina da Vitória*, constitui a referência mais acessível ao público amplo, enquanto os experimentos de Nickerson e Rogers oferecem a base empírica mais rigorosa sobre os efeitos mensuráveis das intervenções comportamentais em campanhas eleitorais. No plano da teoria crítica, Zuboff, Zittrain e Mayer-Schönberger e Cukier constroem enquadramentos normativos que problematizam as implicações democráticas do modelo, questionando se a personalização em escala industrial é compatível com os princípios de autonomia e igualdade que sustentam o processo eleitoral.

Importante se faz ressaltar que a campanha eleitoral contemporânea opera em um ecossistema radicalmente distinto daquele que moldou as teorias clássicas de comunicação política. Durante a maior parte do século XX, as campanhas se organizavam em torno de comícios, jingles radiofônicos, inserções televisivas e cabos eleitorais presenciais: todos instrumentos de comunicação de massa, por natureza indiferenciada, que entregavam a mesma mensagem a públicos amplos e heterogêneos. O século XXI produziu uma inversão fundamental nessa lógica. A comunicação deixou de ser de um para muitos e tornou-se, potencialmente, de um para cada um. Cada eleitor pode, em teoria, receber uma mensagem diferente, calibrada especificamente para suas preferências, medos, valores e vulnerabilidades.

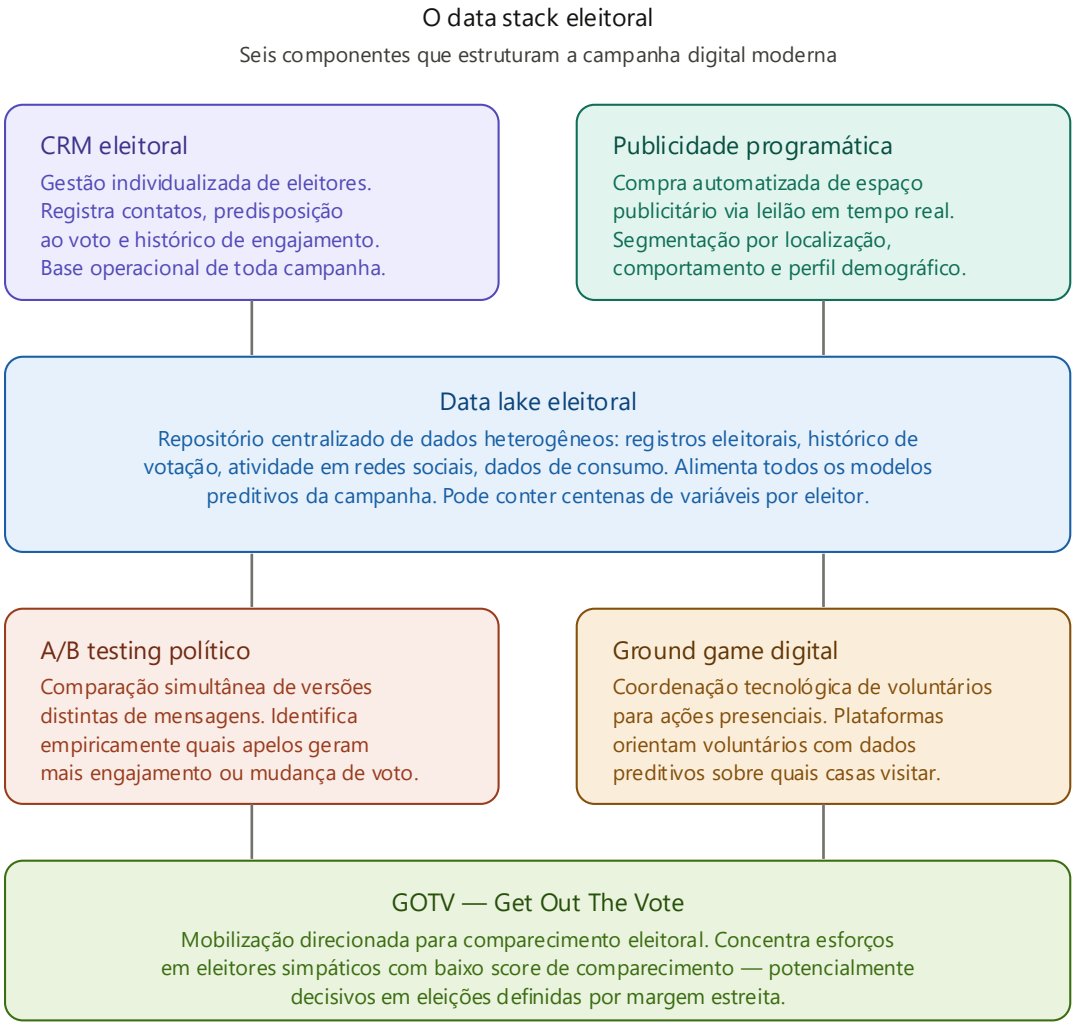
Essa transformação não é apenas quantitativa, uma questão de alcance ou velocidade. É qualitativa: ela altera a própria natureza da relação entre campanha e eleitor, substituindo o apelo coletivo pela interpelação individualizada. E tem uma arquitetura técnica identificável. O coração da campanha digital moderna é o que os profissionais da área denominam *data stack* eleitoral: um conjunto integrado de sistemas de coleta, armazenamento, processamento e análise de dados que alimenta cada decisão estratégica da campanha, desde a alocação de verbas publicitárias até a definição de quais eleitores merecem a visita pessoal de um voluntário.

A campanha digital moderna não é um conjunto de ferramentas isoladas. É um sistema integrado, no qual cada componente alimenta os demais e nenhuma decisão

estratégica relevante ocorre fora do fluxo de dados. Compreender essa arquitetura é condição para compreender como a inteligência artificial opera concretamente no processo eleitoral: não como um recurso adicional, mas como a lógica que conecta e otimiza cada uma das camadas.

O diagrama a seguir organiza os seis componentes centrais desse sistema, da coleta e modelagem de dados até a mobilização final para o comparecimento eleitoral. A leitura de cima para baixo revela não apenas uma sequência técnica, mas uma cadeia estratégica: cada camada depende da anterior e prepara a seguinte. É nessa arquitetura invisível que a eleição contemporânea, em grande medida, se decide.

Imagem 01. Data Stack Eleitoral



Fonte: Elaborado pelo autor com base em Norris (2014), AI Act (2024).

É importante compreender que essas ferramentas não operam de forma isolada. Elas formam um pipeline de inteligência eleitoral no qual dados brutos são progressivamente transformados em decisões estratégicas. Um eleitor que interage com um post de campanha no Facebook gera dados que alimentam o modelo preditivo, que atualiza seu score de persuabilidade, que define se ele receberá publicidade adicional, que por sua vez gera novos dados de comportamento, um ciclo de retroalimentação que, ao longo de meses de campanha, torna os modelos progressivamente mais precisos.

■ **Reflexão**
A questão que se coloca para o direito eleitoral é: em que medida esse ciclo de retroalimentação de dados é compatível com os princípios da igualdade de oportunidades entre candidatos e da autonomia do eleitor? Se uma campanha bem financiada tem acesso a modelos preditivos muito mais precisos que uma campanha com recursos limitados, não estamos diante de uma nova forma de captura do processo eleitoral pelo poder econômico desta vez mediada por algoritmos em vez de cartazes e cabos eleitorais?

MICROTARGETING: CONCEITO, EVOLUÇÃO HISTÓRICA E ARQUITETURA TÉCNICA

O microtargeting eleitoral é, provavelmente, a aplicação mais poderosa e mais controversa da análise de dados em contexto político. Diferentemente da segmentação demográfica tradicional, que agrupa eleitores por características observáveis como faixa etária, gênero, escolaridade, renda e localização geográfica, o microtargeting aspira a algo substancialmente mais ambicioso: identificar o perfil psicológico individual do eleitor e, com base nesse perfil, prever não apenas como ele votará, mas quais argumentos têm maior probabilidade de alterar sua intenção de voto.

A distinção é mais do que técnica. A segmentação demográfica trata o eleitor como membro de um grupo; o microtargeting trata-o como indivíduo singular, com disposições, vulnerabilidades e pontos de inflexão próprios. Essa mudança de escala é o que torna o microtargeting ao mesmo tempo tão eficaz como instrumento de campanha e tão perturbador como fenômeno democrático.

A evolução histórica do microtargeting pode ser periodizada em quatro fases, cada uma marcada por uma ruptura tecnológica que redefine as possibilidades estratégicas disponíveis às campanhas.

A primeira fase, de 1994 a 2004, opera com as mesmas ferramentas do marketing comercial de massa: listas de endereços, pesquisas de opinião agregadas e inserções

publicitárias em mídia tradicional. A segmentação é grosseira e os dados, essencialmente demográficos.

A segunda fase, de 2004 a 2012, incorpora dados comportamentais digitais e introduz o targeting psicográfico inicial. A campanha de reeleição de Barack Obama em 2012 tornou-se o caso de referência dessa transição, ao demonstrar que a combinação entre modelagem preditiva, dados de redes sociais e experimentação comportamental em larga escala produzia vantagens mensuráveis em eleições competitivas.

A terceira fase, de 2012 a 2018, consolida a integração entre redes sociais, modelagem de personalidade e inteligência artificial para targeting em escala individual. É nesse período que o escândalo da Cambridge Analytica expõe, para o grande público, a profundidade das inferências psicográficas que se tornaram tecnicamente possíveis e politicamente operacionalizadas.

A quarta fase, iniciada em 2018 e em curso ao tempo da redação desta obra, representa uma ruptura qualitativa em relação às anteriores. A disponibilização comercial de modelos de linguagem de grande escala introduz o que pode ser denominado personalização generativa: não apenas a entrega de mensagens diferentes para perfis diferentes, mas a geração automática de conteúdo singular para cada receptor.

O que muda não é apenas a precisão do targeting, mas sua natureza: a mensagem política deixa de ser um produto pré-fabricado distribuído de forma segmentada e passa a ser, em teoria, um artefato produzido em tempo real para um único destinatário, com base em seu perfil psicológico, histórico de interações e contexto situacional.

Tabela 18. Periodização do Microtargeting Eleitoral

Fase	Período	Dados Utilizados	Tecnologia Central	Caso Referência
I — Segmentação demográfica	1994-2004	Dados demográficos, registros públicos, pesquisas	Bancos de dados relacionais, GIS	Campanha Clinton 1996
II — Targeting comportamental	2004-2012	Dados de consumo, comportamento digital inicial	Data warehousing, regressão logística	Obama 2008
III — IA e personalidade	2012-2018	Redes sociais, psicografia, big data	Machine Learning, NLP, modelo OCEAN	Obama 2012, Trump 2016

IV — IA Generativa	2018-presente	Dados multimodais, inferência comportamental	LLMs, deep learning, ad tech avançada	Diversas campanhas 2022-2024

Fonte: Elaborado pelo autor com base em Issenberg (2012), Wylie (2019) e Matejic (2018).

A BASE DE DADOS: COLETA, INTEGRAÇÃO E O PERFIL 360° DO ELEITOR

A eficácia do microtargeting é diretamente proporcional à qualidade e à abrangência da base de dados que o alimenta. A base de dados de uma campanha contemporânea de grande porte é construída a partir da integração de pelo menos cinco categorias de fontes: dados de identidade extraídos do cadastro eleitoral do TSE; dados demográficos enriquecidos por registros públicos e informações do IBGE; dados comportamentais digitais, como interações em redes sociais, histórico de navegação e geolocalização; dados de consumo, incluindo histórico de compras e programas de fidelidade; e dados psicográficos inferidos algoritmicamente a partir das categorias anteriores, frequentemente organizados segundo o modelo OCEAN.

A integração dessas fontes por algoritmos de desduplicação e enriquecimento produz o que os profissionais denominam perfil 360 graus do eleitor: uma representação multidimensional que nenhuma pesquisa eleitoral tradicional seria capaz de capturar com equivalente granularidade. É a existência desse perfil que torna possível o microtargeting de alta precisão. E é precisamente essa existência que levanta as questões jurídicas mais sérias em relação à conformidade com a LGPD: o perfil 360 graus é, na sua essência, um agregado de dados pessoais sensíveis tratados sem que o titular tenha, na maioria dos casos, qualquer consciência disso.

LGPD — Art. 5º, II: Dado Sensível

Para os fins desta Lei, considera-se dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural.

Brasil. Lei nº 13.709/2018 (LGPD), art. 5º, II.

A qualificação das opiniões políticas como dados sensíveis tem implicação direta e grave para o microtargeting eleitoral. Qualquer sistema que utilize dados sobre preferências políticas de eleitores, incluindo dados inferidos a partir de comportamento digital como

curtidas em publicações de determinado candidato, está tratando dados sensíveis na acepção do artigo 5º, II, da LGPD. Isso exige, na ausência de outra hipótese legal aplicável, o consentimento específico e destacado do titular, conforme o artigo 11, I, da mesma lei.

A consequência prática desse enquadramento é significativa. A maioria das campanhas eleitorais que realizam targeting psicográfico de alta precisão no Brasil opera sem esse consentimento, o que as expõe a sanções administrativas da Autoridade Nacional de Proteção de Dados. Mais grave, porém, é a dimensão eleitoral da irregularidade: quando o tratamento ilícito de dados sensíveis integra uma estratégia de campanha que influencia a formação de preferências eleitorais, abre-se espaço para representação por abuso de poder perante a Justiça Eleitoral, com consequências que podem alcançar a cassação do diploma.

Há, portanto, uma dupla exposição jurídica. A primeira, no plano do direito à proteção de dados, perante a ANPD. A segunda, no plano do direito eleitoral, perante o TSE e os Tribunais Regionais Eleitorais. Essas duas esferas de responsabilização operam de forma independente e cumulativa, o que significa que a mesma conduta pode ser simultaneamente objeto de processo administrativo sancionador e de ação eleitoral com efeitos sobre a validade do mandato obtido.

O MODELO PREDITIVO: COMO A IA ANTECIPA O COMPORTAMENTO ELEITORAL

A sofisticação técnica que distingue o microtargeting contemporâneo das técnicas de segmentação da geração anterior reside no modelo preditivo: o conjunto de algoritmos que processa a base de dados para gerar predições sobre o comportamento individual de cada eleitor. Na prática de campanha, quatro modelos preditivos são centrais: o modelo de propensão ao voto, que estima a probabilidade de comparecimento; o modelo de preferência, que estima a probabilidade de voto em cada candidato; o modelo de persuabilidade, que estima a probabilidade de mudança de comportamento em resposta à intervenção da campanha; e o modelo de mensagem ótima, que identifica qual tipo de mensagem é mais provável de produzir o comportamento eleitoral desejado em cada eleitor específico.

A contribuição da inteligência artificial a essa teoria clássica é a capacidade de estimar o persuasion score para cada eleitor individualmente, com base em seu perfil psicográfico e comportamental, em vez de atribuí-lo a grupos demográficos grosseiros. Isso representa uma mudança qualitativa na eficiência da alocação de recursos: em vez de campanhar igualmente em todas as circunscrições de um estado competitivo, a campanha

concentra seus recursos nos blocos residenciais, nos horários e nos canais em que os eleitores com maior *persuasion score* estão mais receptivos.

PSICOLOGIA COMPORTAMENTAL E O NUDGE ELEITORAL: CIALDINI, KAHNEMAN E THALER

A integração entre microtargeting e psicologia comportamental representa um dos desenvolvimentos mais significativos e mais problemáticos do marketing político contemporâneo. Três autores são centrais para compreender essa integração: Robert Cialdini, Daniel Kahneman e Richard Thaler. Juntos, eles oferecem não apenas ferramentas analíticas distintas, mas os componentes de uma teoria unificada da persuasão eleitoral contemporânea.

Cialdini, em *Influence: Science and Practice*, identificou seis princípios universais de influência social: reciprocidade, compromisso e consistência, prova social, autoridade, afinidade e escassez. No contexto eleitoral, cada princípio encontra aplicações específicas. A prova social fundamenta o uso de mensagens sobre o comportamento de vizinhos e pares, explorando a tendência humana de calibrar suas escolhas pelo que percebe como norma do grupo.

A autoridade orienta a estratégia de endorsements, ao ancorar a credibilidade do candidato em figuras de referência reconhecidas pelo eleitor-alvo. A afinidade sustenta o targeting personalizado que cria a percepção de que o candidato compreende as particularidades de cada indivíduo, quando na verdade responde a inferências estatísticas sobre seu perfil.

Kahneman, em *Rápido e Devagar*, oferece o enquadramento teórico mais influente para compreender por que as pessoas tomam decisões que contradizem seus próprios interesses declarados. A distinção entre o Sistema 1, pensamento rápido, intuitivo e emocional, e o Sistema 2, pensamento lento, deliberativo e racional, fundamenta a estratégia de comunicação eleitoral que apela a emoções e intuições em lugar de argumentos. Como observa o autor, nossas intuições sobre o mundo derivam de como o interpretamos, e essa interpretação é profundamente moldada pelo estado emocional, pelas associações passadas e pelo contexto imediato em que a informação é apresentada. Mudar o enquadramento muda a escolha, sem mudar a informação.

O microtargeting psicográfico é, nessa perspectiva, um sistema otimizado para identificar os gatilhos do Sistema 1 de cada eleitor e acioná-los no momento de maior suscetibilidade.

Thaler e Sunstein, em *Nudge*, desenvolveram o conceito de arquitetura de escolhas: a ideia de que o contexto em que as opções são apresentadas influencia sistematicamente as decisões, independentemente de seu conteúdo informacional. Aplicado ao ambiente eleitoral digital, o nudge descreve o conjunto de intervenções no design da experiência do usuário que orientam o comportamento sem coerção explícita, tornando determinadas escolhas mais prováveis simplesmente por alterarem as condições em que são feitas.

A articulação entre os três autores produz uma teoria da persuasão eleitoral de considerável poder explicativo. O microtargeting identifica os perfis psicológicos dos eleitores e mapeia seus pontos de suscetibilidade, na linha de Kahneman. Os princípios de influência de Cialdini orientam o design das mensagens para cada perfil, calibrando apelos de acordo com as disposições individuais identificadas.

O nudge digital de Thaler e Sunstein molda o contexto de recepção dessas mensagens para maximizar sua eficácia, intervindo não no conteúdo, mas nas condições em que o eleitor o processa. O resultado é uma arquitetura de persuasão que opera simultaneamente sobre o perfil do receptor, o conteúdo da mensagem e o contexto de sua recepção: três camadas de influência que se reforçam mutuamente e que, em conjunto, representam um grau de sofisticação persuasiva sem precedente na história do marketing político.

A base teórica mais influente dessa abordagem é o modelo psicométrico OCEAN, desenvolvido a partir da psicologia da personalidade. O modelo descreve a personalidade humana em cinco dimensões fundamentais — *Openness* (Abertura), *Conscientiousness* (Conscienciosidade), *Extraversion* (Extroversão), *Agreeableness* (Amabilidade) e *Neuroticism* (Neuroticismo) —, cada uma das quais correlaciona-se de forma previsível com preferências políticas, sensibilidade a diferentes tipos de mensagem e vulnerabilidade a determinados apelos emocionais.

Tabela 19. Modelo OCEAN: Predisposição Política e Apelo Eleitoral Mais Eficaz

Traço OCEAN	Perfil Alto	Predisposição Política	Apelo Eleitoral Mais Eficaz
Abertura (O)	Curioso, progressista, artístico, tolerante à ambiguidade	Partidos progressistas, pautas de inovação e diversidade	Mudança, novidade, futuro, multiculturalismo
Conscienciosidade (C)	Organizado, disciplinado, orientado a regras, responsável	Partidos de centro-direita, conservadorismo fiscal	Ordem, previsibilidade, responsabilidade, tradição

Extroversão (E)	Sociável, assertivo, otimista, energético	Variável: responde a candidatos carismáticos	Entusiasmo coletivo, movimento, comunidade, liderança forte
Amabilidade (A)	Cooperativo, empático, confiante nos outros	Pautas de bem-estar social, solidariedade	Coesão social, cuidado, harmonia, cooperação
Neuroticismo (N)	Ansioso, sensível a ameaças, instável emocionalmente	Vulnerável a discursos de medo e proteção	Segurança, proteção da família, ameaça do outro, estabilidade

Fonte: Adaptado de Kosinski, Stillwell e Graepel (2013); McCrae e Costa (2008)

A relevância do modelo OCEAN para o contexto eleitoral reside em sua capacidade de traduzir traços psicológicos abstratos em variáveis operacionais para estratégias de comunicação. Eleitores com altos níveis de neuroticismo, por exemplo, tendem a responder de forma mais intensa a mensagens baseadas em ameaça ou insegurança, enquanto indivíduos com alta abertura a experiências podem ser mais receptivos a discursos de mudança e inovação. De forma semelhante, níveis elevados de conscienciosidade podem estar associados a maior sensibilidade a apelos de ordem, estabilidade e responsabilidade institucional.

Essa capacidade de associação entre traços de personalidade e estratégias comunicacionais permite que campanhas eleitorais desenvolvam mensagens altamente personalizadas, ajustadas não apenas a características demográficas ou ideológicas, mas a disposições psicológicas profundas dos eleitores. Nesse sentido, o modelo OCEAN funciona como um elo entre a coleta de dados comportamentais e a operacionalização de estratégias de persuasão, consolidando a base científica da campanha algorítmica contemporânea.

O aspecto mais perturbador da aplicação do modelo OCEAN em campanhas eleitorais é a sua capacidade de prever traços de personalidade a partir de dados digitais aparentemente banais. A pesquisa pioneira de Michal Kosinski, David Stillwell e Thore Graepel (2013), publicada na PNAS, demonstrou que curtidas no Facebook são suficientes para predizer, com acurácia estatisticamente significativa, orientação sexual, etnia, religiosidade, preferências políticas e, crucialmente, os cinco traços do modelo OCEAN. Com apenas 10 curtidas, o modelo superava a precisão de colegas de trabalho; com 70 curtidas, superava amigos; com 150 curtidas, superava cônjuges.

■ **Ponto crítico — Inferência de dados sensíveis**

A LGPD brasileira (Lei 13.709/2018) distingue dados pessoais comuns de dados pessoais sensíveis, estes últimos incluindo, expressamente, "opinião política" (art. 5º, II). O tratamento de dados sensíveis requer consentimento específico e destacado, ou enquadramento em hipóteses legais taxativas. O problema do microtargeting baseado no modelo OCEAN é que ele infere dados sensíveis (opinião política, orientação sexual, religião) a partir de dados aparentemente não sensíveis (curtidas, histórico de navegação). A questão jurídica ainda não resolvida: o dado inferido tem o mesmo status legal do dado declarado? A coleta do dado-fonte para fins de inferência de dado sensível exige o mesmo nível de consentimento?

Para além das questões jurídicas, o microtargeting coloca um problema filosófico central para a teoria democrática, que diz respeito à distinção entre persuasão e manipulação. A tradição liberal clássica, representada por John Stuart Mill e, em formulação contemporânea, por John Rawls, parte do pressuposto de que a deliberação democrática se estrutura como um processo de troca pública de razões. Nesse modelo, cidadãos livres e iguais apresentam argumentos acessíveis, submetem-nos ao escrutínio coletivo e podem revisar suas posições à luz de justificativas consideradas razoáveis.

O microtargeting, tal como operado no ambiente digital contemporâneo, desloca esse modelo ao atuar em um registro distinto daquele da racionalidade deliberativa. Em vez de oferecer razões suscetíveis de avaliação consciente, ele se apoia na identificação de disposições psicológicas individuais e na ativação de respostas emocionais específicas, calibradas a partir de dados comportamentais. A comunicação política deixa de ser orientada pela força do melhor argumento e passa a ser estruturada pela eficácia do estímulo mais adequado ao perfil do receptor.

Essa mudança pode ser interpretada, no âmbito da filosofia moral, como uma transição da persuasão para a manipulação. Enquanto a persuasão pressupõe a capacidade do indivíduo de avaliar criticamente as razões apresentadas, preservando sua autonomia decisória, a manipulação opera por meios que contornam ou exploram limitações dessa capacidade, influenciando o comportamento sem engajamento reflexivo pleno. O ponto central não é a presença de influência, que é inerente a qualquer processo comunicativo, mas a forma como essa influência é exercida e o grau em que respeita, ou compromete, a autonomia do agente.

Nesse sentido, a utilização de dados psicográficos para calibrar mensagens políticas levanta a hipótese de uma assimetria epistêmica estrutural entre emissores e receptores. As campanhas passam a dispor de conhecimento aprofundado sobre as vulnerabilidades

cognitivas e emocionais dos eleitores, enquanto estes permanecem, em grande medida, alheios aos mecanismos pelos quais são alvo de influência. Essa assimetria compromete a ideia de igualdade política, uma vez que a capacidade de influenciar o processo deliberativo torna-se desigualmente distribuída entre atores com diferentes níveis de acesso a dados e tecnologia.

A operacionalização dessa arquitetura de influência depende de uma infraestrutura complexa de coleta, agregação e modelagem de dados. A construção de perfis psicográficos não se baseia em uma única fonte de informação, mas na integração de múltiplos fluxos de dados que, combinados, permitem inferências altamente detalhadas sobre o comportamento e as preferências dos indivíduos.

Entre as principais fontes utilizadas nesse processo, destacam-se os dados provenientes de registros eleitorais, disponibilizados por instituições como o Tribunal Superior Eleitoral, que incluem informações básicas de identificação e localização dos eleitores, além de dados sobre comparecimento às urnas em determinados contextos. Esses dados, embora públicos em sua origem, adquirem novo significado quando integrados a outras bases informacionais.

A atividade em redes sociais constitui outra fonte central, abrangendo interações como curtidas, compartilhamentos, comentários, páginas seguidas e tempo de exposição a conteúdos. Esses dados permitem inferir preferências ideológicas, interesses temáticos e padrões de engajamento, funcionando como indicadores indiretos de traços de personalidade e posicionamentos políticos.

Os dados de consumo digital ampliam ainda mais essa capacidade analítica, incluindo histórico de navegação, compras online, assinaturas de serviços e uso de aplicativos. Tais informações revelam hábitos de consumo, padrões socioeconômicos e estilos de vida, que podem ser correlacionados com comportamentos políticos.

A dimensão geoespacial adiciona um nível adicional de granularidade, ao incorporar dados de localização obtidos por dispositivos móveis. Padrões de mobilidade, frequência a determinados espaços e rotinas diárias permitem inferir aspectos como pertencimento religioso, inserção social e redes de sociabilidade.

A atuação de empresas especializadas na agregação e comercialização de dados pessoais, conhecidas como data brokers, intensifica esse processo ao consolidar informações provenientes de múltiplas fontes, incluindo programas de fidelidade, cadastros comerciais e pesquisas de opinião. Esses agentes operam como intermediários centrais na economia de dados, ampliando significativamente o alcance das bases informacionais disponíveis para campanhas políticas.

Por fim, modelos de inferência preditiva permitem extrapolar variáveis não diretamente observáveis, como renda, orientação política ou grau de religiosidade, a partir de proxies comportamentais. Esses modelos transformam dados fragmentados em perfis coerentes, possibilitando a segmentação e o direcionamento de mensagens com alto grau de precisão.

A combinação dessas fontes e técnicas configura um sistema de conhecimento sobre o eleitor que ultrapassa, em muitos casos, a própria autoconsciência do indivíduo sobre suas preferências e predisposições. É nesse ponto que o problema filosófico da manipulação adquire sua máxima intensidade: quando a influência política se baseia em mecanismos que operam abaixo do limiar da consciência reflexiva, a própria ideia de escolha autônoma, que fundamenta a legitimidade democrática, torna-se objeto de questionamento.

Personalização de Mensagens via Algoritmos

A personalização algorítmica de mensagens eleitorais ultrapassa, de forma significativa, a simples adaptação temática do conteúdo aos interesses do eleitor. Trata-se de um processo multidimensional, no qual diferentes camadas de otimização operam de maneira simultânea e coordenada, cada uma orientada por modelos preditivos específicos e alimentada por conjuntos distintos de dados comportamentais.

Nesse contexto, a comunicação política deixa de ser concebida como uma mensagem unitária dirigida a um público amplo e passa a assumir a forma de um sistema dinâmico de variações controladas, no qual cada elemento da mensagem pode ser ajustado em função das características inferidas do receptor. O conteúdo que chega ao eleitor não é, portanto, uma versão fixa de um discurso político, mas o resultado de um processo contínuo de experimentação, segmentação e refinamento algorítmico.

Isso implica que a personalização não se limita ao tema abordado, mas abrange dimensões como o enquadramento narrativo, o tom emocional, o vocabulário utilizado, o formato do conteúdo, o momento de entrega e o canal de distribuição. Cada uma dessas dimensões é passível de otimização independente, com base em métricas de desempenho específicas, como taxa de engajamento, tempo de visualização ou probabilidade de conversão eleitoral.

O efeito agregado dessas múltiplas camadas de personalização é a construção de experiências comunicacionais profundamente individualizadas, nas quais diferentes eleitores são expostos a versões distintas, e por vezes contraditórias, da mesma

candidatura. A campanha deixa de operar no registro da comunicação pública compartilhada e passa a funcionar como um conjunto de interações privadas, calibradas para maximizar a eficácia persuasiva em nível individual.

Na prática, isso significa que dois eleitores geograficamente próximos, inclusive residentes no mesmo domicílio, podem ser alvo de estratégias comunicacionais completamente distintas, concebidas pela mesma equipe de campanha, sem qualquer visibilidade recíproca. Essa assimetria informacional não é um efeito colateral do sistema, mas uma de suas características estruturais, na medida em que a eficácia da personalização depende justamente da ausência de transparência e da impossibilidade de comparação entre as mensagens recebidas por diferentes indivíduos.

Essa fragmentação radical da comunicação política desafia pressupostos fundamentais da democracia representativa, especialmente a ideia de que o debate eleitoral ocorre em um espaço público minimamente compartilhado, no qual as propostas dos candidatos são expostas ao escrutínio coletivo. Ao substituir esse modelo por um sistema de comunicação individualizada e opaca, a personalização algorítmica redefine as condições de visibilidade, accountability e contestação no processo eleitoral.

A complexidade dessa arquitetura de personalização pode ser melhor compreendida por meio de sua decomposição analítica em dimensões operacionais distintas. Cada uma dessas dimensões corresponde a um eixo específico de intervenção algorítmica, no qual diferentes técnicas de modelagem e otimização são aplicadas com objetivos estratégicos próprios, ainda que interdependentes.

A sistematização dessas dimensões permite evidenciar que a personalização não é um processo unitário, mas um conjunto articulado de decisões automatizadas que incidem sobre múltiplos aspectos da comunicação política. O quadro a seguir sintetiza essas dimensões, detalhando o que é objeto de personalização, as tecnologias envolvidas e os objetivos estratégicos associados a cada uma delas.

Tabela 20. Arquitetura Multidimensional da Personalização Algorítmica em Campanhas Eleitorais

Dimensão	O que é personalizado	Tecnologia envolvida	Objetivo
Conteúdo temático	Assunto abordado (economia, segurança, saúde, educação)	Análise de *issue salience*, NLP, *topic modeling*	Gerar ressonância com as preocupações prioritárias do eleitor
Tom emocional	Enquadramento afetivo (esperança vs. medo, orgulho vs. indignação, unidade vs. ameaça)	Análise de sentimento, psicometria OCEAN	Ativar o estado emocional mais receptivo à mensagem

Formato e mídia	Tipo de conteúdo (texto curto, vídeo longo, infográfico, áudio, imagem)	Modelos de engajamento por formato e perfil	Maximizar taxa de abertura, visualização e compartilhamento
Canal de entrega	Plataforma de distribuição (Facebook, Instagram, WhatsApp, e-mail, SMS, YouTube)	Análise de comportamento *cross-platform*	Encontrar o eleitor no ambiente em que está mais receptivo
Timing	Momento de envio (hora do dia, dia da semana, proximidade da eleição)	Modelos de engajamento temporal	Entregar a mensagem no momento de maior receptividade cognitiva
Frequência e sequenciamento	Número e ordem das exposições à mensagem	Modelos de saturação e reforço de mensagem	Evitar saturação e otimizar a repetição para maximizar retenção e persuasão

Fonte: Elaborado pelo autor, com base em Daniel Kahneman (2011), Robert Cialdini (2009), Richard Thaler e Cass Sunstein (2008), Michal Kosinski, David Stillwell e Thore Graepel (2013), e literatura contemporânea sobre microtargeting e campanhas digitais.

A geração automatizada de variações de conteúdo — hoje potencializada por modelos de linguagem de grande escala como GPT-4 e Claude, representa a fronteira mais recente dessa personalização. Se anteriormente uma equipe humana produzia, talvez, 10 ou 20 variações de uma mesma mensagem para diferentes públicos, os sistemas atuais de IA generativa podem produzir centenas ou milhares de variações em minutos, cada uma ajustada a micropadrões linguísticos, referências culturais e registros comunicativos de subgrupos específicos do eleitorado.

■ **Implicação regulatória — Identificação de conteúdo gerado por IA**
A Resolução TSE nº 23.732/2024 exige que material eleitoral produzido com o auxílio de IA seja identificado como tal. O desafio prático é: quando se tem centenas de variações de uma mensagem, cada uma minimamente distinta, a obrigação de identificação se aplica a cada variação individualmente? Como os sistemas de moderação das plataformas detectam e processam essa escala de produção de conteúdo? Essas são perguntas para as quais a regulamentação ainda não tem resposta satisfatória.

Análise Preditiva de Comportamento Eleitoral

A análise preditiva constitui o núcleo epistemológico da campanha orientada por dados. Trata-se do mecanismo pelo qual informações históricas, relativas ao comportamento passado dos eleitores, são convertidas em estimativas probabilísticas sobre comportamentos futuros, permitindo a antecipação de decisões eleitorais em nível individual. Esse deslocamento, da descrição para a predição, altera qualitativamente a lógica da estratégia política, que passa a operar não apenas com diagnósticos do presente, mas com projeções estatisticamente fundamentadas do futuro.

Esse exercício preditivo estrutura, de forma decisiva, quatro grandes domínios de decisão estratégica: a alocação territorial de recursos de campanha, a definição de

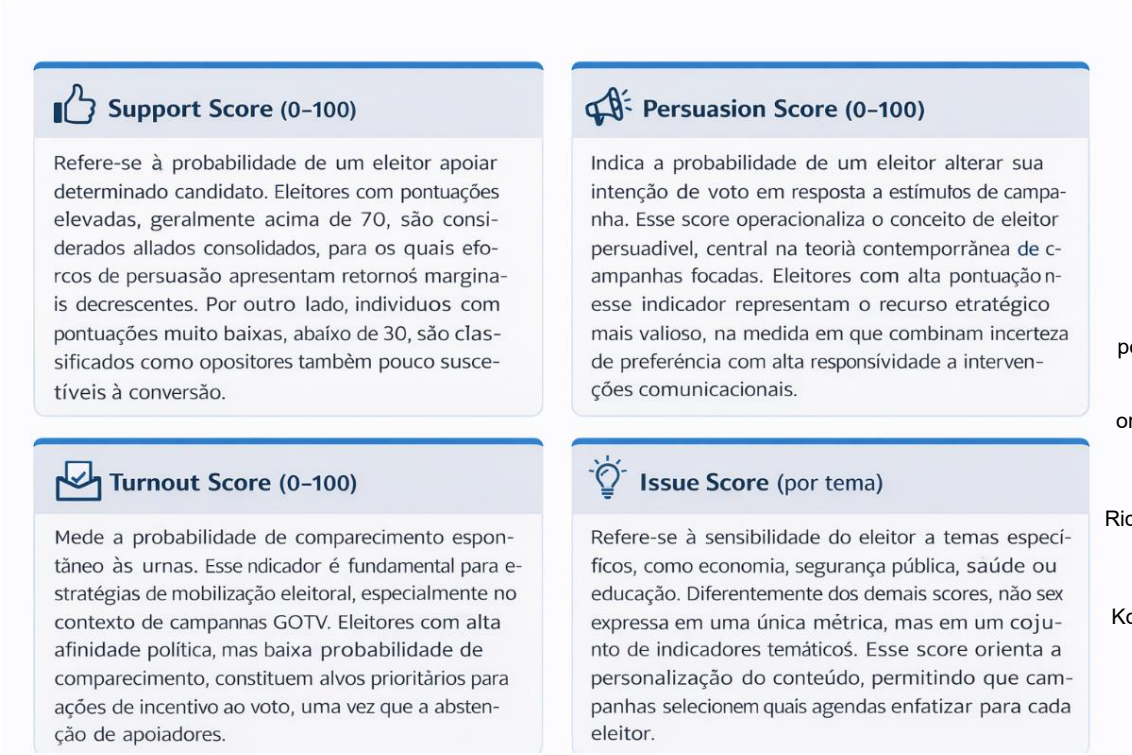
mensagens individualizadas, o planejamento de investimentos em publicidade paga e a priorização de esforços de mobilização eleitoral. Em todos esses casos, a incerteza inerente ao comportamento humano é tratada como variável modelável, cuja variabilidade pode ser reduzida por meio de inferência estatística e aprendizado de máquina.

Os modelos preditivos contemporâneos baseiam-se, em larga medida, em técnicas de ensemble learning, que combinam diferentes algoritmos com o objetivo de produzir estimativas mais estáveis e precisas. Métodos como árvores de decisão, regressão logística, redes neurais e gradient boosting são integrados em arquiteturas híbridas, nas quais as previsões individuais de cada modelo são agregadas por meio de mecanismos de ponderação. A principal vantagem dessa abordagem reside na capacidade de reduzir a variância dos modelos, mitigando o risco de sobreajuste aos dados de treinamento e aumentando a capacidade de generalização para novos contextos eleitorais.

No plano operacional, essas capacidades preditivas são traduzidas em scores eleitorais, que sintetizam, em métricas numéricas, diferentes dimensões do comportamento político dos eleitores. Esses scores funcionam como instrumentos de decisão, permitindo que campanhas priorizem recursos com base em probabilidades estimadas, e não apenas em intuições ou segmentações demográficas tradicionais.

A tipologia mais consolidada desses indicadores compreende quatro classes principais:

Imagem 01.Tipologia dos Scores Preditivos em Campanhas Eleitorais Orientadas por Dados



Fonte: Elaborado pelo autor, com base em literatura sobre campanhas orientadas por dados e microtargeting, incluindo Daniel Kahneman (2011), Richard Thaler e Cass Sunstein (2008), Robert Cialdini (2009), Michal Kosinski et al. (2013), e estudos contemporâneos sobre modelagem preditiva eleitoral.

A utilização combinada desses indicadores permite a construção de estratégias altamente diferenciadas, nas quais cada eleitor é posicionado em um espaço multidimensional de probabilidades. A campanha deixa de operar com categorias amplas e passa a atuar com granularidade individual, ajustando suas intervenções de forma contínua com base em dados atualizados.

Essa capacidade de previsão e segmentação redefine a própria natureza da competição eleitoral. O voto deixa de ser apenas uma expressão de preferência política e passa a ser tratado como uma variável probabilística a ser otimizada. Nesse contexto, a eficácia da campanha não depende apenas da qualidade de suas propostas, mas da precisão com que consegue identificar, influenciar e mobilizar segmentos específicos do eleitorado.

A sofisticação contemporânea dos modelos preditivos alcançou um nível em que campanhas eleitoralmente competitivas operam com verdadeiros sistemas de simulação estratégica em larga escala. Na fase de modelagem, essas campanhas realizam, de forma rotineira, simulações computacionais de milhares de cenários eleitorais alternativos, nos quais variáveis centrais do processo eleitoral são sistematicamente alteradas para estimar seus efeitos sobre o resultado final.

Essas simulações não se limitam a projeções lineares de intenção de voto, mas incorporam múltiplas dimensões de incerteza, incluindo variações nas taxas de comparecimento entre diferentes segmentos demográficos, impactos potenciais de eventos exógenos, como crises econômicas, escândalos políticos ou acontecimentos inesperados, e a eficácia marginal de intervenções de campanha, como publicidade segmentada, mobilização territorial ou ações digitais coordenadas. Trata-se, em termos metodológicos, de uma aplicação avançada de técnicas de simulação estocástica, frequentemente operacionalizadas por métodos como *Monte Carlo*, que permitem explorar distribuições de probabilidade para resultados eleitorais sob diferentes condições.

O objetivo dessas simulações não é prever um único resultado, mas mapear o espaço de possibilidades eleitorais, identificando cenários de maior risco e oportunidades estratégicas. A campanha passa, assim, a operar com uma lógica de gestão probabilística, na qual decisões são tomadas com base na maximização da probabilidade de vitória sob diferentes configurações do ambiente eleitoral.

Um dos casos mais emblemáticos desse nível de sofisticação foi a campanha de Barack Obama em 2012, que implementou uma infraestrutura analítica de grande escala para integrar dados, modelagem e execução estratégica. Durante os meses que

antecederam a eleição, a equipe realizou aproximadamente 66.000 simulações por noite, variando parâmetros críticos do comportamento eleitoral e das ações de campanha. Esse sistema, conhecido internamente como *Project Narwhal*, tornou-se um marco na profissionalização da análise de dados em campanhas políticas, ao demonstrar como a integração entre big data, modelagem preditiva e simulação computacional pode redefinir a forma como estratégias eleitorais são concebidas e executadas.

A operacionalização desses mecanismos cognitivos no contexto das campanhas eleitorais permite observar como princípios abstratos da psicologia comportamental são traduzidos em estratégias comunicacionais concretas. Longe de constituírem apenas achados experimentais em ambientes controlados, esses vieses são sistematicamente incorporados ao design de mensagens políticas, orientando tanto a formulação do conteúdo quanto sua forma de apresentação.

Nesse sentido, a literatura não apenas identifica a existência desses vieses, mas demonstra sua aplicabilidade direta em contextos eleitorais, nos quais decisões são frequentemente tomadas sob condições de incerteza, baixa informação e alta carga emocional. A eficácia das campanhas contemporâneas depende, em grande medida, da capacidade de ativar esses mecanismos de forma estratégica e segmentada.

O quadro a seguir sistematiza os principais vieses cognitivos mobilizados em campanhas eleitorais, explicitando seus mecanismos psicológicos subjacentes, suas aplicações práticas e exemplos concretos de utilização.

Tabela 21. Vieses Cognitivos Aplicados a Campanhas Eleitorais

Viés Cognitivo	Mecanismo Psicológico	Aplicação em Campanhas	Exemplo Concreto
Efeito de enquadramento (Framing)	A mesma informação produz decisões diferentes conforme o modo de apresentação.	Apresentar a mesma política como ganho positivo ou como perda evitada.	'Sua família estará mais segura' vs. 'Sem nossa proposta, sua família continuará em risco.'
Aversão à perda (Loss aversion)	Perdas têm impacto emocional 2x maior que ganhos equivalentes.	Mensagens de ameaça são mais mobilizadoras que mensagens de promessa.	'Tudo que você construiu pode ser perdido' converte mais doações que 'Imagine um futuro melhor.'
Ancoragem (Anchoring)	Primeira informação recebida ancora todas as avaliações subsequentes.	Primeiro número em pedido de doação determina valor médio da doação.	Sugerir R$ 500 como doação-padrão antes de oferecer R$ 100 como opção acessível.

Prova social (Social proof)	Comportamento de pessoas semelhantes valida a decisão individual.	Mostrar quantos já aderiram para criar sensação de movimento crescente.	'1,2 milhão de brasileiros já assinaram. Você também está dentro?'
Escassez (Scarcity)	Recursos percebidos como limitados têm valor subjetivo maior.	Criar senso de urgência temporal para mobilizar ação imediata.	'As próximas 4 horas são decisivas. Cada doação feita agora será dobrada.'
Identidade social (In-group/Out-group)	Pertencimento tribal orienta decisões mais que argumentos racionais.	Ativar identidades religiosas, regionais, étnicas ou de classe como âncora eleitoral.	Mensagens que constroem uma identidade coletiva ameaçada por um outro inimigo.
Mera exposição (Mere exposure)	Familiaridade produz afeição: quanto mais se vê, mais se aprecia.	Frequência de exposição publicitária aumenta avaliação positiva do candidato.	Publicidade de marca antes da comunicação de conteúdo.

Fonte: Compilado a partir de Kahneman (2011), Thaler e Sunstein (2008), Cialdini (1984)

ÉTICA E LIMITES JURÍDICOS: A FRONTEIRA ENTRE PERSUASÃO LEGÍTIMA E MANIPULAÇÃO

A distinção entre persuasão política legítima e manipulação eleitoral ilícita constitui um dos problemas mais complexos e ainda não resolvidos do direito eleitoral contemporâneo. A dificuldade analítica decorre do fato de que a persuasão política, mesmo em suas formas mais tradicionais, sempre incorporou elementos de apelo emocional, exploração de heurísticas cognitivas e adaptação da mensagem a diferentes públicos, sem que tais práticas fossem, por si só, consideradas juridicamente ilícitas ou normativamente incompatíveis com a democracia.

O que o microtargeting psicográfico introduz nesse cenário não é a existência dessas práticas, mas sua reconfiguração em termos de escala, sistematicidade e precisão. A utilização de dados comportamentais massivos, combinada com modelos preditivos e sistemas de distribuição automatizada, permite que estratégias de influência sejam aplicadas de forma individualizada, contínua e praticamente invisível. A ausência de transparência sobre os critérios de segmentação, o conteúdo das mensagens e os mecanismos de entrega agrava esse quadro, dificultando tanto o controle social quanto a fiscalização institucional.

Nesse contexto, um critério doutrinário particularmente relevante para a distinção entre persuasão e manipulação pode ser formulado a partir da literatura filosófica sobre

autonomia, especialmente nas tradições liberais e deliberativas. Trata-se do critério da agência reflexiva, segundo o qual a legitimidade da comunicação política depende de sua capacidade de engajar o eleitor como agente racional, capaz de avaliar razões, ponderar argumentos e revisar suas preferências à luz de justificativas públicas.

A persuasão política legítima, sob esse prisma, é aquela que se dirige à dimensão reflexiva do eleitor, oferecendo informações e argumentos que podem ser submetidos ao escrutínio crítico. Ainda que mobilize elementos emocionais, ela não prescinde da possibilidade de avaliação consciente por parte do destinatário. A manipulação eleitoral, por sua vez, caracteriza-se pela atuação sobre mecanismos não reflexivos do comportamento, explorando vieses cognitivos e vulnerabilidades psicológicas de modo a induzir decisões que não seriam endossadas pelo próprio eleitor em condições de plena consciência e deliberação informada.

Essa distinção, embora conceitualmente robusta, apresenta elevada dificuldade de operacionalização no plano jurídico. O direito eleitoral, tradicionalmente estruturado em torno de categorias como abuso de poder econômico, uso indevido dos meios de comunicação e propaganda irregular, não dispõe de instrumentos plenamente adequados para capturar formas de influência que operam em nível microindividual, distribuídas em redes digitais e mediadas por algoritmos opacos.

A Tribunal Superior Eleitoral, por meio da Resolução nº 23.732/2024, representa um avanço relevante ao introduzir obrigações de transparência e rotulagem de conteúdo, especialmente no que se refere ao uso de inteligência artificial. No entanto, tais medidas atuam predominantemente sobre os efeitos visíveis da comunicação eleitoral, sem alcançar integralmente os processos subjacentes de segmentação, modelagem e personalização que estruturam a influência algorítmica.

Permanece, portanto, uma lacuna normativa significativa entre a sofisticação das práticas de campanha e a capacidade regulatória do ordenamento jurídico. Essa lacuna não é apenas técnica, mas conceitual: enquanto o direito opera com categorias pensadas para um ambiente comunicacional público e relativamente transparente, a campanha algorítmica se desenvolve em um espaço fragmentado, opaco e profundamente assimétrico em termos de informação e poder.

A tradução dessa distinção teórica para o plano normativo exige a construção de critérios operacionais capazes de orientar tanto a interpretação jurídica quanto a atuação institucional. Embora a linha divisória entre persuasão legítima e manipulação ilícita permaneça, em muitos casos, situada em uma zona de indeterminação, é possível

identificar dimensões analíticas que permitem qualificar diferentes práticas comunicacionais segundo seu grau de compatibilidade com os princípios democráticos.

Essas dimensões não operam de forma isolada, mas articulam aspectos relacionados à transparência, veracidade, consentimento informacional, preservação da agência do eleitor e identificação do conteúdo como comunicação eleitoral. A partir dessa perspectiva, torna-se possível estruturar uma tipologia que distingue práticas claramente legítimas, situações intermediárias de ambiguidade normativa e condutas que, pela sua natureza, configuram manipulação eleitoral ilícita.

O quadro a seguir sistematiza esses critérios, oferecendo um referencial analítico para a avaliação das práticas contemporâneas de comunicação política no ambiente digital.

Tabela 22. Critérios de Distinção entre Persuasão Legítima e Manipulação Eleitoral

Dimensão	Persuasão Legítima	Zona Cinzenta	Manipulação Ilícita
Transparência	Identidade do emissor clara	Patrocínio obscurecido	Emissor deliberadamente falso
Veracidade	Conteúdo factualmente correto	Omissão seletiva de contexto	Afirmações factualmente falsas
Consentimento	Dados coletados com consentimento	Dados públicos reprocessados	Dados extraídos sem consentimento
Agência	Apela à capacidade reflexiva	Explora vieses conhecidos	Contorna deliberadamente a reflexividade
Rotulagem	Identificado como publicidade eleitoral	Conteúdo de aparência orgânica	Conteúdo gerado por IA não rotulado

Fonte: Elaborado pelo autor com base em Zuboff (2019), Resolução TSE nº 23.732/2024 e jurisprudência do TSE.

⚡ ATENÇÃO PRÁTICA — Implicação Jurídica para Campanhas Brasileiras

O uso de dados psicográficos inferidos de comportamento digital para microtargeting eleitoral no Brasil enfrenta múltiplas barreiras jurídicas. A LGPD (art. 11, I) exige consentimento específico e destacado para o tratamento de dados que revelem opiniões políticas. A inferência de opiniões políticas a partir de comportamento digital configura tratamento de dados sensíveis mesmo que os dados de origem não sejam individualmente sensíveis. A ausência de consentimento adequado expõe a campanha a sanções da ANPD e, se o tratamento irregular impactar o resultado eleitoral, a representação por abuso de poder perante a Justiça Eleitoral.

O assessor de campanha que contratar empresa de tecnologia sem verificar sua conformidade com a LGPD é solidariamente responsável pelas infrações cometidas.

O problema do uso de dados pessoais no contexto eleitoral pode ser compreendido, em sua formulação mais precisa, como uma manifestação aguda de assimetria informacional e de poder decisório entre emissores e destinatários da comunicação política. Trata-se de uma assimetria não apenas quantitativa, relativa ao volume de dados disponíveis, mas qualitativa, na medida em que envolve conhecimento estruturado sobre preferências, vulnerabilidades e padrões comportamentais dos eleitores, sem correspondente transparência sobre os mecanismos de sua utilização.

Campanhas eleitorais dotadas de recursos financeiros e capacidade tecnológica operam com bases de dados altamente granulares, capazes de capturar dimensões profundas do comportamento individual. Esses dados, frequentemente obtidos por meio de múltiplas fontes e técnicas de inferência, permitem a construção de perfis preditivos que extrapolam aquilo que o próprio eleitor reconhece explicitamente sobre si. Nesse sentido, o poder informacional das campanhas não reside apenas no acesso aos dados, mas na capacidade de transformá-los em conhecimento acionável para fins de influência.

O eleitor, por sua vez, encontra-se em posição estruturalmente desvantajosa. Ele não dispõe de informações sobre os critérios que orientam sua segmentação, não tem visibilidade sobre as diferentes versões de mensagens distribuídas a outros segmentos e, em regra, não possui instrumentos efetivos para compreender ou contestar os processos pelos quais é identificado como alvo de determinadas estratégias comunicacionais. Essa opacidade compromete não apenas a transparência do processo eleitoral, mas a própria possibilidade de exercício de um consentimento informado.

A noção de consentimento, central na regulação de dados pessoais, revela-se particularmente problemática nesse contexto. Em ambientes digitais, o consentimento tende a ser formalizado por meio de mecanismos padronizados, frequentemente compreendidos de forma limitada pelos usuários e desvinculados de contextos específicos de uso, como o eleitoral. Assim, ainda que juridicamente existente, esse consentimento pode carecer de substância material, na medida em que não pressupõe compreensão efetiva das finalidades e consequências do tratamento de dados.

Do ponto de vista normativo, essa configuração tensiona princípios fundamentais da ordem democrática, como a igualdade política e a autonomia do eleitor. A igualdade é afetada na medida em que diferentes atores dispõem de capacidades assimétricas de

influência, baseadas no acesso desigual a dados e tecnologia. A autonomia, por sua vez, é comprometida quando decisões políticas passam a ser moldadas por intervenções que operam abaixo do nível da consciência reflexiva, dificultando a avaliação crítica por parte do indivíduo.

Nesse cenário, a proteção de dados pessoais no contexto eleitoral não pode ser tratada apenas como uma questão de privacidade individual, mas deve ser compreendida como um elemento constitutivo da integridade do processo democrático. A regulação adequada desse campo exige, portanto, não apenas mecanismos de consentimento, mas instrumentos de transparência, auditabilidade e responsabilização que reduzam as assimetrias informacionais e restabeleçam condições mínimas de equilíbrio entre campanhas e eleitores.

A complexidade dessas assimetrias informacionais e das tensões normativas associadas ao uso de dados pessoais em campanhas eleitorais exige uma análise mais concreta das práticas efetivamente adotadas no contexto brasileiro. Para além dos princípios gerais da proteção de dados e da teoria democrática, torna-se necessário examinar como diferentes estratégias de microtargeting são enquadradas pelo ordenamento jurídico e avaliadas sob o prisma ético.

Essa análise revela um cenário marcado por heterogeneidade regulatória, no qual coexistem práticas claramente vedadas, zonas de ambiguidade normativa e condutas formalmente permitidas, mas eticamente problemáticas. A dissociação entre legalidade e legitimidade emerge, nesse contexto, como um dos traços mais característicos da campanha algorítmica contemporânea.

O quadro a seguir sistematiza esse panorama, classificando as principais práticas de microtargeting segundo seu status jurídico no Brasil, sua avaliação ética e observações relevantes sobre sua aplicação concreta.

Tabela 23. Status Jurídico e Ético das Práticas de Microtargeting no Brasil

Prática	Status Jurídico no Brasil	Status Ético	Observações
Publicidade segmentada por dados demográficos	Permitida com declaração ao TSE (Res. 23.610/2019).	Aceitável	Transparência para o regulador existe; para o eleitor, não.
Microtargeting por perfil psicográfico inferido	Zona cinzenta: LGPD exige base legal para dado sensível inferido.	Eticamente contestável	ANPD ainda não se pronunciou definitivamente sobre dados políticos inferidos.

Compra de listas de data brokers sem consentimento	Vedada pela LGPD arts. 7º e 11.	Antiético	Prática comum não acompanhada de fiscalização efetiva.
Dark posts (publicidade invisível ao público geral)	Registro obrigatório pelo TSE desde 2022.	Problemático para a transparência democrática	Visibilidade ao TSE não equivale a visibilidade pública.
Disparo em massa via WhatsApp por agências contratadas	Proibido: abuso de poder econômico (Lei 9.504/1997, art. 222-A).	Antiético	Responsabilização ainda insuficiente na prática.
Perfil psicográfico sem consentimento (modelo Cambridge Analytica)	Ilegal: dado político é dado sensível (LGPD art. 5º, II).	Gravemente antiético	Viola autonomia do eleitor e princípio da autodeterminação informativa.
Conteúdo eleitoral gerado por IA sem identificação	Proibido pela Resolução TSE nº 23.732/2024.	Antiético	Cumprimento depende de capacidade técnica de detecção das plataformas.

Fonte: Elaborado pelo autor com base em Resolução TSE 23.610/2019, LGPD, Lei 9.504/1997, Res. TSE 23.732/2024 e Apostila Curso IA e Eleições módulo 2 (2026).

■ RESUMO DO CAPÍTULO:

Este capítulo demonstrou que o microtargeting eleitoral se consolidou como um elemento estruturante do marketing político contemporâneo, ancorado em fundamentos teóricos robustos provenientes da psicologia comportamental e da ciência de dados. Sua eficácia não é contingente, mas explicável: decorre da capacidade de integrar modelos de decisão individual, vieses cognitivos e análise preditiva em estratégias comunicacionais altamente personalizadas. A articulação entre Robert Cialdini, Daniel Kahneman e Richard Thaler permite compreender os mecanismos pelos quais a persuasão é otimizada no nível individual, enquanto as contribuições de Shoshana Zuboff e Michal Kosinski evidenciam as condições estruturais que tornam esse modelo potencialmente manipulatório, especialmente no que se refere à exploração de dados comportamentais e à opacidade dos sistemas de segmentação. No plano jurídico, a análise indicou que o ordenamento brasileiro apresenta avanços relevantes, notadamente com a atuação do Tribunal Superior Eleitoral por meio da Resolução nº 23.732/2024 e com os parâmetros estabelecidos pela Lei Geral de Proteção de Dados. Ainda assim, tais instrumentos operam de forma parcial diante da complexidade do fenômeno, deixando lacunas significativas na regulação de práticas baseadas em inferência psicográfica, personalização algorítmica e distribuição opaca de conteúdo. A tensão central que atravessa o capítulo pode ser sintetizada na relação entre eficácia estratégica e legitimidade democrática. O microtargeting atinge seu máximo potencial precisamente quando atua sobre dimensões não reflexivas do comportamento, explorando vieses cognitivos e reduzindo o espaço de deliberação consciente. Contudo, é justamente nesse ponto que sua compatibilidade com os princípios democráticos se fragiliza, uma vez que a legitimidade do processo eleitoral depende da preservação da agência reflexiva e da autonomia informacional dos eleitores. Dessa forma, o microtargeting não representa apenas uma inovação técnica, mas um desafio normativo de primeira ordem. A questão que se impõe não é se essas tecnologias serão utilizadas, mas sob quais condições institucionais, regulatórias e éticas seu uso poderá ser compatibilizado com os requisitos mínimos de uma democracia substantiva.

QUESTÕES PARA REFLEXÃO E DEBATE

1. Em que medida o microtargeting psicográfico compromete a autonomia do eleitor ao operar abaixo do nível de sua consciência reflexiva, e é possível compatibilizar essa prática com os pressupostos normativos da democracia deliberativa?
2. A distinção entre persuasão legítima e manipulação eleitoral pode ser operacionalizada juridicamente a partir do critério da agência reflexiva, ou permanece inevitavelmente situada em uma zona de indeterminação normativa?
3. O uso de dados pessoais e inferências psicográficas em campanhas eleitorais configura uma violação da autodeterminação informativa, ou pode ser justificado como extensão legítima das estratégias tradicionais de comunicação política?
4. As capacidades preditivas e de personalização algorítmica criam uma assimetria estrutural de poder entre campanhas e eleitores. Essa assimetria compromete o princípio da igualdade política, ou pode ser mitigada por mecanismos regulatórios adequados?
5. Os instrumentos normativos existentes no Brasil, como a atuação do Tribunal Superior Eleitoral e a Lei Geral de Proteção de Dados, são suficientes para regular o microtargeting eleitoral, ou há necessidade de um novo paradigma regulatório específico para campanhas orientadas por dados?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Microtargeting não é segmentação demográfica. A diferença muda tudo.

Segmentar por faixa etária ou renda é uma coisa. Construir um perfil psicológico individual e prever quais argumentos têm maior probabilidade de alterar a intenção de voto de uma pessoa específica é outra completamente diferente. O microtargeting contemporâneo opera nesse segundo nível. Para profissionais de campanha, isso significa que o investimento em dados comportamentais e modelagem psicográfica tem retorno estratégico muito superior ao de uma segmentação demográfica tradicional. Para assessores jurídicos, significa que qualquer dado usado para inferir opinião política é dado sensível sob a LGPD, com exigências de consentimento específico que a maioria das campanhas ignora.

2. Seu data stack determina o teto da sua campanha.

Uma campanha sem CRM eleitoral estruturado, sem integração de dados comportamentais e sem modelo preditivo opera com uma fração da eficiência possível. Não é uma questão de preferência tecnológica: é uma questão de capacidade de alocar recursos onde eles produzem resultado. O data stack não substitui a política, o candidato ou a relação com o eleitor. Mas ele determina, com crescente precisão, quem recebe qual mensagem, quando, em qual canal e com qual apelo emocional. Campanhas que constroem esse sistema cedo têm vantagem acumulada ao longo de meses. Campanhas que improvisam na reta final disputam em desvantagem estrutural.

3. Kahneman, Cialdini e Thaler não são teoria. São o manual de operação da campanha digital.

Os princípios de influência de Cialdini, a distinção entre Sistema 1 e Sistema 2 de Kahneman e o conceito de arquitetura de escolhas de Thaler e Sunstein descrevem exatamente o que os algoritmos de

recomendação e os sistemas de microtargeting fazem na prática. Prova social, aversão à perda, ancoragem, enquadramento emocional: cada um desses mecanismos tem aplicação direta no design de mensagens eleitorais. A diferença é que, no ambiente digital, eles não são aplicados por um redator numa sala de reunião. São operacionalizados em larga escala, de forma automatizada, para milhões de eleitores simultaneamente, com ajuste contínuo baseado em dados de comportamento em tempo real.

4. O modelo OCEAN transforma traços de personalidade em estratégia de mensagem.

Eleitores com alto neuroticismo respondem mais intensamente a mensagens de ameaça e insegurança. Eleitores com alta abertura a experiências são mais receptivos a discursos de mudança e inovação. Eleitores com alta conscienciosidade tendem a valorizar ordem, previsibilidade e responsabilidade institucional. Com apenas 70 curtidas no Facebook, um algoritmo consegue prever esses traços com precisão superior à de amigos próximos. Para campanhas com capacidade de integrar essa modelagem, isso significa adaptar não apenas o tema da mensagem, mas seu tom emocional, seu vocabulário e seu enquadramento narrativo ao perfil psicológico de cada segmento. Para a regulação, significa que dados aparentemente banais de comportamento digital são, na prática, dados sensíveis sendo processados sem consentimento.

5. Dark posts, disparos de WhatsApp e perfis psicográficos sem consentimento têm enquadramentos jurídicos distintos. Conhecer a diferença é obrigação do assessor.

Publicidade segmentada por dados demográficos com declaração ao TSE é permitida. Microtargeting por perfil psicográfico inferido é zona cinzenta, com a LGPD exigindo base legal para dado sensível que a maioria das campanhas não consegue demonstrar. Compra de listas de data brokers sem consentimento é vedada pela LGPD. Disparo em massa via WhatsApp por agências contratadas é proibido como abuso de poder econômico. Conteúdo gerado por IA sem rotulagem é proibido pela Resolução TSE 23.732/2024. O assessor que contrata uma empresa de tecnologia sem verificar sua conformidade com a LGPD é solidariamente responsável pelas infrações. Ignorância sobre o enquadramento jurídico não é defesa: é agravante.

6. A linha entre persuasão legítima e manipulação ilícita é conceitual, mas tem consequências práticas reais.

Persuasão legítima se dirige à capacidade reflexiva do eleitor: oferece argumentos que podem ser avaliados, mesmo que use apelos emocionais. Manipulação opera sobre mecanismos não reflexivos, explorando vieses cognitivos e vulnerabilidades psicológicas para induzir decisões que o eleitor não endossaria em condições de plena consciência. O microtargeting psicográfico de alta precisão, especialmente quando combinado com dados obtidos sem consentimento e conteúdo gerado por IA sem identificação, aproxima-se estruturalmente da segunda categoria. Isso não é apenas um problema filosófico: é um critério que a jurisprudência eleitoral está progressivamente incorporando para avaliar condutas que, embora não se enquadrem nos tipos tradicionais, comprometem a integridade do processo eleitoral.

7. 66 mil simulações por noite. Esse é o nível de sofisticação com que campanhas competitivas operam.

A campanha de Obama em 2012 realizava dezenas de milhares de simulações computacionais diárias, variando parâmetros de comportamento eleitoral para estimar efeitos sobre o resultado final. Isso não é ficção científica nem privilégio exclusivo de superpotências: é a fronteira do que campanhas com acesso a dados, modelos preditivos e capacidade analítica já fazem. No Brasil, campanhas competitivas de âmbito nacional já operam com nível similar de sofisticação em modelagem e simulação. O que separa uma campanha que usa bem esses recursos de uma que os usa mal não é apenas tecnologia: é a qualidade dos dados, a competência da equipe analítica e, crescentemente, a capacidade de operar dentro dos limites legais sem perder eficácia estratégica.

CAPÍTULO 5

Cambridge Analytica

Anatomia de um Escândalo, Paradigma Global e Implicações para o Direito Eleitoral Brasileiro

O Escândalo na Perspectiva Acadêmica

O escândalo Cambridge Analytica, revelado em março de 2018 pelas reportagens de Carole Cadwalladr e Emma Graham-Harrison no *The Guardian* e no *The Observer*, rapidamente transcendeu o espaço do jornalismo investigativo para se tornar objeto de análise acadêmica intensa nas áreas de ciência política, direito, comunicação e ética da tecnologia. Mais do que um episódio isolado de violação de dados, o caso funcionou como catalisador de um debate que já se acumulava há anos sobre a compatibilidade entre as práticas emergentes de marketing político orientado por dados e os fundamentos normativos da democracia liberal.

A literatura acadêmica que se formou em torno do caso pode ser organizada em três linhas de investigação, cada uma com pressupostos metodológicos e implicações normativas distintos.

A Primeira Linha: Análise Técnica e o Problema da Magnitude

A primeira linha de investigação é a da análise técnica, que examina os métodos de coleta e processamento de dados e questiona, com base em evidência empírica, a magnitude real do impacto eleitoral atribuído à Cambridge Analytica. Essa linha é representada, de forma paradigmática, pelos trabalhos de Guess, Nagler e Tucker (2019), que investigaram a relação entre exposição a desinformação e mudança efetiva de comportamento eleitoral durante as eleições norte-americanas de 2016.

A contribuição central desses autores é metodológica antes de ser substantiva: eles demonstram que a narrativa dominante sobre o poder transformador do microtargeting psicográfico sobre o resultado eleitoral carece de sustentação empírica robusta. A exposição a conteúdo desinformativo, embora real e mensurável, concentrou-se de forma desproporcional em segmentos do eleitorado que já apresentavam predisposições políticas consolidadas, o que limita significativamente seu potencial de conversão. Eleitores indecisos, que seriam o alvo estratégico mais relevante de qualquer campanha de persuasão, mostraram-se menos expostos ao tipo de conteúdo que a Cambridge Analytica alegava ser capaz de influenciar.

Esse resultado tem implicações teóricas relevantes que vão além da avaliação retrospectiva de uma campanha específica. Ele ressuscita, em novo contexto, a tese clássica de Lazarsfeld sobre os efeitos limitados da comunicação política de massa: a ideia de que a exposição a mensagens políticas tende a reforçar predisposições existentes mais do que a convertê-las.

Se essa tese se sustenta mesmo no ambiente de microtargeting de alta precisão, então parte da narrativa sobre o poder disruptivo da Cambridge Analytica pode ter sido construída tanto pela empresa para valorizar seus serviços quanto pela cobertura jornalística em busca de uma explicação causal simples para um resultado eleitoral perturbador.

Isso não significa, contudo, que o escândalo seja irrelevante do ponto de vista normativo. O problema pode não residir primariamente nos efeitos eleitorais diretos, mas na natureza das práticas em si: na coleta de dados sem consentimento, na opacidade dos mecanismos de influência e na ausência de qualquer prestação de contas sobre o que foi feito com os dados de 87 milhões de usuários.

Nesse sentido, a análise técnica, ao questionar a eficácia do microtargeting, paradoxalmente reforça a análise normativa: se o problema não é que a Cambridge Analytica conseguiu manipular a eleição, o problema é que ela tentou, com recursos e métodos que o ordenamento jurídico ainda não estava equipado para alcançar.

Ainda no campo da análise técnica, os trabalhos de Albright (2016) sobre redes de desinformação e os de Howard e Kollanyi (2016) sobre o papel dos bots na amplificação de conteúdo político oferecem uma perspectiva complementar. Esses estudos demonstram que a infraestrutura técnica de uma campanha de influência algorítmica não se resume à modelagem psicográfica, mas envolve uma arquitetura distribuída de amplificação coordenada, na qual agentes automatizados e redes de contas inautênticas desempenham papel central na criação de aparência de organicidade para conteúdo estrategicamente produzido.

A Segunda Linha: Análise Regulatória e os Limites do Direito

A segunda linha de investigação é a da análise regulatória, que examina as respostas jurídicas e institucionais ao caso e avalia a adequação dos marcos normativos vigentes para lidar com o tipo de prática que o escândalo expôs. As contribuições de Rubinstein (2014) e Solove (2013) são particularmente relevantes porque foram produzidas antes do escândalo, o que lhes confere o estatuto de diagnóstico antecipado das vulnerabilidades que o caso viria a materializar.

Solove, em sua teoria pragmática da privacidade, argumenta que os frameworks tradicionais de proteção de dados, centrados na ideia de consentimento informado como mecanismo primário de proteção, são estruturalmente insuficientes para lidar com os riscos que emergem da agregação de dados em larga escala. O problema não é que os usuários não consentiram com cada coleta individual de dado, mas que ninguém consente com o perfil que emerge da combinação de milhares de dados aparentemente triviais. A privacidade, nessa perspectiva, não é uma propriedade de dados isolados, mas uma propriedade relacional que emerge de padrões de agregação e inferência. O escândalo Cambridge Analytica é, em larga medida, uma ilustração empírica precisa dessa tese.

Rubinstein, por sua vez, examina os limites da autorregulação no campo da publicidade comportamental e antecipa a insuficiência dos mecanismos de opt-out como instrumento de proteção efetiva da autonomia do usuário. Sua análise é relevante para o contexto eleitoral porque demonstra que, mesmo quando usuários têm acesso formal a mecanismos de recusa, as assimetrias de informação e de poder entre plataformas e indivíduos tornam esses mecanismos estruturalmente ineficazes para a maioria da população.

No plano das respostas institucionais ao escândalo, a investigação conduzida pela Information Commissioner's Office do Reino Unido e as audiências no Congresso norte-

americano com Mark Zuckerberg produziram um registro documental de grande relevância analítica. Ambos os processos evidenciaram não apenas as falhas das empresas envolvidas, mas as limitações das próprias autoridades regulatórias, que demonstraram dificuldade em formular perguntas tecnicamente precisas sobre sistemas cuja arquitetura não compreendiam plenamente. Essa assimetria de conhecimento entre reguladores e regulados é, ela própria, um problema normativo de primeira ordem, e sua superação exige investimento institucional em capacidade técnica que a maioria dos sistemas regulatórios ainda não realizou de forma adequada.

No contexto brasileiro, o escândalo Cambridge Analytica teve um papel indireto mas relevante na aceleração do processo legislativo que culminou na aprovação da Lei Geral de Proteção de Dados em 2018. A exposição pública das consequências da ausência de regulação adequada criou uma janela de oportunidade política que facilitou a tramitação de um projeto que até então avançava de forma lenta no Congresso Nacional. Compreender essa conexão é importante porque ela revela que o direito não apenas reage a eventos, mas é moldado pela percepção pública de riscos que episódios como o da Cambridge Analytica tornam concretos e inteligíveis para além dos círculos especializados.

A Terceira Linha: Análise Normativa e a Crise da Legitimidade Democrática

A terceira linha de investigação é a da análise normativa, que questiona as implicações do caso para a teoria democrática e para os fundamentos de legitimidade dos processos eleitorais contemporâneos. É nessa linha que o debate atinge sua maior profundidade teórica e suas implicações mais amplas.

O enquadramento de Shoshana Zuboff sobre capitalismo de vigilância constitui a referência mais influente e, ao mesmo tempo, mais ambiciosa dessa linha. Para Zuboff, o escândalo Cambridge Analytica não é uma anomalia nem um desvio de conduta de atores específicos: é a expressão sintomática de um modelo econômico estruturalmente orientado à extração e mercantilização de dados comportamentais para fins de modificação de comportamento em escala. O capitalismo de vigilância, em sua formulação, não apenas coleta dados sobre o que as pessoas fazem: ele opera para prever e moldar o que farão, transformando a experiência humana em matéria-prima para produtos de certeza comportamental vendidos a mercados de futuros de comportamento.

Nessa perspectiva, a Cambridge Analytica não inventou um novo modelo: ela aplicou ao campo político uma lógica que as grandes plataformas já operavam no campo comercial. A diferença é que, quando essa lógica incide sobre o processo eleitoral, ela não apenas

viola a privacidade individual, mas compromete o mecanismo central pelo qual as democracias produzem legitimidade: a formação autônoma de preferências políticas por cidadãos livres.

Essa tese encontra tensão produtiva com a perspectiva de Éveline Lubbers e outros pesquisadores que, sem negar os problemas identificados por Zuboff, questionam se o conceito de capitalismo de vigilância captura com precisão as especificidades do caso eleitoral ou se, ao contrário, tende a produzir uma narrativa de onipotência tecnológica que paradoxalmente legitima o poder que pretende criticar. Se os algoritmos são apresentados como capazes de manipular eleições com precisão cirúrgica, a narrativa pode, inadvertidamente, reforçar a percepção de que a resistência individual é inútil e que a agência política do eleitor comum é ilusória.

A tensão entre essas perspectivas é teoricamente fecunda porque aponta para um problema genuíno na análise normativa do microtargeting: a dificuldade de calibrar adequadamente a magnitude dos efeitos sem incorrer, de um lado, no minimalismo que subestima riscos reais, ou, de outro, no alarmismo que superestima capacidades e produz desempoderamento cívico. Encontrar esse equilíbrio analítico é uma das tarefas mais importantes da pesquisa nesse campo e uma das condições para que a regulação seja proporcional, eficaz e compatível com os valores que pretende proteger.

Um terceiro vetor da análise normativa é representado pelos trabalhos de Luciano Floridi sobre ética da informação e pelos de Michael Sandel sobre os limites morais do mercado. Ambos oferecem instrumentos conceituais para argumentar que certas dimensões da experiência humana, entre elas a formação de preferências políticas, não devem ser tratadas como mercadorias, independentemente do consentimento formal dos envolvidos.

Essa perspectiva não resolve os problemas jurídicos imediatos, mas fornece o fundamento filosófico para regulações que vão além da proteção de dados individuais e alcançam a dimensão coletiva e institucional do problema democrático.

A Cambridge Analytica no Ecossistema do Poder: Arquitetura Organizacional e Modelo de Negócios

Compreender o escândalo Cambridge Analytica exige, antes de tudo, resistir à tentação de tratá-lo como o produto de uma empresa excêntrica operando nas margens do sistema. A Cambridge Analytica não era uma anomalia: era a expressão mais visível de um ecossistema mais amplo, que combina capacidade técnica de coleta e processamento de

dados, expertise em operações de influência psicológica e demanda política por instrumentos de persuasão de alta precisão. Compreender sua arquitetura organizacional é compreender como esse ecossistema funciona e por que ele tende a se reproduzir independentemente da sorte de qualquer empresa específica.

Das Operações de Influência ao Marketing Eleitoral: A SCL Group como Matriz

A Cambridge Analytica foi fundada em 2013 como subsidiária norte-americana da SCL Group, empresa britânica com sede em Londres e histórico documentado de atuação em operações de influência psicológica em contextos de segurança nacional e em processos eleitorais em países em desenvolvimento. Essa genealogia institucional é central para compreender a natureza da Cambridge Analytica e não deve ser reduzida a um detalhe biográfico de menor importância.

A SCL Group, fundada na década de 1990, especializou-se no que seus documentos internos denominavam ISTAR, acrônimo para Intelligence, Surveillance, Target Acquisition and Reconnaissance, aplicado não a contextos militares convencionais, mas a ambientes políticos e sociais. A empresa desenvolveu metodologias de análise de populações, identificação de vulnerabilidades psicológicas coletivas e design de campanhas de comunicação orientadas a modificar comportamentos em contextos de conflito político e instabilidade institucional. Entre seus clientes documentados figuravam governos, forças armadas e partidos políticos em países da África, Ásia e América Latina, onde a linha entre consultoria eleitoral e operação de influência tende a ser mais porosa do que em democracias consolidadas.

Esse histórico é relevante por duas razões. A primeira é que ele revela que as técnicas aplicadas pela Cambridge Analytica em eleições norte-americanas e europeias não foram desenvolvidas originalmente para contextos democráticos maduros, mas testadas e refinadas em ambientes com menor capacidade regulatória e menor resistência institucional. A segunda é que ele situa a empresa não como um player do setor de marketing político convencional que adotou técnicas inovadoras, mas como uma organização com cultura e métodos oriundos de um campo, o das operações de influência em contextos de segurança, no qual a manipulação não é um efeito colateral indesejado, mas um objetivo explícito.

A transição da SCL Group para o mercado eleitoral norte-americano, materializada na criação da Cambridge Analytica com apoio financeiro de Robert Mercer e orientação estratégica de Steve Bannon, representou uma tentativa de aplicar essa expertise a um

contexto de altíssimo valor político e, ao mesmo tempo, de elevada visibilidade regulatória. O resultado foi a colisão entre métodos desenvolvidos para ambientes de baixa accountability e um contexto no qual, eventualmente, a exposição pública seria inevitável.

O Modelo de Negócios: Três Competências em Cadeia

O modelo de negócios da Cambridge Analytica combinava três competências distintivas que, articuladas em sequência, constituíam uma proposta de valor singular no mercado de consultoria eleitoral.

A primeira competência era a coleta de dados pessoais em escala, realizada por múltiplas vias que incluíam parcerias com data brokers, desenvolvimento de aplicativos que capturavam dados de usuários e suas redes de contatos, compra de bases de dados comerciais e, no caso mais emblemático, a exploração da API de amigos do Facebook para acesso não autorizado a perfis de usuários que nunca consentiram com qualquer coleta. A capacidade de agregar dados de fontes heterogêneas e construir perfis individuais de alta granularidade era apresentada pela empresa como sua principal vantagem competitiva em relação a concorrentes que operavam com bases de dados mais limitadas.

A segunda competência era a modelagem psicográfica pelo modelo OCEAN, que convertia os dados coletados em perfis de personalidade individualizados organizados segundo as cinco dimensões do modelo: Abertura, Conscienciosidade, Extroversão, Amabilidade e Neuroticismo. Essa competência era apresentada como o elemento de diferenciação científica da empresa, que se posicionava não como uma agência de publicidade política convencional, mas como uma organização com capacidade de inferência psicológica baseada em evidência empírica. A pesquisa de Kosinski, Stillwell e Graepel, publicada na PNAS em 2013, era frequentemente citada como validação científica dessa abordagem, ainda que os próprios autores tenham posteriormente se distanciado de qualquer associação com a Cambridge Analytica e contestado a utilização de seu trabalho como fundamento para as alegações da empresa.

A terceira competência era a produção e distribuição de mensagens políticas personalizadas com base nos perfis gerados. Essa etapa fechava o ciclo entre dado comportamental e intervenção comunicacional: os perfis psicográficos informavam não apenas o conteúdo das mensagens, mas seu tom emocional, seu vocabulário, seu formato e o canal de distribuição mais adequado para cada segmento do eleitorado. A empresa alegava ser capaz de produzir variações de mensagem calibradas para cada uma das

dimensões do modelo OCEAN, maximizando a ressonância de cada comunicação com as disposições psicológicas específicas de seu receptor.

A articulação dessas três competências em uma oferta integrada era o que diferenciava a Cambridge Analytica de concorrentes que operavam apenas em partes isoladas desse processo. Uma agência de publicidade tradicional podia produzir mensagens segmentadas, mas não dispunha da infraestrutura de dados para identificar os perfis com a granularidade que a Cambridge Analytica alegava ter. Uma empresa de análise de dados podia construir modelos preditivos sofisticados, mas não tinha a capacidade criativa e de distribuição para converter esses modelos em campanhas operacionais. A proposta de valor da Cambridge Analytica era precisamente a integração vertical dessas capacidades em uma única organização.

O Posicionamento no Ecossistema: Dependências, Parceiros e Vulnerabilidades

A Cambridge Analytica não operava de forma autônoma. Ela era um nó em um ecossistema mais amplo de atores interdependentes, cuja análise é essencial para compreender tanto como o escândalo foi possível quanto por que práticas similares tendem a persistir mesmo após o colapso de uma empresa específica.

O primeiro eixo de dependência era o das plataformas de dados. A capacidade da Cambridge Analytica de construir perfis psicográficos em escala dependia fundamentalmente do acesso a dados gerados por plataformas como o Facebook, que concentravam volumes sem precedente de informação comportamental sobre centenas de milhões de usuários. Essa dependência estrutural criava uma relação assimétrica: a Cambridge Analytica precisava do Facebook para operar, mas o Facebook precisava de empresas como a Cambridge Analytica para monetizar seu ecossistema de desenvolvedores. A API de amigos, cujo abuso viabilizou a coleta de 87 milhões de perfis, existia porque o Facebook tinha interesse econômico em manter um ecossistema de aplicativos de terceiros que ampliasse o engajamento na plataforma. A vulnerabilidade explorada pela Cambridge Analytica não era um bug: era uma feature cujos riscos foram subestimados ou negligenciados.

O segundo eixo de dependência era o dos clientes políticos. A Cambridge Analytica era uma empresa de serviços, e sua sobrevivência dependia da disposição de campanhas eleitorais e partidos políticos em contratar seus serviços. Essa dependência é relevante porque situa a responsabilidade pelo escândalo de forma mais distribuída do que a narrativa dominante, centrada na empresa, tende a sugerir. Os clientes que contrataram a Cambridge Analytica tinham, em graus variáveis, conhecimento sobre os métodos utilizados e interesse

nos resultados produzidos. A questão sobre em que medida campanhas políticas que contrataram a empresa podem ser responsabilizadas por práticas irregulares de coleta de dados é uma das que a análise regulatória ainda não respondeu de forma satisfatória.

O terceiro eixo de dependência era o dos pesquisadores e acadêmicos cujo trabalho foi instrumentalizado pela empresa. O caso de Aleksandr Kogan é paradigmático: um pesquisador universitário com acesso legítimo a dados do Facebook para fins de pesquisa científica transferiu esses dados para uma empresa comercial, violando os termos de uso da plataforma e os princípios éticos que regem a pesquisa com dados humanos. Esse episódio levanta questões sobre os mecanismos de governança da pesquisa acadêmica em ciência de dados e sobre a porosidade entre a produção de conhecimento científico e sua instrumentalização por interesses comerciais e políticos.

A Dissolução da Empresa e a Persistência do Modelo

A Cambridge Analytica encerrou suas operações em maio de 2018, dois meses após a revelação do escândalo. A dissolução formal da empresa, contudo, não equivale ao desaparecimento do modelo que ela representava. Parte de sua liderança e de seu pessoal técnico migrou para novas organizações com nomes diferentes, mas com propostas de valor estruturalmente similares.

O ecossistema de data brokers que fornecia dados à empresa continuou operando. As plataformas que viabilizaram a coleta em escala continuaram existindo, ainda que com restrições adicionais às suas APIs. E a demanda política por instrumentos de persuasão baseados em dados comportamentais não apenas persistiu como se intensificou, alimentada pela percepção, correta ou não, de que quem domina os dados domina a eleição.

Esse padrão de dissolução formal seguida de reprodução informal é característico de setores nos quais a vantagem competitiva reside em conhecimento tácito e em redes de relacionamento, mais do que em ativos físicos ou marcas registradas. A lição mais importante do caso Cambridge Analytica, do ponto de vista regulatório, pode não ser sobre essa empresa específica, mas sobre a necessidade de regular o modelo, e não apenas os atores que o operacionalizam em determinado momento. Enquanto a regulação se concentrar em responsabilizar empresas individuais após a ocorrência de danos, o ecossistema que torna esses danos possíveis permanecerá estruturalmente intacto.

O Mecanismo Técnico: Como os Dados Foram Coletados Sem Consentimento

A análise do mecanismo técnico pelo qual a Cambridge Analytica obteve dados de aproximadamente 87 milhões de usuários do Facebook revela algo mais perturbador do que uma simples violação de privacidade em larga escala. Ela expõe uma vulnerabilidade estrutural inscrita na própria arquitetura das plataformas de redes sociais, cuja lógica econômica incentiva a abertura de dados a desenvolvedores externos ao mesmo tempo em que cria condições para o abuso sistêmico desse acesso. Compreender esse mecanismo em profundidade é condição para compreender por que episódios similares continuam possíveis mesmo após as reformas introduzidas pelo escândalo.

Tabela 24. Fluxo de Dados no Esquema Cambridge Analytica

Etapa	Ator	Ação	Volume	Legalidade
1	Usuários do Facebook	Instalação voluntária do quiz	270 mil perfis	Legal (consentimento formal)
2	API do Facebook	Acesso aos dados de amigos	87 milhões de perfis	Ilegal (violação TOS Facebook)
3	Aleksandr Kogan	Transferência para CA	87 milhões de perfis	Ilegal (violação TOS e acordos acadêmicos)
4	Cambridge Analytica	Modelagem psicográfica OCEAN	87 milhões de perfis	Ilegal (uso comercial não autorizado)
5	Campanhas clientes	Targeting eleitoral personalizado	Segmentos eleitorais identificados	Juridicamente contestado

Fonte: Elaborado pelo autor com base em Cadwalladr e Graham-Harrison (2018) e Wylie (2019).

O instrumento central da coleta foi a chamada API de amigos do Facebook, vigente até 2015. Para compreender seu funcionamento e suas implicações, é necessário entender o modelo econômico que a justificava e os incentivos que a tornaram possível.

O Facebook, como toda plataforma de redes sociais baseada em publicidade, tem interesse econômico fundamental em maximizar o engajamento dos usuários. Uma das estratégias para ampliar esse engajamento é o desenvolvimento de um ecossistema de aplicativos de terceiros que utilizem os dados da plataforma para oferecer serviços adicionais aos usuários, criando razões para que permaneçam no ambiente da plataforma por mais tempo e com maior intensidade. Para viabilizar esse ecossistema, o Facebook disponibilizava aos desenvolvedores credenciados uma interface de programação, a API, que permitia o acesso a dados dos usuários que autorizassem o aplicativo.

Até 2015, essa API operava segundo uma lógica de acesso em cascata que, em retrospecto, era uma vulnerabilidade de proporções extraordinárias: quando um usuário autorizava um aplicativo, esse aplicativo obtinha acesso não apenas aos dados do usuário que o havia instalado, mas também aos dados de todos os seus amigos na plataforma, independentemente de qualquer consentimento por parte desses amigos. A autorização de um único usuário abria uma janela de acesso que se estendia, de forma invisível, por toda a sua rede de contatos.

A justificativa técnica para esse design era a melhoria da experiência do usuário: um aplicativo de jogo social, por exemplo, precisava de acesso aos dados dos amigos do usuário para exibir placares comparativos ou enviar convites. A justificativa econômica era igualmente clara: quanto mais dados os desenvolvedores pudessem acessar, mais sofisticados e envolventes seriam os aplicativos que construiriam, e mais tempo os usuários passariam na plataforma. O que essa lógica desconsiderava, ou desconsiderou convenientemente, era que ela transformava cada usuário que instalasse qualquer aplicativo em um vetor involuntário de exposição de dados de toda a sua rede de contatos.

Essa arquitetura não era desconhecida dos engenheiros do Facebook. Era uma decisão de design consciente, tomada em contexto de pressão competitiva intensa para ampliar o ecossistema de desenvolvedores e aumentar o engajamento. A decisão de mantê-la por anos, apesar de evidências crescentes de potencial para abuso, é em si mesma um objeto relevante de análise normativa sobre as responsabilidades das plataformas na governança de suas infraestruturas técnicas.

Tabela 25. Cambridge Analytica: As Quatro Dimensões do Escândalo

Dimensão	O que Aconteceu	Consequência Jurídica	Lição para o Brasil
Coleta de dados	87 milhões de perfis coletados via app com autorização de 270.000, exploração de API do Facebook que permitia acesso a dados de amigos não instaladores.	Facebook: multa US$ 5 bilhões (FTC, 2019). ICO britânico: £500.000. Nenhuma condenação criminal individual.	A LGPD exige consentimento específico e destacado para dados sensíveis (opinião política). Coleta indireta via amigos sem consentimento seria ilegal no Brasil desde 2020.
Perfil psicográfico	Modelo OCEAN inferido a partir de curtidas, relacionando comportamento digital a traços de personalidade e	Dado político inferido (opinião política) = dado sensível pela LGPD, art. 5º, II. Sem base legal para	A inferência de opinião política a partir de dados não sensíveis exige consentimento específico. Questão jurídica ainda não

	vulnerabilidades psicológicas específicas.	tratamento sem consentimento.	resolvida definitivamente pela ANPD.
Microtargeting psicográfico	Mensagens calibradas para explorar vulnerabilidades específicas de cada perfil OCEAN, especialmente neuroticismo alto (suscetibilidade a medo) para narrativas de ameaça.	Não há tipo penal específico nos EUA ou no Brasil. Na teoria, pode configurar manipulação eleitoral ilícita se demonstrado dolo e nexo causal.	Resolução TSE 23.732/2024 proíbe sistemas que utilizem técnicas subliminares ou deliberadamente manipulativas, mas o threshold probatório é elevado.
Voter suppression digital	Identificação de eleitores democratas potencialmente persuasíveis para descampaign: envio de narrativas negativas sobre Clinton para desestimulá-los a votar.	Voter suppression é crime federal nos EUA. No Brasil, desestímulo ao voto por meios fraudulentos pode configurar crime eleitoral.	Art. 222-A da Lei 9.504/1997 proíbe disparo em massa para fins ilícitos, incluindo desestímulo fraudulento ao comparecimento.

Fonte: Elaborado pelo autor com base em Cadwalladr e Graham-Harrison (2018), Wylie (2019)

O Aplicativo de Kogan: Engenharia do Consentimento Fraturado

O pesquisador Aleksandr Kogan, professor da Universidade de Cambridge especializado em psicometria e comportamento digital, desenvolveu um aplicativo denominado "thisisyourdigitallife", apresentado aos usuários como um quiz de personalidade com propósito científico. O aplicativo foi instalado voluntariamente por aproximadamente 270 mil usuários, que consentiram, nos termos da autorização padrão da API, com o acesso do aplicativo aos seus dados do Facebook.

Esse consentimento, formalmente existente para os 270 mil usuários que instalaram o aplicativo, era, na formulação de Solove, um consentimento fraturado: juridicamente válido em sua forma, mas substancialmente vazio em seu conteúdo. Os usuários concordaram com os termos de uso do aplicativo, que eram apresentados, como é prática universal nesse tipo de interface, em formato de texto longo e tecnicamente denso que praticamente nenhum usuário lê integralmente. Estudos sobre o comportamento de usuários diante de termos de serviço digitais demonstram consistentemente que a taxa de leitura efetiva desses documentos é próxima de zero, o que transforma o consentimento formal em uma ficção jurídica cujo único efeito prático é transferir responsabilidade do operador para o usuário.

Mais grave, contudo, é a situação dos aproximadamente 87 milhões de usuários cujos dados foram coletados sem que tivessem instalado o aplicativo ou dado qualquer forma de consentimento, ainda que formal. Esses usuários eram amigos dos 270 mil que haviam instalado o aplicativo, e seus dados foram acessados exclusivamente em virtude da API de amigos, sem qualquer interação sua com o sistema. Do ponto de vista da teoria do consentimento, sua situação é qualitativamente distinta: não se trata de consentimento fraturado, mas de ausência absoluta de consentimento. Eles não participaram de nenhuma decisão, não foram informados de nenhuma coleta e não tinham qualquer mecanismo para impedir o acesso aos seus dados.

Essa distinção é juridicamente relevante porque ela revela que o escândalo não envolve apenas uma violação dos termos de uso de uma plataforma, mas uma coleta de dados de dezenas de milhões de pessoas sem qualquer base legal identificável. Sob o marco da LGPD, que exige base legal para todo tratamento de dados pessoais, e especialmente para dados sensíveis como opiniões políticas, a conduta seria categoricamente ilegal. Sob o GDPR europeu, que estava em processo de finalização no período dos fatos, ela violaria múltiplos princípios fundamentais, incluindo licitude, lealdade, transparência e limitação de finalidade.

A coleta dos dados por Kogan foi realizada sob o pretexto de pesquisa acadêmica, o que lhe conferia acesso à API em condições que não seriam disponíveis a um desenvolvedor comercial sem esse enquadramento. Os termos de uso do Facebook para desenvolvedores proibiam expressamente a transferência de dados obtidos por meio da API para terceiros e sua utilização para fins comerciais. Kogan transferiu os dados coletados para a Cambridge Analytica em violação direta dessas disposições.

Essa violação tem implicações que vão além da responsabilidade individual de Kogan. Ela expõe a fragilidade dos mecanismos de governança da pesquisa acadêmica com dados de plataformas digitais, que dependem fundamentalmente da boa-fé dos pesquisadores e da capacidade das plataformas de monitorar o uso que é feito dos dados que disponibilizam. Nenhuma dessas condições estava presente no caso: Kogan agiu em violação às normas aplicáveis, e o Facebook não tinha sistemas efetivos para detectar ou impedir a transferência não autorizada de dados.

O episódio também levanta questões sobre a responsabilidade da Cambridge Analytica na cadeia de custódia dos dados. A empresa sabia, ou tinha condições de saber, que os dados que estava adquirindo haviam sido coletados por um pesquisador acadêmico com acesso a uma API destinada a fins de pesquisa, não a fins comerciais. A *due diligence* mínima sobre a proveniência dos dados que adquiria revelaria imediatamente a

irregularidade da transferência. A ausência dessa *due diligence,* ou sua realização e subsequente desconsideração dos resultados, configura, no mínimo, negligência grave, e possivelmente cumplicidade ativa na violação.

A Construção dos Perfis Psicográficos: Do Dado Bruto ao Instrumento de Influência

Os dados coletados por Kogan e transferidos para a Cambridge Analytica incluíam informações diversas dos perfis do Facebook: curtidas em páginas e publicações, dados demográficos, redes de amigos, localização geográfica e, em muitos casos, conteúdo de publicações e mensagens privadas acessíveis via API. Esse conjunto heterogêneo de informações era, em si mesmo, um dado bruto sem valor estratégico imediato. O valor era gerado pelo processamento que a Cambridge Analytica aplicava a esses dados para produzir perfis psicográficos organizados segundo o modelo OCEAN.

O processo de modelagem psicográfica a partir de dados do Facebook seguia, em linhas gerais, a metodologia desenvolvida por Kosinski, Stillwell e Graepel, cujas implicações foram discutidas nos capítulos anteriores. As curtidas em páginas e publicações eram tratadas como indicadores de traços de personalidade, e modelos de aprendizado de máquina eram treinados para inferir os cinco scores do modelo OCEAN a partir desses indicadores. O resultado era um perfil numérico que situava cada usuário em um espaço multidimensional de personalidade, permitindo segmentações e personalizações de mensagem em nível individual.

É importante notar que a precisão dessas inferências é, ela própria, um objeto de controvérsia científica. Os modelos preditivos treinados com dados do Facebook demonstraram capacidade estatisticamente significativa de inferir traços de personalidade em contextos de pesquisa controlada, mas a transposição dessas capacidades para um contexto de campanha eleitoral em larga escala introduz variáveis adicionais que limitam a validade preditiva dos modelos. A Cambridge Analytica, em suas comunicações comerciais, tendia a apresentar suas capacidades em termos que superestimavam consistentemente o que a evidência científica disponível sustentava.

Essa inflação das capacidades alegadas tem implicações analíticas relevantes. De um lado, ela sugere que o impacto eleitoral efetivo das operações da empresa pode ter sido inferior ao que a narrativa dominante do escândalo pressupõe, na linha do argumento de Guess, Nagler e Tucker. De outro, ela revela que a empresa operava em um mercado no qual a demanda política por certeza e por instrumentos de controle era suficientemente

intensa para criar disposição de pagamento por capacidades que não eram inteiramente verificáveis.

Esse mercado da ilusão de precisão é, em si mesmo, um fenômeno com implicações normativas relevantes: ele incentiva a produção e comercialização de tecnologias cuja eficácia é exagerada, o que distorce as decisões de investimento das campanhas e alimenta uma narrativa pública de onipotência algorítmica que tem efeitos próprios sobre o comportamento eleitoral e a confiança nas instituições democráticas.

O Legado Técnico: O que Mudou e o que Permanece

Em resposta ao escândalo, o Facebook encerrou a API de amigos em 2015 e introduziu restrições adicionais ao acesso de desenvolvedores após a revelação pública de 2018. Essas mudanças representam um avanço técnico real na redução de uma vulnerabilidade específica. Contudo, elas não alteram as condições estruturais que tornaram o escândalo possível.

A lógica econômica que incentiva plataformas a abrir seus dados a desenvolvedores externos permanece intacta. A assimetria entre o volume de dados que as plataformas acumulam sobre seus usuários e a capacidade dos usuários de compreender e controlar esse acúmulo continua sendo uma característica estrutural do modelo de negócios das grandes plataformas. A dependência de mecanismos de consentimento formal como instrumento primário de proteção continua produzindo a ficção jurídica do usuário informado que lê termos de serviço e toma decisões autônomas sobre o uso de seus dados.

O que mudou, de forma mais significativa, foi o ambiente regulatório. O GDPR europeu, plenamente em vigor desde maio de 2018, introduziu um regime de responsabilização muito mais robusto, com multas de até 4% do faturamento global para violações graves. A LGPD brasileira, aprovada no mesmo ano, seguiu trajetória similar. Esses marcos regulatórios não eliminam as vulnerabilidades técnicas, mas alteram o cálculo de risco para plataformas e desenvolvedores que considerem explorá-las.

O legado mais duradouro do caso Cambridge Analytica, contudo, pode não ser técnico nem regulatório, mas epistemológico. Ele tornou amplamente compreensível, para além dos círculos especializados, que dados digitais aparentemente banais, como curtidas em publicações do Facebook, contêm informação suficiente para inferir dimensões profundas da personalidade e das preferências políticas dos indivíduos. Essa compreensão pública alterou permanentemente o debate sobre privacidade digital, tornando difícil

sustentar, como era comum antes de 2018, que dados comportamentais digitais são informações triviais sem relevância para a autonomia e a dignidade das pessoas.

⚖ JURISPRUDÊNCIA TSE — AIJE nº 0601346-36.2022.6.00.0000

EMENTA:

Ação de Investigação Judicial Eleitoral. Uso de empresa estrangeira de análise de dados para construção de perfis psicográficos de eleitores brasileiros sem consentimento adequado nos termos da LGPD. Targeting eleitoral baseado em dados sensíveis (opiniões políticas inferidas) sem base legal adequada. Impulsionamento pago não declarado como despesa de campanha.

RESULTADO:

O TSE reconheceu a competência para investigar o uso de dados pessoais em campanhas eleitorais em articulação com a LGPD, estabelecendo precedente relevante para a articulação entre as jurisdições da Justiça Eleitoral e da ANPD em matérias de proteção de dados aplicadas ao processo eleitoral.

■ RESUMO DO CAPÍTULO:

Este capítulo analisou o escândalo Cambridge Analytica como fenômeno simultaneamente técnico, organizacional, jurídico e normativo, recusando tanto a narrativa de onipotência algorítmica quanto a minimização de suas implicações democráticas. A análise demonstrou que o caso não deve ser lido como uma anomalia produzida por atores excêntricos operando nas margens do sistema, mas como a expressão sintomática de um ecossistema mais amplo, estruturalmente orientado à extração e mercantilização de dados comportamentais para fins de influência política. A revisão da literatura acadêmica revelou três linhas de investigação com pressupostos e implicações distintos. A análise técnica, representada pelos trabalhos de Guess, Nagler e Tucker, questiona a magnitude real do impacto eleitoral atribuído à empresa, ressuscitando em novo contexto a tese clássica de Lazarsfeld sobre os efeitos limitados da comunicação política de massa. A análise regulatória, ancorada nas contribuições de Solove e Rubinstein, demonstra que os frameworks tradicionais de proteção de dados baseados em consentimento individual são estruturalmente insuficientes para lidar com os riscos que emergem da agregação de dados em larga escala. A análise normativa, cujo eixo central é o enquadramento de Zuboff sobre capitalismo de vigilância, situa o escândalo não como desvio, mas como manifestação de uma lógica econômica que, quando aplicada ao processo eleitoral, compromete o mecanismo central pelo qual as democracias produzem legitimidade. A análise da arquitetura organizacional da Cambridge Analytica revelou uma empresa cuja cultura e métodos eram oriundos não do marketing político convencional, mas das operações de influência psicológica desenvolvidas pela SCL Group em contextos de segurança nacional e em eleições em países com menor capacidade regulatória. Esse histórico é central para compreender por que práticas que seriam inaceitáveis em contextos democráticos maduros foram aplicadas sem os filtros éticos e jurídicos que deveriam tê-las impedido. O exame do mecanismo técnico de coleta de dados expôs uma vulnerabilidade estrutural inscrita na própria arquitetura do Facebook, cuja lógica econômica incentivava a abertura de dados a desenvolvedores externos ao mesmo tempo em que criava condições para o abuso sistêmico desse acesso. A distinção entre o consentimento fraturado dos 270 mil usuários que instalaram o aplicativo de Kogan e a ausência absoluta de consentimento dos 87 milhões cujos dados foram acessados indiretamente é juridicamente fundamental e revela que o escândalo não envolve apenas violação de termos de uso, mas coleta de dados sem qualquer base legal identificável. O legado do caso para o direito eleitoral brasileiro é concreto e imediato. A LGPD, cuja aprovação foi acelerada em parte pelo impacto público do escândalo, qualifica opiniões políticas como dados sensíveis e exige consentimento específico e destacado para seu

tratamento. O precedente estabelecido pelo TSE na AIJE nº 0601346-36.2022.6.00.0000 reconheceu a competência da Justiça Eleitoral para investigar o uso de dados pessoais em campanhas eleitorais em articulação com a LGPD, abrindo caminho para a responsabilização de práticas similares às da Cambridge Analytica no contexto brasileiro. A questão central que o capítulo deixa em aberto é precisamente a que orienta a parte seguinte desta obra: em que medida o ordenamento jurídico brasileiro está equipado para responder, em tempo útil, a um ecossistema de influência que, como o caso demonstrou, se dissolve e se reconstitui mais rapidamente do que qualquer resposta regulatória convencional consegue alcançar.

QUESTÕES PARA REFLEXÃO E DEBATE

1. A tese de Guess, Nagler e Tucker sobre os efeitos limitados do microtargeting psicográfico sobre o comportamento eleitoral é metodologicamente sólida ou enfrenta limitações que comprometem suas conclusões? Em que medida a dificuldade de medir efeitos de influência subliminar pode subestimar sistematicamente o impacto real dessas práticas?

2. A teoria pragmática da privacidade de Solove, segundo a qual a privacidade é uma propriedade relacional que emerge de padrões de agregação e não de dados isolados, implica que os frameworks de consentimento individual são inadequados como instrumento primário de proteção de dados? Que modelo alternativo seria mais compatível com essa perspectiva?
3. O enquadramento de Zuboff sobre capitalismo de vigilância oferece uma explicação estrutural para o escândalo Cambridge Analytica, mas enfrenta a crítica de produzir uma narrativa de onipotência tecnológica que pode gerar desempoderamento cívico. Como calibrar a análise crítica das práticas de microtargeting sem incorrer nem no minimalismo que subestima riscos nem no alarmismo que paralisa a agência política?

4. A genealogia da Cambridge Analytica como subsidiária da SCL Group, especializada em operações de influência em contextos de baixa accountability regulatória, tem implicações para a avaliação da culpabilidade de seus dirigentes. Em que medida o histórico organizacional de uma empresa pode ser relevante para a responsabilização jurídica de suas condutas?

5. A distinção entre o consentimento fraturado dos usuários que instalaram o aplicativo de Kogan e a ausência absoluta de consentimento dos 87 milhões de usuários afetados indiretamente tem consequências jurídicas distintas sob a LGPD? Como o direito brasileiro deveria tratar essas duas categorias de titulares em um caso hipotético análogo?

6. O conceito de mercado da ilusão de precisão, utilizado para descrever o ambiente em que a Cambridge Analytica operava, sugere que campanhas políticas podem ser induzidas a investir em tecnologias cuja eficácia é sistematicamente exagerada. Que mecanismos regulatórios ou de mercado poderiam reduzir essa assimetria informacional entre fornecedores de tecnologia eleitoral e seus clientes políticos?

7. O precedente estabelecido pelo TSE na AIJE nº 0601346-36.2022.6.00.0000 reconheceu a articulação entre a jurisdição eleitoral e a LGPD. Quais são os limites e as potencialidades dessa articulação? Em que casos a competência da Justiça Eleitoral e da ANPD pode ser concorrente, e como resolver eventuais conflitos entre as duas esferas de responsabilização?

8. A dissolução formal da Cambridge Analytica não implicou o desaparecimento do modelo que ela representava. Que tipo de regulação seria necessário para alcançar o ecossistema, e não apenas os atores que o operacionalizam em determinado momento? O direito comparado oferece algum modelo promissor nessa direção?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. A Cambridge Analytica não foi uma exceção. Foi um sintoma. E o sintoma persiste.

A empresa encerrou operações em maio de 2018. Parte de sua liderança migrou para novas organizações com nomes diferentes, mas propostas de valor estruturalmente idênticas. O ecossistema de data brokers que fornecia dados continuou operando. As plataformas continuaram existindo. E a demanda política por instrumentos de persuasão baseados em dados comportamentais não apenas persistiu como se intensificou. Para profissionais de campanha, isso significa que as práticas expostas pelo escândalo não desapareceram: apenas ficaram menos visíveis. Para assessores jurídicos, significa que o risco de responsabilização por práticas similares é real e crescente, especialmente com a LGPD em vigor e a jurisprudência do TSE se consolidando.

2. O consentimento que seu usuário deu não cobre o que você está fazendo com os dados dele.

O escândalo Cambridge Analytica foi possível porque 270 mil usuários consentiram com algo que abriu o acesso a dados de 87 milhões de pessoas que não consentiram com nada. Esse mecanismo não existe mais no Facebook, mas a lógica subjacente persiste em outras formas: consentimentos genéricos que cobrem finalidades específicas não informadas, bases de dados compradas de terceiros sem verificação da proveniência, dados comportamentais usados para inferir opiniões políticas sem que o usuário soubesse que estava fornecendo essa informação. Sob a LGPD, cada uma dessas situações é uma exposição jurídica. O assessor que não verifica a cadeia de custódia dos dados que sua campanha utiliza é solidariamente responsável pelas irregularidades que essa cadeia contém.

3. A imprecisão do microtargeting não é uma defesa jurídica.

A evidência acadêmica sugere que o impacto eleitoral efetivo da Cambridge Analytica foi provavelmente inferior ao que a empresa alardeava e ao que a narrativa jornalística sugeriu. Isso não tem relevância jurídica para a avaliação das condutas. A coleta de dados sem consentimento é ilegal independentemente de ter produzido mudança de comportamento eleitoral. O tratamento de dados sensíveis sem base legal é ilegal independentemente de ter funcionado. A tentativa de manipular eleitores por meio de mensagens calibradas para explorar vulnerabilidades psicológicas é eticamente problemática independentemente de sua eficácia. A ineficácia não exonera; ela apenas reduz, eventualmente, a gravidade dos efeitos para fins de quantificação de danos.

4. Voter suppression digital é crime no Brasil. E é mais fácil de executar do que parece.

Uma das dimensões menos discutidas do caso Cambridge Analytica foi a estratégia de voter suppression: identificar eleitores potencialmente favoráveis ao adversário e enviar-lhes conteúdo destinado a desestimulá-los a comparecer às urnas. No Brasil, o artigo 222-A da Lei 9.504/1997 proíbe disparos em massa para fins ilícitos, e o desestímulo fraudulento ao comparecimento pode configurar crime eleitoral. Com as ferramentas de segmentação disponíveis hoje, uma campanha pode identificar com razoável precisão os eleitores menos motivados do adversário e direcionar narrativas desanimadoras especificamente a eles, sem que isso seja visível para ninguém além do emissor e do receptor. O monitoramento desse tipo de estratégia é um dos maiores desafios operacionais da fiscalização eleitoral contemporânea.

5. O TSE tem competência para investigar o uso de dados pessoais em campanhas. Use isso.

O precedente estabelecido na AIJE 0601346-36.2022.6.00.0000 reconheceu a competência da Justiça Eleitoral para investigar o uso de dados pessoais em campanhas eleitorais em articulação com a LGPD. Isso significa que campanhas que identificarem indícios de que o adversário está utilizando dados psicográficos obtidos sem consentimento, comprando listas de data brokers irregulares ou construindo perfis de opiniões políticas sem base legal têm um caminho institucional disponível: a representação perante o TSE ou os Tribunais Regionais Eleitorais, que podem acionar a ANPD de forma coordenada. A articulação entre as duas esferas ainda está sendo construída, mas o precedente existe e pode ser invocado.

6. A SCL Group operou em países com menos regulação antes de chegar às democracias consolidadas. Esse padrão ainda existe.

As técnicas que a Cambridge Analytica aplicou nas eleições norte-americanas de 2016 foram testadas e refinadas pela SCL Group em eleições de países com menor capacidade regulatória e menor resistência institucional. Esse padrão de desenvolvimento de ferramentas de influência em ambientes permissivos para posterior aplicação em mercados mais regulados não é exclusivo desse caso. Empresas de consultoria eleitoral que operam no Brasil e que têm histórico de atuação em eleições de países com regulação fraca merecem due diligence reforçada sobre seus métodos, suas fontes de dados e sua conformidade com o marco jurídico brasileiro. A origem das ferramentas importa para a avaliação dos riscos que elas carregam.

7. O que você não sabe sobre os dados que está usando pode custar o mandato.

A Cambridge Analytica sabia, ou tinha condições de saber, que os dados que adquiriu de Kogan haviam sido obtidos em violação dos termos de uso do Facebook. A ausência de due diligence sobre a proveniência dos dados foi tratada como evidência de cumplicidade, não como excludente de responsabilidade. No Brasil, o assessor de campanha que contrata uma empresa de análise de dados sem verificar sua conformidade com a LGPD, sem exigir documentação sobre a base legal do tratamento de dados sensíveis e sem auditar a proveniência das bases utilizadas está assumindo um risco jurídico que pode se materializar tanto em sanções da ANPD quanto em ação eleitoral com efeitos sobre a validade do mandato. Ignorância sobre o que o fornecedor faz com os dados não é defesa: é negligência.

CAPÍTULO 6

Bots, Chatbots e Avatares Digitais

Automação Política, Simulação de Presença e os Limites da Identidade Digital nas Eleições

Automação Política e Teoria Democrática

O fenômeno da automação política por agentes digitais e seu impacto sobre a qualidade do debate democrático constitui um campo de investigação acadêmica em rápida maturação, cuja trajetória é marcada pela tensão entre a velocidade com que as práticas evoluem e o ritmo necessariamente mais lento com que a pesquisa empírica e a reflexão normativa conseguem acompanhá-las.

Essa defasagem não é trivial: ela significa que boa parte da literatura disponível descreve um fenômeno que, ao tempo de sua publicação, já havia se transformado de forma significativa. Compreender essa limitação estrutural é o primeiro passo para uma leitura crítica e produtiva do que a pesquisa nessa área já sabe e do que ainda permanece em aberto.

A literatura organiza-se em três linhas de investigação com pressupostos metodológicos e implicações normativas distintos, que se complementam sem se sobrepor inteiramente.

A Linha Empírica: Mapeando a Presença e o Impacto dos Bots

A primeira linha é a empírica, que mapeia a presença e o impacto de agentes automatizados em plataformas digitais durante períodos eleitorais. Os trabalhos fundadores dessa linha, publicados entre 2016 e 2017, estabeleceram as bases metodológicas e os primeiros diagnósticos substantivos sobre a escala do fenômeno.

Ferrara et al. (2016) produziram o estudo seminal sobre bots do Twitter, identificando padrões comportamentais que distinguem contas automatizadas de contas humanas e estimando que entre 9% e 15% dos usuários ativos do Twitter à época eram, na verdade, bots. Esse resultado, embora sujeito a revisões metodológicas posteriores, teve impacto duradouro por tornar visível, com precisão estatística, a escala da presença de agentes não humanos em plataformas que se apresentavam como espaços de conversação entre pessoas reais.

Howard e Kollanyi (2016), em análise focada no referendo do Brexit e nas eleições norte-americanas de 2016, demonstraram que uma parcela desproporcional do conteúdo político de alta circulação no Twitter era gerada por um número relativamente pequeno de

contas automatizadas altamente ativas. Sua contribuição mais significativa foi metodológica: eles demonstraram que a influência dos bots não se mede pelo número de contas, mas pela concentração de atividade, e que um conjunto pequeno de bots bem posicionados pode exercer influência desproporcionalmente grande sobre a percepção pública de quais temas e narrativas estão dominando o debate.

Bessi e Ferrara (2016), ao examinar especificamente a eleição presidencial norte-americana de 2016, identificaram que cerca de 19% da discussão política no Twitter no período pré-eleitoral era gerada por bots, e que esses agentes automatizados estavam distribuídos de forma relativamente equilibrada entre as duas principais correntes políticas, embora com padrões de comportamento distintos. Esse resultado tensionou a narrativa simplista de que os bots eram instrumentos exclusivos de um lado do espectro político e introduziu a necessidade de análises mais nuançadas sobre quem os opera, com quais objetivos e com quais efeitos específicos sobre diferentes segmentos do eleitorado.

A linha empírica enfrenta, contudo, uma limitação estrutural que os próprios pesquisadores reconhecem: os métodos de detecção de bots são continuamente desafiados pela evolução das técnicas de automação. À medida que os sistemas de detecção se tornam mais sofisticados, os operadores de bots adaptam seus sistemas para simular comportamento humano com maior fidelidade. Esse ciclo de corrida armamentista entre detecção e evasão significa que os dados empíricos disponíveis tendem a subestimar a presença real de agentes automatizados, na medida em que os mais sofisticados são precisamente os que escapam à detecção.

A Linha Técnica: Construção, Detecção e Classificação

A segunda linha é a técnica, que descreve os métodos de construção, detecção e classificação de bots políticos. O trabalho de Varol et al. (2017) é a referência central nessa linha, ao propor um sistema de detecção baseado em aprendizado de máquina que integra múltiplas dimensões do comportamento de uma conta, incluindo frequência de publicação, padrões de interação, características do perfil e análise linguística do conteúdo produzido.

A contribuição técnica dessa linha vai além da detecção: ela mapeia a arquitetura dos sistemas de automação política, revelando que a sofisticação das operações de bot varia enormemente, desde sistemas rudimentares que publicam conteúdo em intervalos regulares até arquiteturas complexas que utilizam modelos de linguagem para gerar texto original, sistemas de análise de sentimento para calibrar o tom das respostas e algoritmos de agendamento para simular os padrões temporais do comportamento humano.

Essa variabilidade técnica tem implicações diretas para a análise jurídica: ela significa que não existe um único fenômeno que possa ser denominado "uso de bots", mas um espectro de práticas com diferentes graus de sofisticação, diferentes capacidades de impacto e diferentes perfis de responsabilização. Uma conta que publica automaticamente links para notícias é tecnicamente um bot, mas seu funcionamento e seus efeitos são radicalmente distintos dos de um sistema que gera respostas originais em linguagem natural para interagir com eleitores reais em tempo real. O direito eleitoral, ao regular o uso de automação, precisa ser suficientemente preciso para capturar essa diversidade sem criar normas que, por serem excessivamente amplas, proíbam práticas legítimas de automatização de processos comunicacionais.

A Linha Normativa: Automação e os Fundamentos da Democracia

A terceira linha é a normativa, e é nela que o debate alcança sua maior profundidade teórica. Sua questão central pode ser formulada nos seguintes termos: o que significa, para a qualidade da democracia, que uma parcela significativa do que circula como opinião pública em plataformas digitais seja gerada por agentes não humanos, operados por atores com interesses específicos e sem qualquer transparência sobre sua natureza artificial?

Yochai Benkler, em *A Riqueza das Redes* (2006), oferece o contraponto teórico mais influente para compreender a gravidade dessa questão. Benkler argumenta que a comunicação digital em rede cria, pela primeira vez na história, condições estruturais para uma esfera pública genuinamente descentralizada, na qual qualquer indivíduo pode produzir e distribuir conteúdo em escala comparável à de organizações com recursos muito superiores. Essa promessa democrática, no entanto, pressupõe que os participantes do espaço público digital sejam, de fato, participantes: agentes com perspectivas, interesses e razões próprias, cuja contribuição ao debate enriquece a deliberação coletiva precisamente porque representa uma voz genuína.

Os bots políticos subvertem essa premissa de forma estrutural. Ao inserir no espaço público digital vozes que não representam nenhum indivíduo real, nenhuma perspectiva genuína e nenhum interesse autêntico, eles distorcem o sinal mais fundamental que a esfera pública democrática precisa ser capaz de processar: a distribuição real de opiniões na sociedade. Quando eleitores percebem que determinada narrativa domina o debate nas redes sociais, eles inferem, de forma razoável, que essa dominância reflete uma distribuição real de preferências. Se essa percepção é artificialmente construída por agentes automatizados, o eleitor está sendo induzido a uma leitura falsa da realidade social,

com consequências potencialmente significativas sobre sua própria disposição a expressar opiniões, a votar e a perceber a legitimidade do resultado eleitoral.

Esse fenômeno, que a literatura denomina percepção de norma social artificialmente manipulada, é particularmente relevante porque opera sobre um mecanismo cognitivo profundamente enraizado no comportamento humano: a tendência a ajustar posições e comportamentos com base no que se percebe como opinião majoritária do grupo de referência. Quando essa percepção é sistematicamente distorcida por automação, o efeito sobre a deliberação democrática é análogo ao de uma pesquisa de opinião fraudulenta exibida em escala massiva: ele não apenas informa incorretamente, mas altera ativamente o objeto que pretende medir.

A linha normativa enfrenta, contudo, uma tensão que os pesquisadores mais honestos reconhecem: a dificuldade de distinguir, no plano conceitual, entre automação política ilegítima e formas legítimas de amplificação de mensagens políticas que sempre existiram na política. Partidos políticos sempre mobilizaram militantes para distribuir material de campanha, defender candidatos em debates públicos e amplificar mensagens. A diferença introduzida pelos bots não é, portanto, a existência de amplificação organizada, mas a ausência de qualquer vínculo entre a amplificação e agentes humanos reais. É precisamente essa ruptura do vínculo entre voz e agente que constitui o núcleo do problema democrático, e é ela que precisa ser capturada com precisão pela regulação.

Taxonomia dos Agentes Automatizados

A literatura sobre automação política utiliza nomenclatura nem sempre consistente, o que introduz ambiguidade tanto na análise acadêmica quanto no debate jurídico. Expressões como "bot", "conta falsa", "perfil automatizado" e "agente artificial" são frequentemente usadas de forma intercambiável em contextos nos quais as distinções técnicas têm consequências relevantes para a avaliação jurídica das condutas. Para os propósitos desta obra, fixam-se quatro definições precisas, organizadas de acordo com o grau crescente de sofisticação técnica e complexidade normativa.

O bot político, em sentido estrito, é uma conta em plataforma de rede social operada total ou predominantemente por software, projetada para publicar, compartilhar, curtir ou responder a conteúdo político de forma autônoma, sem intervenção humana individualizada em cada ação. Em sua forma mais simples, o bot político é um sistema de automação que executa rotinas predefinidas: publica determinados tipos de conteúdo em horários específicos, interage com determinadas contas segundo critérios algorítmicos e amplifica

narrativas de acordo com palavras-chave ou hashtags programadas. Sua identificação é relativamente simples quando opera de forma rudimentar, pois seus padrões de comportamento, como frequência de publicação incompatível com atividade humana ou ausência de variação na linguagem, são detectáveis por sistemas automatizados de análise.

O cyborg digital representa um grau superior de sofisticação e um desafio analítico e jurídico substancialmente maior. Trata-se de uma conta que combina operação humana com amplificação automatizada: um agente humano real produz ou seleciona conteúdo, e sistemas automatizados amplificam esse conteúdo de forma coordenada, simulando engajamento orgânico. A detecção do cyborg é consideravelmente mais complexa do que a do bot puro, precisamente porque os padrões de comportamento humano presentes na operação mascaram a automação subjacente. Do ponto de vista jurídico, o cyborg levanta questões adicionais sobre responsabilização: a presença de um operador humano identificável torna mais difícil sustentar que a conta é "automatizada" em sentido estrito, mas a automação que amplifica suas mensagens pode configurar as mesmas irregularidades eleitorais que o bot puro.

O chatbot eleitoral é um sistema de diálogo automatizado projetado para interagir com eleitores em linguagem natural, simulando uma conversa com um agente humano. Diferentemente do bot político tradicional, que opera de forma unidirecional, o chatbot é concebido para responder a perguntas, fornecer informações sobre candidatos e programas, mobilizar eleitores para o comparecimento e, em algumas implementações mais sofisticadas, conduzir conversas de persuasão individualizadas com base no perfil do interlocutor.

A introdução de modelos de linguagem de grande escala transformou radicalmente as capacidades dos *chatbots* eleitorais: sistemas que até recentemente produziam respostas visivelmente artificiais passaram a ser capazes de manter conversas indistinguíveis das humanas em qualidade linguística. Essa transformação coloca em primeiro plano a questão da identificação obrigatória: o eleitor que interage com um chatbot tem o direito de saber que está conversando com uma máquina, e a omissão dessa informação configura, sob a Resolução TSE nº 23.732/2024, uma irregularidade eleitoral.

O avatar digital é a categoria tecnicamente mais avançada e normativamente mais perturbadora. Trata-se de uma representação sintética de um candidato real, construída com tecnologias de síntese de imagem e voz, capaz de gerar conteúdo audiovisual no qual o candidato aparentemente se manifesta, sem que tenha efetivamente se manifestado. O avatar digital pode reproduzir a aparência física, a voz, os gestos e os padrões de linguagem

do candidato com fidelidade suficiente para enganar observadores não treinados. Em sua versão mais sofisticada, combinada com modelos de linguagem de grande escala, o avatar pode responder a perguntas, adaptar seu discurso a diferentes contextos e produzir conteúdo original que o candidato real nunca produziu. A vedação expressa de deepfakes e simulações fraudulentas de candidatos na Resolução TSE nº 23.732/2024 é uma resposta direta a essa categoria, cujo potencial de dano à integridade do processo eleitoral é o mais elevado dentre todos os agentes automatizados aqui descritos.

A distinção entre essas quatro categorias não é apenas taxonômica: ela tem consequências diretas para a avaliação jurídica das condutas, para a identificação dos responsáveis e para a definição das sanções aplicáveis. Uma campanha que utiliza bots para amplificar conteúdo enfrenta um perfil de responsabilização diferente de uma campanha que produz avatares digitais fraudulentos do candidato adversário. Tratar essas situações como instâncias do mesmo fenômeno, como a nomenclatura imprecisa frequentemente induz, é um erro analítico com consequências práticas relevantes tanto para a acusação quanto para a defesa em processos eleitorais.

Tabela 26. Tipologia Completa dos Agentes Automatizados Políticos

Tipo	Definição	Capacidade Operacional	Detectabilidade	Risco Eleitoral Principal
Bot simples	Conta 100% automatizada por script simples. Sem intervenção humana.	Volume massivo (milhares de posts/hora), baixa adaptabilidade.	Alta: padrões temporais, de linguagem e de comportamento anômalos.	Inflação artificial de trending topics, métricas e percepção de popularidade.
Bot sofisticado (ML-based)	Automação com modelos de linguagem: responde a contexto, aprende padrões humanos, adapta comportamento.	Volume alto e adaptabilidade: pode manter conversas convincentes, variar horários, simular comportamento humano.	Média a baixa: modelos avançados são difíceis de distinguir de humanos.	Disseminação de desinformação convincente, criação de consenso artificial.
Cyborg	Conta humana que utiliza ferramentas de automação para amplificar alcance: agendamento automático,	Autenticidade humana mais eficiência da automação: maior credibilidade que bots puros.	Baixa: parece e ocasionalmente é humano; a automação é parcial.	Astroturfing credível, infiltração de grupos, coordenação inautêntica de

	respostas semiautomáticas, templates.			aparência espontânea.
Sockpuppet	Identidade falsa completamente operada por humano real, sem automação técnica.	Alta adaptabilidade, baixo volume: um humano opera como outra pessoa.	Muito baixa quando bem construída.	Infiltração de grupos adversários, criação de dissidência falsa, manipulação de líderes de opinião.
Troll farm	Operação organizada de múltiplos operadores humanos coordenados, cada um com múltiplos perfis.	Escala industrial com qualidade humana: combinação mais perigosa.	Muito baixa: operação profissional com procedimentos definidos.	Campanhas de influência, polarização sistemática, criação de narrativas falsas persistentes.

Fonte: Elaborado pelo autor com base Ferrara et al. (2016).

Bots Políticos: Funcionamento Técnico e Detecção

Compreender o funcionamento técnico de um bot político sofisticado é condição necessária para avaliá-lo juridicamente com precisão. A tendência de tratar bots como uma categoria homogênea, definida apenas pela presença de automação, obscurece diferenças técnicas que têm implicações diretas sobre a capacidade de impacto, a detectabilidade e o perfil de responsabilização jurídica de cada sistema.

Um bot que publica links automaticamente a partir de um feed RSS e um sistema que gera respostas originais em linguagem natural, adapta seu comportamento ao contexto em tempo real e coordena centenas de contas para maximizar amplificação são tecnicamente distintos ao ponto de serem, na prática, fenômenos diferentes que merecem tratamento analítico e jurídico diferenciado.

Arquitetura de um Bot Político Sofisticado

A construção de um bot político sofisticado envolve quatro componentes técnicos interdependentes, cada um com função específica na cadeia de operação do sistema. O primeiro componente é o mecanismo de coleta e monitoramento ambiental, responsável

por rastrear em tempo real o ambiente informacional das plataformas em que o bot opera. Esse componente utiliza as APIs das plataformas para monitorar hashtags, tópicos em tendência, palavras-chave relevantes e o comportamento de contas específicas, produzindo um fluxo contínuo de dados sobre o estado atual do debate público. A sofisticação desse componente determina a capacidade de resposta do sistema: bots rudimentares monitoram palavras-chave fixas predefinidas; sistemas avançados utilizam modelos de processamento de linguagem natural para identificar tópicos emergentes e avaliar seu potencial de engajamento antes de decidir se e como intervir.

O segundo componente é o módulo de geração de conteúdo, que traduz os dados coletados pelo primeiro componente em publicações, respostas e interações. Em sua versão mais simples, esse módulo opera com *templates* predefinidos, preenchendo variáveis com informações extraídas do monitoramento. Em versões intermediárias, utiliza técnicas de parafraseamento para variar a formulação de mensagens recorrentes, reduzindo a detectabilidade por sistemas que identificam bots pela repetição literal de conteúdo.

Em sua versão mais avançada, o módulo de geração integra modelos de linguagem de grande escala que produzem texto original, contextualmente adequado e linguisticamente indistinguível de conteúdo humano. Essa última versão representa uma ruptura qualitativa em relação às anteriores: ela elimina o principal indicador técnico de automação que os sistemas de detecção historicamente utilizavam, que era a repetição ou a baixa qualidade linguística do conteúdo produzido.

O terceiro componente é o módulo de gestão de identidade, responsável por criar e manter a aparência de contas humanas autênticas. Esse componente inclui a geração de fotos de perfil convincentes por meio de redes neurais generativas adversariais, conhecidas como GANs, que produzem imagens de rostos humanos que nunca existiram e que são praticamente indistinguíveis de fotografias reais. Inclui também a construção de histórico de atividade simulando os ritmos do comportamento humano: variação nos horários de publicação, alternância entre conteúdo político e pessoal, padrões de engajamento compatíveis com os de uma conta humana ativa. Em operações sofisticadas, esse componente é alimentado por perfis psicográficos que definem a "personalidade" de cada conta do bot, garantindo coerência temática e estilística ao longo do tempo.

O quarto componente é o módulo de distribuição e amplificação, que coordena múltiplas contas para maximizar o impacto de narrativas específicas. Esse componente implementa estratégias de amplificação coordenada: contas do sistema interagem entre si para criar a aparência de engajamento orgânico, elevam determinados conteúdos às listas

de tendências por meio de curtidas e compartilhamentos coordenados e criam a ilusão de que determinadas opiniões são amplamente compartilhadas quando na verdade estão sendo artificialmente amplificadas. A coordenação entre as contas é gerenciada de forma a simular comportamento independente: as interações são distribuídas ao longo do tempo e variadas em tipo para evitar os padrões de sincronia que os sistemas de detecção utilizam como indicadores de coordenação inautêntica.

Métodos de Detecção e suas Limitações

A detecção de bots políticos baseia-se na análise de padrões de comportamento que divergem do que é esperado de usuários humanos. Os principais indicadores utilizados pelos sistemas de detecção organizam-se em quatro dimensões. A dimensão temporal analisa a frequência e a regularidade das publicações. Contas humanas apresentam variação natural nos horários de atividade, com períodos de inatividade compatíveis com sono, trabalho e outras atividades da vida cotidiana.

Bots rudimentares publicam com frequência e regularidade incompatíveis com esse padrão, operando em horários em que usuários humanos raramente estão ativos ou mantendo cadências de publicação uniformes que não apresentam a variabilidade característica do comportamento humano. Bots sofisticados, contudo, incorporam deliberadamente essa variabilidade, tornando a análise temporal insuficiente como critério isolado de detecção.

A dimensão relacional examina a estrutura das redes de seguidores e seguidos e os padrões de interação. Contas de bot frequentemente apresentam relações desproporcionais entre seguidores e seguidos, histórico de interações limitado a um conjunto restrito de contas e ausência das relações recíprocas e diversificadas que caracterizam redes sociais humanas autênticas. A análise de grafos de rede pode identificar clusters de contas que interagem predominantemente entre si, sugerindo coordenação inautêntica. Operações sofisticadas, contudo, investem na construção de redes de aparência orgânica, estabelecendo conexões com contas humanas reais ao longo do tempo para diluir os indicadores de artificialidade.

A dimensão linguística analisa o conteúdo das publicações em busca de padrões inconsistentes com a produção humana. Repetição literal ou quase-literal de formulações, ausência de erros tipográficos e variações linguísticas naturais, coerência temática excessiva e ausência de conteúdo pessoal ou contextualmente situado são indicadores clássicos de automação. A introdução de modelos de linguagem de grande escala

comprometeu seriamente a eficácia dessa dimensão: sistemas baseados em GPT-4 ou Claude produzem texto com qualidade linguística, variação estilística e contextualização situacional indistinguíveis da produção humana, eliminando os marcadores que os sistemas de detecção linguística historicamente identificavam.

A dimensão comportamental examina padrões de resposta a eventos e a velocidade de reação a conteúdo novo. Bots que respondem a eventos em intervalos de tempo incompatíveis com a leitura e o processamento humano, ou que apresentam velocidade de resposta uniforme independentemente do horário, revelam sua natureza automatizada por esse indicador. Sistemas avançados incorporam delays variáveis e simulam tempo de leitura antes de responder, tornando também essa dimensão insuficiente como critério isolado.

A consequência prática dessas limitações é que a detecção confiável de bots sofisticados requer a combinação de múltiplas dimensões de análise e o acesso a dados que as plataformas frequentemente não disponibilizam a pesquisadores independentes ou a autoridades eleitorais. Essa assimetria de acesso entre operadores de bots e sistemas de detecção é, ela própria, um problema de governança que nenhuma solução técnica isolada pode resolver: ela exige obrigações legais de transparência e acesso a dados que marcos regulatórios como o DSA europeu começam a introduzir, mas que a maioria dos sistemas eleitorais ainda não foi capaz de implementar de forma efetiva.

Chatbots Eleitorais: Usos Lícitos e Riscos de Simulação de Identidade

O chatbot eleitoral é a categoria de agente automatizado com maior ambivalência normativa, precisamente porque seus usos legítimos e seus usos problemáticos compartilham a mesma infraestrutura técnica e frequentemente se distinguem apenas pela intenção do operador e pelo design específico do sistema. Essa ambivalência exige uma abordagem analítica que não se limite a proibir ou permitir a categoria em abstrato, mas que identifique com precisão as características que distinguem um uso aceitável de um uso manipulativo.

Os usos lícitos dos chatbots eleitorais são substantivos e merecem reconhecimento antes da análise de seus riscos. O atendimento automatizado a eleitores com informações factuais sobre propostas de governo, posições do candidato e logística eleitoral, como locais de votação, documentos necessários e procedimentos para justificativa de ausência, representa um ganho de eficiência comunicacional que não implica qualquer comprometimento da integridade do processo eleitoral.

A mobilização de voluntários e apoiadores por meio de sistemas automatizados de agendamento e coordenação é igualmente legítima. A coleta estruturada de feedback de eleitores, quando realizada com consentimento informado e sem uso dos dados para fins não declarados, pode contribuir para campanhas mais responsivas às preocupações do eleitorado. Todos esses usos são lícitos desde que a natureza automatizada do sistema seja clara e imediatamente identificável para o eleitor que interage com ele.

Os usos problemáticos concentram-se em três categorias. A primeira é a simulação de interação direta com o próprio candidato, que induz o eleitor a acreditar que está em contato com um ser humano real, e especificamente com a figura pública que representa a candidatura. Essa simulação compromete a autonomia decisória do eleitor de forma particularmente grave, porque explora o capital de confiança e de proximidade que a percepção de contato direto com um candidato gera.

A segunda é a coleta de dados pessoais dos eleitores sem consentimento informado, que transforma a interface de atendimento em instrumento de perfilamento. A terceira é a disseminação de desinformação por meio de respostas incorretas ou enganosas, que pode ocorrer tanto por design deliberado quanto por limitações técnicas do sistema que o operador não corrigiu adequadamente.

Framework Ético para Chatbots Eleitorais: Seis Critérios

A questão central não é se chatbots devem ser permitidos em campanhas eleitorais. Eles já estão sendo utilizados de forma crescente, e há argumentos legítimos para seu uso em funções informativas e logísticas. A questão é como devem ser desenhados e operados para que o ganho de eficiência que proporcionam não se converta em instrumento de manipulação. A fronteira entre uso aceitável e uso problemático pode ser delimitada por seis critérios fundamentais, cuja observância cumulativa é condição necessária para que um chatbot eleitoral seja considerado compatível com os princípios de integridade do processo democrático e com o marco regulatório vigente.

O primeiro critério é o da identificação plena e imediata. O sistema deve identificar-se como automatizado no primeiro contato e em toda interação subsequente em que essa natureza possa não estar clara para o interlocutor. Formulações como "Sou o assistente virtual da campanha X, um sistema automatizado” são necessárias, mas insuficientes se apresentadas de forma a não ser percebidas: a identificação deve ser destacada, posicionada no início da interação e formulada em linguagem inequívoca, não enterrada em texto longo ou apresentada em fonte reduzida. Sob a Resolução TSE nº 23.732/2024,

a ausência de identificação adequada configura irregularidade eleitoral com consequências sancionatórias para a campanha responsável pelo sistema.

O segundo critério é o do conteúdo verificável e não manipulativo. O chatbot deve limitar suas respostas a informações verificáveis sobre o programa de governo, as posições do candidato e a logística eleitoral. Não deve utilizar técnicas de urgência artificial, escassez, apelos emocionais programados ou padrões de design que induzam o eleitor a tomar decisões sem reflexão adequada. Essa restrição não proíbe que o chatbot apresente o candidato de forma favorável: campanhas legítimas podem e devem comunicar suas posições de forma persuasiva. O que é vedado é a utilização de mecanismos que contornem a capacidade reflexiva do eleitor, explorando vulnerabilidades cognitivas identificadas por modelagem psicográfica.

O terceiro critério é o da ausência de simulação de empatia ou vínculo emocional. Sistemas de inteligência artificial generativa contemporâneos são capazes de simular empatia, compreensão e vínculo emocional com convincente fidelidade. Essa capacidade, que tem aplicações legítimas em contextos terapêuticos e educacionais, é eticamente inaceitável em contexto eleitoral, precisamente porque explora a tendência humana de avaliar positivamente e de votar em quem percebe como próximo, compreensivo e genuinamente interessado em suas preocupações. Um chatbot que finge compartilhar as angústias do eleitor, celebrar suas conquistas ou preocupar-se com seu bem-estar individual não está exercendo persuasão legítima: está manipulando uma resposta emocional cuja natureza artificial o eleitor não tem condições de identificar.

O quarto critério é o do consentimento informado para coleta de dados. Toda coleta de dados pessoais durante a interação com o chatbot deve ser precedida de consentimento livre, específico e informado, nos termos da LGPD. Dados coletados durante interações com chatbot de campanha que revelem, explicita ou implicitamente, a opinião política do eleitor traz consigo dados sensíveis para os fins do artigo 5º, II da LGPD, e seu tratamento sem consentimento específico configura violação com potencial de responsabilização tanto perante a ANPD quanto perante a Justiça Eleitoral. O design do chatbot deve tornar clara, antes de qualquer coleta, a finalidade específica do uso dos dados e as consequências da recusa em fornecê-los.

O quinto critério é o do direito à transferência humana. O eleitor deve sempre poder solicitar atendimento por um agente humano real, e essa solicitação deve ser atendida em prazo compatível com a natureza da demanda. A recusa a identificar-se como sistema automatizado quando questionado diretamente, ou a recusa a transferir para um humano quando solicitado, constitui violação ética grave e potencialmente ilegal, na medida em que

priva o eleitor do direito de saber com quem está interagindo e de escolher a modalidade de interação que prefere. Campanhas que operam chatbots sem capacidade de transferência humana ou sem disponibilizar essa informação ao eleitor assumem um risco jurídico significativo.

O sexto critério é o da auditabilidade e preservação de registros. Todas as interações do chatbot com eleitores devem ser registradas em logs completos, preservados pelo período mínimo estabelecido pela regulação eleitoral vigente e disponibilizados às autoridades eleitorais competentes mediante requisição fundamentada. A opacidade das conversas de chatbot é estruturalmente incompatível com a transparência que o processo eleitoral requer: se uma campanha utiliza um sistema automatizado para interagir com eleitores em escala, as autoridades de fiscalização precisam ter condições de verificar o que foi dito, a quem foi dito e com quais efeitos verificáveis. A ausência de registros auditáveis não é apenas um problema regulatório: é um indicador de que o sistema foi desenhado para operar fora do alcance do controle institucional, o que é, em si mesmo, um indício de irregularidade.

A observância cumulativa desses seis critérios não garante que um chatbot eleitoral seja isento de críticas éticas mais amplas. Há perspectivas que questionam a legitimidade do uso de qualquer sistema automatizado em interações que o eleitor percebe como pessoais, independentemente de qualquer identificação formal. Essa perspectiva tem fundamento: mesmo um chatbot que se identifica claramente como sistema automatizado pode produzir efeitos de vínculo e de confiança que não seriam produzidos se o eleitor soubesse que está interagindo com um algoritmo. O que os seis critérios oferecem é um threshold mínimo de aceitabilidade, abaixo do qual o uso de chatbots eleitorais é claramente incompatível com os princípios democráticos, e acima do qual o debate sobre seus limites permanece em aberto.

⚖ JURISPRUDÊNCIA TSE — Representação TSE nº 0604567-22.2024.6.00.0000

EMENTA:

Representação por propaganda eleitoral irregular. Uso de chatbot que simulava interação direta com o candidato, respondendo a eleitores em primeira pessoa sem identificação como sistema automatizado. Coleta de dados de eleitores sem consentimento adequado para os fins da LGPD.

RESULTADO:

O TSE determinou a imediata cessação do funcionamento do chatbot, a eliminação dos dados coletados sem consentimento e aplicou multa ao candidato pela propaganda eleitoral irregular. O caso estabeleceu o entendimento de que a simulação de identidade humana por chatbot eleitoral configura propaganda irregular independentemente da vontade do candidato de ser pessoalmente percebido como interlocutor.

Avatares Digitais e a Regulação da Identidade Sintética

O avatar digital é a categoria de agente automatizado com o regime jurídico mais complexo e, ao mesmo tempo, com o maior potencial de dano à integridade do processo eleitoral. Sua complexidade normativa decorre de uma característica técnica que o distingue fundamentalmente das demais categorias: enquanto bots e chatbots são sistemas que atuam em nome de uma campanha ou de um ator político identificável, o avatar digital é um sistema que atua sob a identidade de uma pessoa real, simulando sua aparência, sua voz e sua expressão de forma suficientemente convincente para induzir o eleitor a acreditar que está diante de uma manifestação genuína do candidato representado. Essa diferença não é apenas de grau: é de natureza. O problema que o avatar coloca não é apenas o da automação não identificada, mas o da falsificação de identidade política em escala industrial.

O Regime Jurídico: Autorização, Rotulagem e Vedações

A Resolução TSE nº 23.732/2024 estabelece para os avatares digitais um regime de autorização condicionada, que parte do reconhecimento de que a tecnologia tem usos legítimos no contexto eleitoral e que sua proibição absoluta seria desproporcional e de difícil enforcement. O regime estrutura-se em torno de três exigências cumulativas.

A primeira exigência é a autorização expressa do candidato representado. O uso de um avatar digital que reproduz a aparência e a voz de um candidato real depende de consentimento explícito e documentado do próprio candidato ou de seu representante legal autorizado. Essa exigência é tecnicamente óbvia, mas juridicamente fundamental: ela estabelece que a identidade de um candidato não é um bem público livremente utilizável por qualquer ator, incluindo sua própria campanha, sem um ato formal de consentimento. A ausência de autorização não é uma irregularidade processual menor: é o elemento que transforma o uso do avatar de prática permitida em infração eleitoral.

A segunda exigência é a rotulagem clara e imediata do conteúdo como representação sintética gerada por inteligência artificial. Essa exigência opera sobre o mesmo fundamento que orienta toda a política de identificação de conteúdo gerado por IA da Resolução: o eleitor tem o direito de saber, antes de processar qualquer mensagem política, se o que está recebendo é uma manifestação genuína do candidato ou uma representação sintética produzida por sistemas automatizados. A rotulagem não deve ser

apresentada em formato que minimize sua visibilidade: legendas em fonte pequena, avisos transitórios que aparecem por frações de segundo ou identificações inseridas apenas no código-fonte da publicação não atendem ao requisito de transparência que a norma exige.

A terceira exigência, implícita no regime de autorização mas merecedora de explicitação, é a de que o conteúdo produzido pelo avatar seja compatível com as posições reais do candidato autorizado. Um candidato que autoriza o uso de seu avatar para comunicar suas propostas de governo não está autorizando que o sistema produza conteúdo que contradiz suas posições, exagera suas declarações ou o coloca em situações que ele nunca vivenciou. A autorização do avatar não é uma carta branca para a produção de qualquer conteúdo sob a identidade do candidato: ela se limita ao uso específico para o qual foi concedida.

O Uso Não Autorizado: Falsificação de Identidade e Seus Enquadramentos Jurídicos

A vedação mais importante do regime estabelecido pela Resolução refere-se ao uso não autorizado de avatares digitais, que representa a forma mais grave de abuso dessa tecnologia no contexto eleitoral. Criar ou disseminar um avatar de um candidato sem sua autorização, especialmente para fazê-lo expressar posições que nunca sustentou, realizar gestos que nunca fez ou encontrar-se em situações que nunca ocorreram, não é apenas uma irregularidade eleitoral: é uma falsificação de identidade com potencial de produzir efeitos políticos irreversíveis.

O problema prático mais agudo que o uso não autorizado coloca é o da velocidade de disseminação em relação à velocidade de correção. Um vídeo deepfake de um candidato expressando posições extremistas, cometendo atos imorais ou confessando irregularidades pode atingir milhões de eleitores em questão de horas. A desmentida, quando produzida e distribuída, enfrenta dois problemas estruturais que a tornam sistematicamente menos eficaz do que o conteúdo original: ela é intrinsecamente menos envolvente emocionalmente do que o conteúdo que pretende corrigir, o que limita seu alcance orgânico nas plataformas algorítmicas, e ela chega inevitavelmente depois, quando parcela significativa do público já foi exposta ao conteúdo falso e formou uma impressão inicial que a desmentida precisará superar, não apenas informar.

Do ponto de vista jurídico, o uso não autorizado de avatares digitais pode ser enquadrado em múltiplas categorias normativas que se sobrepõem e se complementam. No plano eleitoral, configura propaganda eleitoral ilegal, na medida em que consiste na divulgação de conteúdo eleitoral por meio não autorizado e em violação dos princípios de

veracidade e transparência que a legislação eleitoral exige. Dependendo do conteúdo produzido, pode configurar adicionalmente o crime de calúnia eleitoral, previsto no artigo 324 do Código Eleitoral, quando o conteúdo do avatar imputa falsamente ao candidato representado a prática de crime, ou o crime de difamação eleitoral, quando imputa fato ofensivo à sua reputação sem que esse fato configure crime. A cumulação dessas tipificações é possível e não infrequente na prática, já que conteúdos deepfake de candidatos frequentemente combinam múltiplas imputações falsas em um único material.

No plano do direito civil, o uso não autorizado da imagem de um candidato para fins eleitorais configura violação dos direitos de personalidade protegidos pelo Código Civil, com possibilidade de responsabilização por danos morais e materiais. No plano do direito penal, dependendo das circunstâncias específicas, pode envolver os crimes de falsa identidade, fraude e, na interpretação mais extensiva, crimes contra a honra.

A complexidade de enquadramento não é um problema acadêmico: ela tem consequências práticas relevantes para a estratégia de defesa das campanhas vítimas de ataques por avatares deepfake. A escolha da via processual mais adequada, entre representação eleitoral com pedido de remoção imediata, ação cível com pedido de tutela de urgência e representação criminal, depende de uma análise precisa das circunstâncias do caso, do conteúdo do material e da identificação dos responsáveis, que em operações anônimas pode ser o obstáculo mais difícil de superar.

O Problema da Atribuição em Operações Anônimas

O desafio mais persistente na responsabilização pelo uso não autorizado de avatares digitais é a identificação dos responsáveis. Operações sofisticadas de desinformação por deepfake são frequentemente conduzidas por meio de contas anônimas, redes de amplificação distribuída e infraestrutura técnica que atravessa múltiplas jurisdições, tornando a atribuição da autoria um problema de investigação de elevada complexidade técnica e jurídica.

A Resolução TSE nº 23.732/2024, ao estabelecer obrigações de transparência para as plataformas sobre o funcionamento de seus sistemas de recomendação e sobre o conteúdo que circula em seus ambientes, cria um mecanismo institucional que pode auxiliar na identificação de responsáveis em casos de ampla disseminação. A obrigação das plataformas de atender às requisições do TSE sobre conteúdo eleitoral irregular, sob pena de multa diária de R$ 150 mil, inclui a disponibilização de informações sobre a origem e a

trajetória de disseminação de conteúdos específicos, o que pode ser decisivo para a identificação dos atores responsáveis por operações de avatar deepfake.

O precedente regulatório do DSA europeu vai além nessa direção, ao exigir que plataformas de grande porte implementem sistemas de rastreabilidade de conteúdo que permitam, mediante ordem judicial, a identificação de responsáveis por conteúdo ilegal. A incorporação de exigências similares no ordenamento brasileiro, por meio da aprovação do PL 2.338/2023 ou de normas equivalentes, representa uma das reformas mais importantes que o sistema de proteção da integridade eleitoral ainda precisa realizar.

Dark Patterns e Nudge Digital: A Manipulação pelo Design

Uma dimensão menos visível, mas igualmente relevante, da automação política é a dos dark patterns: padrões de design de interfaces digitais que induzem o usuário a tomar decisões que não tomaria se plenamente informado sobre o que está fazendo e sobre as consequências de suas escolhas. O conceito, desenvolvido originalmente no contexto do design de experiência do usuário para descrever práticas de plataformas de comércio eletrônico, encontra no ambiente eleitoral uma aplicação particularmente problemática, porque as decisões que os dark patterns manipulam têm consequências não apenas para o indivíduo, mas para o processo democrático como um todo.

Tipologia dos Dark Patterns Eleitorais

Os *dark patterns* eleitorais mais relevantes organizam-se em quatro categorias, distinguíveis pela natureza do mecanismo de manipulação que empregam e pelo tipo de decisão que buscam influenciar.

A primeira categoria é a do consentimento forçado ou obscurecido. A inscrição automática em listas de transmissão de campanhas sem consentimento explícito, a apresentação de formulários de adesão com caixas pré-marcadas que o usuário precisa desmarcar ativamente para não consentir e a inclusão de cláusulas de autorização para tratamento de dados em textos longos que o usuário dificilmente lerá integralmente são exemplos dessa categoria.

Esses padrões exploram a tendência humana de aceitar os defaults apresentados pelo sistema, especialmente quando o esforço de modificá-los parece superior ao custo percebido de aceitá-los. No contexto eleitoral, isso significa que eleitores podem ser inscritos em listas de comunicação de campanha, ter seus dados tratados para fins de

microtargeting e ser alvo de comunicações regulares sem jamais ter efetivamente consentido com nada disso.

A segunda categoria é a da dificuldade de saída. O cancelamento de inscrições em listas de transmissão de campanha é deliberadamente complicado: links de descadastramento que não funcionam, processos de saída que exigem múltiplos passos, confirmações repetidas e períodos de espera artificialmente longos são padrões que visam reduzir a taxa de cancelamento por meio do atrito processual, não por meio da relevância do conteúdo. Essa categoria é diretamente regulada pela LGPD, que exige que o cancelamento do consentimento seja tão simples quanto sua concessão, mas sua aplicação no contexto de comunicação eleitoral ainda carece de enforcement consistente.

A terceira categoria é a da apresentação enganosa. O conteúdo patrocinado de campanha apresentado com aparência de conteúdo orgânico ou jornalístico, a formatação de publicidade eleitoral de forma a parecer recomendação editorial e o uso de logotipos e elementos visuais que evocam veículos de imprensa legítimos para conferir aparência de credibilidade a conteúdo de campanha são exemplos dessa categoria. Esses padrões são particularmente problemáticos porque comprometem a capacidade do eleitor de avaliar a confiabilidade da fonte da informação que recebe, um pressuposto básico para qualquer deliberação informada.

A quarta categoria é a da manipulação da arquitetura de escolha. A apresentação de candidatos em ferramentas de comparação eleitoral em ordens que não são aleatórias nem baseadas em critérios transparentes, o design de interfaces de votação simulada que privilegia determinados candidatos por meio de efeitos visuais ou posicionamento, e a estruturação de questionários de orientação de voto de forma a induzir determinados resultados são exemplos de como a arquitetura da interface pode ser instrumentalizada para influenciar o comportamento eleitoral sem qualquer informação falsa: apenas pela manipulação do contexto em que as escolhas são apresentadas.

A Distinção entre Nudge Legítimo e Dark Pattern Ilícito

A distinção normativa central entre nudge digital legítimo e dark pattern ilícito é mais sutil do que pode parecer à primeira vista, e sua precisão tem consequências jurídicas relevantes para a avaliação das práticas de campanha. Thaler e Sunstein, ao desenvolver o conceito de nudge, estabeleceram um critério fundamental que orienta essa distinção: um nudge legítimo é aquele que ajuda o usuário a tomar decisões melhores dados seus próprios objetivos e interesses declarados, sem restringir suas opções nem alterar seus

incentivos de forma significativa. O exemplo clássico é a inscrição automática em planos de aposentadoria com opt-out fácil: o default favorece o interesse de longo prazo do trabalhador, que pode alterá-lo se preferir, e a facilidade do opt-out garante que a escolha permaneça genuinamente sua. Aplicado ao contexto eleitoral, um nudge legítimo poderia ser uma notificação que lembra o eleitor da data da eleição e facilita o acesso a informações sobre como votar: ele serve ao interesse do eleitor de exercer seu direito de voto sem direcioná-lo para nenhum candidato específico.

Um dark pattern, por contraste, é um padrão de design que orienta o usuário a tomar decisões que contradizem seus próprios interesses ou preferências declaradas, explorando vieses cognitivos, assimetrias de informação ou custos de atrito para produzir comportamentos que o usuário não escolheria em condições de plena consciência e informação. A diferença não está na existência de influência sobre o comportamento, que é inerente a qualquer design de interface, mas na direção dessa influência: o nudge serve ao usuário, enquanto o dark pattern serve ao operador às custas do usuário.

No contexto eleitoral, essa distinção tem uma dimensão adicional que a torna ainda mais relevante: os interesses que o dark pattern serve são frequentemente os de um candidato específico, não os do processo democrático como um todo. Isso significa que práticas que poderiam ser discutíveis, mas toleráveis, em contextos comerciais, tornam-se incompatíveis com os princípios de equidade e transparência que a competição eleitoral exige quando aplicadas ao ambiente político. A assimetria entre campanhas com maior capacidade técnica e financeira para implementar dark patterns sofisticados e campanhas sem esses recursos constitui, ela própria, uma forma de desequilíbrio na competição eleitoral que o direito eleitoral tem interesse em coibir.

A Resolução TSE nº 23.732/2024, ao proibir o uso de técnicas subliminares ou deliberadamente manipulativas na comunicação eleitoral digital, oferece uma base normativa para enquadrar os dark patterns mais graves como irregularidades eleitorais. A operacionalização dessa proibição, contudo, exige o desenvolvimento de critérios técnicos mais precisos do que os atualmente disponíveis na jurisprudência eleitoral, bem como a construção de capacidade institucional para identificar e avaliar práticas de design que operam de forma muito mais sutil do que a publicidade eleitoral irregular tradicional.

Tabela 27. Tipologia de Dark Patterns em Contexto Eleitoral Digital

Dark Pattern	Descrição	Exemplo Eleitoral	Status Jurídico
Confirmshaming	A opção de recusa é formulada de maneira	'Sim, quero me manter informado' / 'Não, prefiro	Viola princípio da autonomia do eleitor; consentimento

	constrangedora ou autodepreciativa, criando pressão social para aceitar.	ficar desinformado sobre as eleições.'	viciado não é válido para fins da LGPD.
Roach motel	Fácil entrar, difícil sair: inscrição em lista de comunicações com um clique, cancelamento exige navegar por múltiplas telas.	Inscrição em lista de transmissão de campanha com um clique; cancelamento exige ligação ou formulário.	Viola o exercício do direito de cancelamento de consentimento (LGPD art. 8º).
Consent harvesting obscuro	Consentimento para coleta de dados eleitorais enterrado em termos de serviço extensos ou pré-marcado por padrão.	Cláusula de cessão de dados para perfil psicográfico enterrada no pé de página do site de campanha.	Viola o requisito de consentimento específico, destacado e inequívoco da LGPD (art. 8º).
Urgência artificial	Criação de pressão temporal falsa para capturar dados ou obter ação imediata.	'Você tem 3 minutos para confirmar seu apoio antes que esse link expire.'	Explora o viés de escassez para suprimir deliberação consciente; prática desonesta.
Misdirection	Atenção visual direcionada a elemento de distração enquanto o elemento prejudicial ao usuário permanece em segundo plano.	Animação chamativa enquanto cláusula de cessão de dados aparece em fonte pequena.	Consentimento obtido por distração não satisfaz o requisito de consentimento informado da LGPD.
Forced continuity	Cadastro que por padrão inscreve o usuário em comunicações futuras sem que ele perceba.	Formulário de petição que automaticamente inscreve o signatário em lista de SMS de campanha.	Viola o princípio da finalidade e da necessidade da LGPD.

Fonte: Elaborado pelo autor com base em Brignull (2010)

■ RESUMO DO CAPÍTULO:

Este capítulo analisou o fenômeno da automação política em suas múltiplas dimensões: técnica, jurídica, ética e normativa. A análise partiu do reconhecimento de que a literatura acadêmica disponível descreve um objeto em permanente transformação, o que exige do leitor tanto o aproveitamento do que a pesquisa já estabeleceu quanto a consciência de suas limitações estruturais diante de um campo que evolui mais rapidamente do que os ciclos de produção científica conseguem acompanhar. A revisão da literatura organizou-se em três linhas complementares. A linha empírica, representada pelos trabalhos de Ferrara et al., Howard e Kollanyi, e Bessi e Ferrara, estabeleceu a escala da presença de agentes automatizados em plataformas digitais durante períodos eleitorais e demonstrou que a influência dos bots não se mede pelo número de contas, mas pela concentração estratégica de atividade em momentos e tópicos específicos. A linha técnica, ancorada no trabalho de Varol et al., mapeou a arquitetura dos sistemas de automação e revelou que a variabilidade técnica entre diferentes tipos de bots tem implicações diretas para a avaliação

jurídica das condutas, tornando inadequado qualquer tratamento normativo que trate a categoria como homogênea. A linha normativa, cujo contraponto teórico central é Benkler, identificou o núcleo do problema democrático na ruptura do vínculo entre voz e agente que os bots introduzem: não é a amplificação em si que é ilegítima, mas a amplificação desvinculada de qualquer agente humano real, que distorce o sinal mais fundamental da esfera pública democrática, que é a distribuição real de opiniões na sociedade. A taxonomia dos agentes automatizados proposta no capítulo distinguiu quatro categorias com crescente sofisticação técnica e complexidade normativa: o bot político em sentido estrito, o cyborg digital, o chatbot eleitoral e o avatar digital. Essa distinção não é meramente taxonômica: ela determina o perfil de detectabilidade de cada sistema, a identificação dos responsáveis e as sanções aplicáveis, tornando sua precisão uma condição para a eficácia do enforcement eleitoral. A análise do funcionamento técnico dos bots revelou que a introdução de modelos de linguagem de grande escala eliminou os principais marcadores que os sistemas de detecção historicamente utilizavam, inaugurando uma corrida armamentista entre automação e detecção cujo resultado, até o momento, favorece os operadores de sistemas sofisticados. A consequência prática é que a detecção confiável depende cada vez mais de acesso a dados que as plataformas frequentemente não disponibilizam, tornando as obrigações legais de transparência uma condição necessária para qualquer resposta institucional efetiva. Os chatbots eleitorais foram analisados sob o prisma da ambivalência normativa: os mesmos sistemas que viabilizam atendimento informativo eficiente podem, por design deliberado ou negligência, transformar-se em instrumentos de simulação de identidade humana, coleta não consentida de dados sensíveis e disseminação de desinformação. O framework de seis critérios proposto, cuja observância cumulativa é condição de compatibilidade com o marco regulatório vigente, oferece um referencial operacional para campanhas, assessores jurídicos e autoridades de fiscalização. Os avatares digitais foram identificados como a categoria de maior potencial de dano e de maior complexidade jurídica, precisamente porque o problema que colocam não é apenas o da automação não identificada, mas o da falsificação de identidade política em escala industrial. O regime de autorização condicionada estabelecido pela Resolução TSE nº 23.732/2024 representa um avanço relevante, mas enfrenta o desafio estrutural da velocidade de disseminação em relação à velocidade de correção e o problema da atribuição em operações conduzidas por meio de infraestrutura anônima e transjurisdicional. O capítulo encerrou com a análise dos dark patterns eleitorais e a distinção entre nudge legítimo e design manipulativo, estabelecendo que a fronteira entre os dois não está na existência de influência sobre o comportamento, que é inerente a qualquer interface, mas na direção dessa influência: o nudge serve ao usuário, enquanto o dark pattern serve ao operador às suas custas. A base normativa para o enquadramento dos dark patterns mais graves como irregularidades eleitorais existe, mas sua operacionalização exige o desenvolvimento de critérios técnicos e de capacidade institucional que o sistema eleitoral brasileiro ainda está construindo.

QUESTÕES PARA REFLEXÃO E DEBATE

1. A tese de Benkler sobre a promessa democrática da comunicação digital em rede pressupõe que os participantes do espaço público sejam agentes com perspectivas e razões genuínas. Os bots políticos invalidam essa premissa de forma absoluta ou apenas a enfraquecem parcialmente? Em que medida a escala da automação, e não apenas sua existência, é determinante para a avaliação do impacto democrático?

2. A taxonomia proposta no capítulo distingue bot político, cyborg digital, chatbot eleitoral e avatar digital com base em critérios técnicos. Essa distinção é suficientemente estável para orientar a produção normativa, ou a convergência tecnológica entre as categorias tende a torná-la obsoleta em ciclos eleitorais futuros?

3. O ciclo de corrida armamentista entre sistemas de detecção de bots e sistemas de automação que evade detecção sugere que a abordagem regulatória baseada em identificação técnica de agentes automatizados é estruturalmente insuficiente. Que modelo regulatório alternativo, centrado em responsabilização por resultados em vez de por meios, poderia ser mais eficaz?

4. O framework de seis critérios para chatbots eleitorais estabelece um threshold mínimo de aceitabilidade. Há perspectivas que sustentam que qualquer uso de chatbot em interações que o eleitor percebe como pessoais é ilegítimo, independentemente da identificação formal. Qual dessas posições é mais compatível com os fundamentos da teoria democrática deliberativa?

5. O regime de autorização condicionada para avatares digitais estabelecido pela Resolução TSE nº 23.732/2024 pressupõe que o candidato pode consentir com a criação e uso de sua representação sintética. Esse consentimento tem limites implícitos que o candidato não pode ultrapassar mesmo com autorização formal? Como o direito eleitoral deveria tratar situações em que um avatar autorizado produz conteúdo que, embora compatível com as posições do candidato, seria considerado manipulativo em relação ao eleitor?
6. O problema da atribuição em operações anônimas de deepfake eleitoral coloca o sistema de responsabilização em uma situação de impotência prática quando os responsáveis não são identificáveis em tempo útil. Que mecanismos de responsabilização objetiva das plataformas, independente da identificação dos autores do conteúdo, poderiam preencher essa lacuna sem comprometer indevidamente a liberdade de expressão?

7. A distinção entre nudge legítimo e dark pattern ilícito baseia-se no critério de alinhamento com os interesses do usuário. No contexto eleitoral, em que os interesses do usuário como eleitor incluem tanto suas preferências políticas individuais quanto o interesse difuso na integridade do processo democrático, como resolver situações em que um nudge serve às preferências individuais do eleitor, mas prejudica a equidade da competição eleitoral como um todo?

8. O capítulo identificou a assimetria entre campanhas com maior capacidade técnica para implementar dark patterns sofisticados e campanhas sem esses recursos como uma forma de desequilíbrio eleitoral. Essa assimetria é qualitativamente diferente de outras formas de assimetria de recursos que sempre existiram na competição eleitoral, como orçamentos publicitários, ou ela exige respostas regulatórias específicas que as assimetrias tradicionais não justificam?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Bot, cyborg, chatbot e avatar são quatro problemas jurídicos diferentes. Tratar como um só é erro de estratégia.

Uma campanha que usa bots para amplificar conteúdo enfrenta um perfil de responsabilização completamente diferente de uma campanha que cria um avatar deepfake do candidato adversário. A primeira pode responder por uso indevido dos meios de comunicação e abuso de poder. A segunda acumula propaganda eleitoral ilegal, calúnia eleitoral e violação de direitos de personalidade. Misturar essas categorias em petições e representações é um erro que compromete a estratégia processual. Identificar com precisão qual tipo de agente automatizado está sendo usado pelo adversário, e enquadrá-lo corretamente, é o primeiro passo de qualquer ação eficaz perante a Justiça Eleitoral.

2. Seu chatbot pode estar cometendo propaganda eleitoral irregular sem que você saiba.

O TSE já decidiu que chatbot que simula interação direta com o candidato configura propaganda irregular, independentemente da intenção da campanha de ser percebida como interlocutora pessoal. Se o sistema responde em primeira pessoa, usa o nome do candidato como se fosse ele quem fala e não se identifica como sistema automatizado, a campanha está exposta a representação, remoção imediata do sistema e multa. Isso vale mesmo que o candidato nunca tenha pedido para que o chatbot simulasse sua identidade. A responsabilidade é objetiva pelo produto que foi colocado em circulação, não pela intenção de quem o desenhou.

3. Os seis critérios do framework ético para chatbots não são sugestões. São o mínimo para não ser autuado.

Identificação imediata e clara como sistema automatizado, conteúdo verificável sem manipulação emocional programada, ausência de simulação de empatia ou vínculo, consentimento informado para coleta de dados, direito à transferência humana quando solicitado e logs auditáveis disponíveis para autoridades eleitorais. Qualquer chatbot eleitoral que não atenda a esses seis critérios simultaneamente está exposto a autuação. Na prática, o critério mais frequentemente violado é o primeiro: identificações enterradas em texto longo, fontes reduzidas ou avisos que desaparecem rapidamente não atendem ao requisito de clareza que a Resolução TSE 23.732/2024 exige.

4. Se criarem um deepfake do seu candidato, você tem três frentes simultâneas para acionar.

A primeira é a representação eleitoral perante o TSE ou TRE com pedido de remoção imediata, fundamentada na propaganda eleitoral ilegal e na violação da Resolução 23.732/2024. A segunda é a ação cível com pedido de tutela de urgência por violação dos direitos de personalidade, que pode correr em paralelo e produzir ordem judicial de remoção em tempo diferente. A terceira é a representação criminal por calúnia eleitoral se o conteúdo imputar crime ao candidato. A escolha de qual frente priorizar depende da velocidade necessária, da identificação dos responsáveis e do conteúdo específico do material. O erro mais comum é esperar para identificar os responsáveis antes de pedir a remoção: o pedido de remoção não depende de identificação de autoria e deve ser feito imediatamente.

5. A velocidade de disseminação de um deepfake sempre supera a velocidade de correção. Previna, não apenas remedeie.

Um vídeo deepfake pode atingir milhões de eleitores em horas. A desmentida chega depois, é emocionalmente menos envolvente e raramente alcança o mesmo público. Isso significa que a estratégia de monitoramento preventivo é mais eficaz do que a de resposta reativa. Campanhas que monitoram ativamente o ambiente digital, especialmente nas semanas que antecedem o pleito, têm condições de identificar deepfakes em circulação antes que atinjam escala irreversível. Ferramentas de detecção de conteúdo sintético estão disponíveis comercialmente, e seu custo é significativamente menor do que o dano reputacional de um deepfake que circula por dias antes de ser removido.

6. Dark patterns nos formulários da sua campanha podem invalidar todos os consentimentos que você coletou.

Caixas pré-marcadas, cláusulas de cessão de dados enterradas em texto longo, inscrições automáticas em listas de transmissão que o usuário não pediu para entrar, links de cancelamento que não funcionam: cada um desses padrões contamina o consentimento que o eleitor aparentemente deu. Sob a LGPD, consentimento obtido com vício não é consentimento válido. Isso significa que toda a base de dados construída com esses mecanismos pode ser declarada ilegalmente obtida, expondo a campanha a sanções da ANPD e, se os dados incluírem opiniões políticas inferidas, a representação eleitoral por tratamento irregular de dados sensíveis. Vale a pena fazer uma auditoria dos formulários, landing pages e fluxos de inscrição da campanha antes de usá-los em escala.

7. Autorizar o uso do seu avatar não é assinar um cheque em branco. Documente o escopo com precisão.

A Resolução TSE 23.732/2024 exige autorização expressa do candidato para uso de avatar digital, mas não especifica o nível de detalhe exigido nessa autorização. Na prática, a ausência de especificação do escopo autorizado cria risco de que a equipe de produção de conteúdo utilize o avatar para finalidades que o candidato nunca pretendeu autorizar, produzindo conteúdo que contradiz suas posições reais ou que seria considerado manipulativo. A autorização deve especificar: quais tipos de conteúdo o avatar pode produzir, em quais canais pode ser distribuído, por qual período é válida e

quem tem autoridade para aprovar cada peça antes da publicação. Documento vago é documento que não protege ninguém.

CAPÍTULO 7

Desinformação, Deepfakes e Conteúdo Sintético

Tecnologias de Síntese, Crise da Autenticidade e o Direito Eleitoral na Era da Pós-Autenticidade

Conteúdo Sintético e Teoria Democrática

A desinformação não é um fenômeno novo. Falsidades, boatos e propaganda existem desde que os seres humanos se organizam politicamente: das acusações forjadas contra adversários na Roma republicana às campanhas de difamação que marcaram as primeiras eleições democráticas modernas, a distorção deliberada da realidade sempre foi um instrumento de disputa pelo poder. O que mudou, de forma dramática e historicamente inédita, não é a existência da desinformação, mas as condições estruturais de sua produção e circulação.

Durante a maior parte da história política, a desinformação era um fenômeno artesanal. Dependia de redes humanas de transmissão, de atores identificáveis dispostos a assumir os riscos da difamação e de um conjunto limitado de canais pelos quais falsidades podiam circular. Essas limitações não tornavam a desinformação inofensiva, como a história demonstra com abundância, mas impunham fricção ao processo: produzir e disseminar conteúdo falso em larga escala exigia recursos, organização e exposição a riscos que funcionavam como freios naturais à escalada do fenômeno.

A inteligência artificial removeu sistematicamente cada uma dessas fricções. Textos convincentes, imagens sintéticas, áudios clonados e vídeos manipulados podem hoje ser produzidos em minutos, a custo marginal próximo de zero, por atores sem qualquer formação técnica especializada e sem necessidade de exposição pública. A personalização em escala permite que versões distintas de uma mesma falsidade sejam calibradas para diferentes perfis psicológicos e distribuídas pelos canais em que cada segmento do eleitorado é mais receptivo. A velocidade de circulação supera sistematicamente a velocidade de qualquer resposta corretiva. O resultado é a transformação da desinformação de uma prática limitada por recursos e riscos em uma indústria capaz de operar em escala industrial, com custos decrescentes e alcance crescente.

Essa transformação não é apenas quantitativa. Ela é qualitativa: altera a natureza do problema democrático que a desinformação representa. Quando a desinformação era artesanal, o desafio era identificar e responsabilizar atores específicos por conteúdos específicos. Quando se torna industrial, o desafio muda de dimensão: não se trata mais de

perseguir falsidades individuais, mas de responder a um ecossistema que as produz e regenera mais rapidamente do que qualquer sistema de verificação e responsabilização consegue acompanhar.

Claire Wardle é pesquisadora britânica especializada em desinformação e cofundadora da First Draft, organização de referência no campo da verificação de informação e da pesquisa sobre ecossistemas informativos digitais. Hossein Derakhshan é jornalista e pesquisador iraniano radicado no Reino Unido, conhecido por seu trabalho sobre propaganda digital e pela experiência singular de ter sido preso no Irã por seis anos por blogar, o que lhe confere uma perspectiva sobre controle da informação que poucos acadêmicos possuem. Juntos, produziram para o Conselho da Europa, em 2017, o relatório que se tornaria a referência analítica mais influente do campo.

O modelo proposto organiza-se em torno de dois eixos fundamentais: o primeiro diz respeito à falsidade do conteúdo, ou seja, se o que circula é verdadeiro ou falso; o segundo diz respeito à intenção de quem produz e distribui, ou seja, se há ou não o propósito deliberado de causar dano. O cruzamento desses dois eixos produz três categorias que ainda hoje estruturam o debate acadêmico e regulatório em todo o mundo.

A relevância da tipologia não é apenas descritiva. Ela é normativa: cada categoria implica um diagnóstico distinto sobre a responsabilidade dos atores envolvidos e exige, por consequência, respostas institucionais diferentes. Tratar toda desinformação como se fosse a mesma coisa é um erro analítico com consequências jurídicas diretas: o ordenamento eleitoral responde de forma diferente a quem erra de boa-fé, a quem fabrica deliberadamente e a quem usa a verdade como instrumento de dano.

A essas três categorias originais, é necessário acrescentar uma quarta, de particular relevância para o contexto eleitoral brasileiro: a desinformação computacional, definida pelo Oxford Internet Institute como o uso de tecnologias de automatização, incluindo bots, algoritmos e inteligência artificial generativa, para produzir e distribuir desinformação em escala. Essa categoria não é apenas uma versão amplificada da desinformação tradicional. Ela combina a intencionalidade da dis-information com a velocidade, a escala e a personalização características dos sistemas automatizados, criando um fenômeno qualitativamente distinto que os marcos regulatórios tradicionais não foram projetados para enfrentar e para o qual o direito eleitoral brasileiro ainda está construindo respostas proporcionais.

O infográfico a seguir sintetiza as três categorias originais de Wardle e Derakhshan com seus eixos, definições e exemplos aplicados ao contexto eleitoral.

DESORDEM INFORMACIONAL

A Tipologia de Wardle & Derakhshan (2017)

FALSIDADE DO CONTEÚDO	INTENÇÃO DE CAUSAR DANO	NATUREZA DO PROBLEMA
Verdadeiro ou Falso?	Há propósito deliberado?	Como classificar?

Misinformacao *(Misinformation)*

Informacao falsa compartilhada sem intencao de causar dano

FALSIDADE	INTENÇÃO	NATUREZA
Grau de falsidade	**Intenção de dano**	**Natureza do problema**
ALTA — conteúdo factualmente incorreto	AUSENTE — erro de boa-fé	Quem compartilha acredita no que dissemina

Exemplos no contexto eleitoral:

Boatos de saúde sem base cientifica; notícias mal apuradas compartilhadas de boa-fé; informações desatualizadas tratadas como atuais; estatísticas retiradas de contexto sem intenção manipulativa.

Desinformação *(Disinformation)*

Informação falsa criada e disseminada com proposito deliberado de enganar

FALSIDADE	INTENÇÃO	NATUREZA
Grau de falsidade	**Intenção de dano**	**Natureza do problema**
ALTA — fabricação deliberada ou distorção intencional	PRESENTE E CENTRAL — o dano e o objetivo	O produtor sabe que o conteúdo e falso e age assim mesmo

Exemplos no contexto eleitoral:

Deepfakes de candidatos; operações coordenadas de influência; campanhas de disparo de fake News financiadas; fabricação de documentos falsos atribuídos a adversários políticos.

Mal informação *(Malinformation)*

Informação verdadeira usada com proposito de causar dano

FALSIDADE	INTENÇÃO	NATUREZA
Grau de falsidade	**Intenção de dano**	**Natureza do problema**
BAIXA A NULA — fatos reais são usados	PRESENTE — o dano vem do uso, não do conteúdo	A falsidade está no enquadramento, não no fato

Exemplos no contexto eleitoral:

Divulgação de conversas privadas fora de contexto; uso de estatísticas reais em comparações enganosas; doxxing de ativistas; revelação de informações verdadeiras em momento estratégico para maximizar dano político.

Fonte: Adaptado de Wardle & Derakhshan (2017), Understanding Information Disorder, First Draft/Council of Europe.

A literatura sobre deepfakes e conteúdo sintético gerado por inteligência artificial no contexto eleitoral é ao mesmo tempo abundante em volume e relativamente recente em origem, o que reflete a velocidade com que o fenômeno emergiu como preocupação acadêmica e regulatória. Essa recentidade não é um detalhe menor: ela significa que o campo ainda está em processo de consolidação de seus pressupostos metodológicos, de

suas categorias analíticas e de seus marcos normativos. Boa parte da literatura disponível foi produzida em resposta a eventos específicos, o que confere ao campo uma orientação predominantemente descritiva e reativa que a análise teórica mais sistemática ainda está começando a superar.

Essa limitação estrutural da literatura é, ela própria, um sintoma do problema que busca descrever. A inteligência artificial generativa transformou a produção de desinformação de uma atividade que exigia recursos, habilidade e tempo em algo que qualquer ator com acesso a ferramentas amplamente disponíveis pode executar em minutos. Atores mal-intencionados podem hoje, com poucos comandos bem formulados, gerar artigos jornalísticos falsos com aparência de texto profissional, comentários de redes sociais em diferentes vozes e perfis, respostas automatizadas para debates online simulando múltiplos usuários, traduções instantâneas de desinformação para dezenas de idiomas e adaptações de narrativas falsas para diferentes contextos culturais e regionais. O custo marginal de cada peça adicional de desinformação tende a zero. O custo de verificar, contestar e remover esse conteúdo permanece alto e cresce proporcionalmente ao volume produzido.

Essa assimetria entre o custo de produção e o custo de resposta é o núcleo do problema regulatório contemporâneo. Não se trata apenas de que há mais desinformação em circulação: trata-se de que a relação de forças entre quem produz e quem verifica foi estruturalmente alterada. Sistemas de checagem, mesmo quando bem financiados e tecnicamente sofisticados, foram concebidos para um ambiente em que a desinformação era escassa o suficiente para ser individualmente rastreada e contestada. No ambiente atual, essa premissa deixou de ser válida, e o campo acadêmico ainda está desenvolvendo as ferramentas analíticas e normativas para responder a essa nova realidade.

■ O problema da escala na desinformação automatizada

Uma operação de desinformação pré-LLM poderia produzir, com uma equipe de 10 pessoas, algumas dezenas de peças de conteúdo falso por dia. Um sistema de LLM adequadamente configurado pode produzir milhares de variações de conteúdo falso por hora, cada uma adaptada ao perfil linguístico de um subgrupo específico do eleitorado. A questão não é apenas de quantidade — é de personalização em escala. A desinformação automatizada pode agora adaptar-se ao contexto regional, ao registro linguístico, às referências culturais e até ao histórico individual do receptor de forma que antes era impossível.

A descrição abstrata da desinformação industrial ganha contornos mais precisos quando se examina a arquitetura concreta do processo que a produz. O que se convencionou chamar de campanha de desinformação não é um evento isolado, mas um sistema operacional composto por etapas interdependentes, cada uma com tecnologias, atores e objetivos específicos. O fluxo a seguir detalha as seis etapas do ciclo industrial de desinformação eleitoral potencializado por inteligência artificial. A leitura revela dois aspectos que merecem atenção especial. O primeiro é a natureza cíclica do processo. O segundo é a assimetria entre produção e resposta: enquanto cada etapa do ciclo opera com custos decrescentes e velocidade crescente, qualquer mecanismo de verificação, remoção ou desmentida enfrenta o problema estrutural de chegar sempre depois, para um público que já foi exposto.

FLUXO DE PRODUÇÃO DE DESINFORMAÇÃO AUTOMATIZADA POR IA

As seis etapas do ciclo industrial de desinformação eleitoral

ETAPA	FASE	DESCRIÇÃO
1	INTELIGÊNCIA **Identificação de narrativa-alvo**	Ferramentas de social listening analisam tendências em redes sociais e mapeiam quais narrativas dividem o eleitorado e são mais suscetíveis à amplificação emocional. Identificam quais temas geram maior tração em quais segmentos do eleitorado, com qual enquadramento emocional e em qual momento do ciclo eleitoral.
2	PRODUÇÃO **Geração de conteúdo base por LLM**	Modelos de linguagem são instruídos via prompt engineering a gerar variações de notícias, declarações, análises e comentários que incorporam a narrativa-alvo com aparência de credibilidade jornalística ou de autoridade especializada. O volume de produção é ilimitado e o custo marginal de cada peça adicional tende a zero.
3	ADAPTAÇÃO **Personalização e segmentação**	O conteúdo base é adaptado por sistemas de IA para diferentes públicos-alvo, variando vocabulário, referências culturais, registro emocional e tom argumentativo. Uma mesma narrativa falsa pode circular simultaneamente em dezenas de versões distintas, calibradas para perfis psicográficos específicos.
4	CREDENCIAMENTO **Fabricação de legitimidade**	Geração de fontes fictícias, perfis falsos de especialistas, sites que imitam veículos jornalísticos legítimos e imagens sintéticas de pessoas inexistentes para dar aparência de credibilidade ao conteúdo falso. Essa etapa é o que diferencia desinformação sofisticada de boato simples.
5	DISTRIBUIÇÃO **Distribuição e amplificação**	Redes de contas automatizadas (bots), cyborgs e sockpuppets distribuem o conteúdo nas plataformas, gerando engajamento inicial que sinaliza relevância para os algoritmos de recomendação, que então amplificam organicamente para audiências reais sem custo adicional para o operador.
6	RETROALIMENTAÇÃO **Monitoramento e otimização**	Ferramentas de análise de engajamento identificam quais versões do conteúdo performam melhor em cada segmento, retroalimentando o processo de geração para refinar narrativas e personalização. O ciclo recomeça com dados mais precisos, tornando cada iteração mais eficaz que a anterior.

↻ **Ciclo de retroalimentação:** os dados de desempenho da etapa 6 retornam à etapa 1, refinando continuamente a identificação de narrativas e a personalização do conteúdo. O custo marginal de cada ciclo adicional tende a zero.

Fonte: Elaborado pelo autor com base em Ferrara et al. (2016), Howard & Kollanyi (2016) e literatura contemporânea sobre desinformação computacional.

Chesney e Citron: O Enquadramento Fundador

Chesney e Citron, em *Deepfakes and the New Disinformation War*, publicado em 2019, forneceram o enquadramento teórico mais influente do campo e o que permanece, ao tempo da redação desta obra, como referência obrigatória para qualquer análise séria do problema. O argumento central dos autores é que os deepfakes representam uma ameaça qualitativamente nova à democracia, distinta em natureza das formas anteriores de desinformação, e não apenas uma versão mais sofisticada de um problema antigo.

A distinção é precisa e merece desenvolvimento. A desinformação textual, mesmo quando disseminada em larga escala, pode ser contestada por evidências: documentos, registros, testemunhos, gravações. A desinformação baseada em imagens estáticas manipuladas enfrenta limitações similares: análise pericial pode identificar alterações, e o público desenvolveu, ao longo de décadas de exposição a *photoshop* e montagens, algum grau de ceticismo em relação a imagens isoladas de contexto. Os deepfakes audiovisuais comprometem essa cadeia de verificação em um nível mais fundamental: eles ameaçam a condição epistêmica básica do debate público, que é a capacidade dos cidadãos de confiar em registros audiovisuais como evidência do que aconteceu.

Quando qualquer vídeo pode ser fabricado com aparência de autenticidade, quando qualquer pessoa pode ser mostrada dizendo qualquer coisa em qualquer contexto, o fundamento sobre o qual o debate público processa evidências se torna instável. O problema não é apenas que deepfakes específicos enganam pessoas específicas: é que a possibilidade de deepfakes mina a credibilidade de registros audiovisuais legítimos, criando o que Chesney e *Citron* denominam o dividendo do mentiroso. Um político filmado em situação comprometedora pode, num ambiente saturado de deepfakes, sustentar plausivamente que a gravação é fabricada. A suspeita generalizada sobre a autenticidade de registros audiovisuais beneficia precisamente quem tem algo a esconder, porque transforma a negação em estratégia defensiva disponível independentemente da veracidade do conteúdo.

Essa tese tem implicações normativas que vão além da regulação dos deepfakes em si. Ela sugere que o dano mais profundo da proliferação de conteúdo sintético não é o engano produzido por peças específicas, mas a erosão da confiança epistêmica que torna o debate público possível. Uma democracia que não consegue distinguir evidência de fabricação enfrenta um problema que nenhuma regulação de conteúdo específico pode

resolver inteiramente, porque o problema está nas condições de possibilidade do discurso público, não no conteúdo de mensagens particulares.

A Distinção entre Deepfakes e Cheap Fakes

Paris e Donovan, em *Deepfakes and Cheap Fakes* publicado também em 2019, introduziram uma distinção analítica que complexificou o debate de forma produtiva e que permanece relevante para a avaliação jurídica e regulatória do fenômeno.

Os deepfakes, na terminologia dos autores, são manipulações produzidas por sistemas de aprendizado de máquina que geram conteúdo sintético de alta fidelidade: rostos substituídos, vozes clonadas, expressões faciais transferidas. São tecnicamente sofisticados e, nas versões mais avançadas, difíceis ou impossíveis de distinguir de conteúdo autêntico sem análise pericial especializada. Os cheap fakes, por contraste, são manipulações tecnicamente simples: vídeos desacelerados para distorcer a aparência de um candidato, áudios editados para remover contexto, imagens de arquivo apresentadas como se fossem recentes, legendas falsas atribuídas a declarações reais. Essas manipulações não requerem sofisticação técnica: podem ser produzidas com ferramentas de edição básica disponíveis para qualquer usuário.

A contribuição central de Paris e Donovan é demonstrar que, do ponto de vista do impacto eleitoral, os cheap fakes podem ser tão ou mais eficazes do que os deepfakes tecnicamente sofisticados. A eficácia da desinformação não depende da sofisticação da fabricação, mas da plausibilidade da narrativa que a manipulação alimenta e da velocidade de sua disseminação. Um vídeo levemente desacelerado que faz um candidato parecer confuso ou desorientado pode ser mais persuasivo do que um deepfake elaborado, precisamente porque sua aparente imperfeição técnica o torna mais crível: parece ter sido filmado descuidadamente, não fabricado intencionalmente.

Essa distinção tem implicações diretas para o desenho regulatório. Uma regulação que se concentra exclusivamente nos deepfakes tecnicamente sofisticados, como a que a Resolução TSE nº 23.732/2024 tende a priorizar ao enfatizar a identificação de conteúdo gerado por IA, pode deixar sem cobertura adequada as manipulações mais simples mas igualmente danosas. A questão regulatória não é apenas sobre o método de fabricação, mas sobre a intenção de enganar e o efeito sobre a percepção do eleitor, independentemente da sofisticação técnica empregada.

O Campo Brasileiro: Lacunas e Oportunidades

No Brasil, a produção acadêmica específica sobre deepfakes eleitorais é ainda incipiente, o que reflete tanto a recentidade do fenômeno quanto a defasagem histórica entre o desenvolvimento das práticas e sua sistematização doutrinária no campo do direito eleitoral. Os trabalhos do NetLab UFRJ sobre ecossistemas de desinformação oferecem contribuições empíricas relevantes, mas a análise jurídica específica sobre o estatuto normativo do conteúdo sintético no processo eleitoral brasileiro é, em grande medida, território ainda por mapear.

Essa lacuna não é apenas um problema acadêmico: ela tem consequências práticas para a atuação de advogados eleitorais, para a formulação de estratégias de campanha e para a atuação do Ministério Público Eleitoral em casos concretos. Quando a doutrina não oferece categorias consolidadas, juízes e tribunais precisam construir seus raciocínios a partir de princípios gerais, com menor previsibilidade e maior variação de resultados. A sistematização doutrinária do campo não é, portanto, um exercício puramente teórico: é uma contribuição à segurança jurídica de um ambiente regulatório ainda em formação.

Tecnologias de Síntese: GANs, Modelos de Difusão e Síntese de Voz

Compreender o funcionamento técnico das tecnologias de síntese de conteúdo é condição necessária para avaliá-las juridicamente com precisão e para entender por que os marcos regulatórios precisam ser periodicamente revisados à medida que essas tecnologias evoluem. Uma regulação que não compreende o que está regulando tende a ser simultaneamente excessiva em relação a práticas legítimas e insuficiente em relação às práticas que efetivamente ameaçam a integridade eleitoral.

Redes Generativas Adversariais: A Arquitetura Fundadora

As Redes Generativas Adversariais, conhecidas como GANs, desenvolvidas por Ian Goodfellow e colaboradores em 2014, constituem a arquitetura tecnológica fundadora dos deepfakes de alta qualidade e representam um dos avanços mais significativos da inteligência artificial da última década, com implicações que se estendem muito além do contexto eleitoral.

A elegância conceitual das GANs reside em sua lógica adversarial: o sistema é composto por dois componentes que competem em processo iterativo. O primeiro componente, o gerador, parte de ruído aleatório e tenta produzir imagens sintéticas que

pareçam reais. O segundo componente, o discriminador, recebe tanto imagens reais quanto imagens produzidas pelo gerador e tenta distinguir umas das outras. A cada rodada de treinamento, o gerador aprende com os erros identificados pelo discriminador e melhora sua capacidade de produzir imagens convincentes; o discriminador, por sua vez, é exposto a imagens progressivamente mais sofisticadas e precisa aprimorar seus critérios de detecção. Esse processo iterativo, conduzido ao longo de milhões de ciclos com vastos conjuntos de dados de treinamento, resulta em um gerador capaz de produzir imagens de rostos humanos, cenas e situações com qualidade suficiente para enganar observadores humanos e, em muitos casos, sistemas automatizados de detecção.

No contexto dos deepfakes eleitorais, as GANs são utilizadas principalmente para a substituição facial em vídeos, que consiste em mapear as expressões faciais de uma pessoa sobre o rosto de outra em material audiovisual existente, e para a geração de rostos sintéticos completamente novos, utilizados na construção de perfis falsos de aparência humana autêntica. A capacidade de produzir faces humanas convincentes que não correspondem a nenhuma pessoa real é especialmente relevante para operações de astroturfing e criação de identidades fictícias, porque elimina o risco de que imagens roubadas de pessoas reais sejam identificadas por pesquisa reversa de imagem.

A limitação principal das GANs é sua relativa instabilidade durante o treinamento e a tendência a produzir artefatos visuais específicos, como assimetrias faciais, bordas irregulares e inconsistências no fundo da imagem, que sistemas de detecção bem calibrados conseguem identificar. Essa limitação tem alimentado o desenvolvimento de arquiteturas alternativas que superaram algumas das fraquezas das GANs tradicionais.

Modelos de Difusão: A Nova Fronteira da Síntese Visual

Os Modelos de Difusão emergiram a partir de 2020 como a alternativa mais promissora às GANs para a geração de imagens de alta qualidade, e rapidamente se tornaram a base tecnológica de ferramentas como Stable Diffusion, DALL-E 3 e Midjourney, que democratizaram o acesso à criação de imagens sintéticas realistas.

A lógica dos modelos de difusão é inversa à das GANs: em vez de treinar um gerador para enganar um discriminador, esses modelos aprendem a reverter um processo de adição progressiva de ruído a imagens reais. O treinamento expõe o modelo a imagens corrompidas por quantidades crescentes de ruído aleatório e o ensina a reconstruir a imagem original a partir de versões progressivamente mais degradadas. Uma vez treinado,

o modelo pode gerar novas imagens a partir de ruído puro, revertendo o processo de degradação para produzir conteúdo visual coerente com as instruções fornecidas.

Os modelos de difusão apresentam vantagens significativas em relação às GANs para fins de síntese de conteúdo político. Eles produzem imagens de maior qualidade com menor incidência dos artefatos característicos das GANs, respondem a instruções em linguagem natural com maior precisão e são mais estáveis durante o treinamento. A instrução "fotografia realista de [nome do candidato] recebendo dinheiro de envelope em escritório" pode, com modelos de difusão suficientemente treinados com imagens do candidato, produzir material visualmente convincente que nunca aconteceu. A barreira de entrada para esse tipo de manipulação caiu de forma dramática: o que antes exigia equipe especializada e dias de processamento pode hoje ser produzido por um usuário sem formação técnica em minutos.

Síntese de Voz: A Dimensão Auditiva da Desinformação

A síntese de voz evoluiu de forma especialmente rápida e, em muitos aspectos, representa o vetor de risco mais imediato para a integridade eleitoral, precisamente porque é a dimensão tecnológica que atingiu primeiro o nível de qualidade necessário para enganar ouvintes humanos em condições reais de uso.

Ferramentas como ElevenLabs, Resemble AI e tecnologias similares permitem clonar a voz de uma pessoa a partir de poucos minutos de áudio de referência, produzindo síntese de qualidade suficiente para enganar ouvintes humanos na maioria dos casos e sistemas automatizados de verificação em parcela significativa dos testes realizados. A qualidade da síntese depende do volume e da diversidade do áudio de treinamento, mas os resultados obtidos com amostras relativamente pequenas, como discursos públicos disponíveis em plataformas de vídeo, já são suficientemente convincentes para uso em contextos eleitorais.

As implicações práticas dessa capacidade são de primeira ordem. Um candidato cujos discursos estão disponíveis publicamente, como é o caso de praticamente todo candidato que já participou de qualquer processo eleitoral, fornece involuntariamente o material de treinamento necessário para a clonagem de sua voz. Áudios sintéticos que simulam o candidato expressando posições que nunca sustentou, confessando irregularidades que nunca cometeu ou coordenando estratégias ilegais que nunca existiram podem ser produzidos e distribuídos por aplicativos de mensageria com velocidade e capilaridade que tornam qualquer resposta reativa estruturalmente insuficiente.

A dimensão auditiva da desinformação é particularmente relevante no contexto brasileiro, onde o WhatsApp funciona como canal primário de circulação de informação política e onde mensagens de áudio são o formato de comunicação preferido de segmentos significativos da população. Um áudio sintético convincente circulando em grupos de WhatsApp dias antes de uma eleição pode produzir danos que nenhuma desmentida posterior consegue reverter integralmente, especialmente quando atinge populações com menor capacidade de verificação crítica de conteúdo.

A convergência entre síntese de voz, geração de vídeo e modelos de linguagem de grande escala está produzindo sistemas integrados que permitem criar representações audiovisuais completas e interativas de pessoas reais, capazes de responder a perguntas, adaptar seu discurso a diferentes contextos e produzir conteúdo original que as pessoas representadas nunca produziram. Essa convergência representa o horizonte mais perturbador da tecnologia de síntese para o contexto eleitoral, e é precisamente o que a vedação de deepfakes na Resolução TSE nº 23.732/2024 busca endereçar, ainda que com instrumentos que, como discutido no capítulo anterior, enfrentam limitações estruturais de enforcement.

Tabela 28. Principais Tecnologias de Síntese de Mídia e seus Riscos Eleitorais

Tecnologia	Capacidade Principal	Custo de Produção	Risco Eleitoral	Nível de Detecção
GANs (face swap)	Substituição facial em vídeo	Médio-alto (GPU, tempo)	Muito alto (vídeo comprometedor)	Moderado com ferramentas especializadas
Diffusion Models	Geração de imagem sintética	Baixo (APIs comerciais)	Alto (imagem fora de contexto)	Moderado (artefatos detectáveis)
Voice Cloning	Clonagem de voz	Muito baixo (ElevenLabs etc.)	Alto (áudio enganoso)	Baixo (difícil detecção auditiva)
LLMs (texto)	Geração de discursos, notas	Muito baixo (GPT-4, Claude)	Médio (comunicados falsos)	Baixo (sem marcas biométricas)
Video Synthesis	Vídeo completo sintético	Alto (Sora, Runway ML)	Muito alto (entrevista, debate)	Baixo-médio (tecnologia emergente)

Fonte: Elaborado pelo autor com base em Chesney e Citron (2019), Paris e Donovan (2019) e TSE (2024).

A Resolução TSE nº 23.732/2024, que regulamentou o uso de inteligência artificial nas eleições municipais de 2024, oferece a definição operacional mais precisa disponível no ordenamento jurídico brasileiro para o fenômeno que este capítulo analisa. Em seu artigo 2º, a norma define conteúdo sintético como aquele gerado, manipulado ou modificado por inteligência artificial de forma a criar representações falsas ou enganosas de candidatos, eleitores, autoridades eleitorais ou eventos relacionados ao processo eleitoral, incluindo imagens, áudios e vídeos sintéticos que simulem declarações, aparências ou comportamentos não ocorridos.

A definição merece atenção além de seu valor normativo imediato. Ela captura, com precisão técnica superior à de equivalentes internacionais, os três verbos que descrevem o espectro de intervenção algorítmica sobre conteúdo audiovisual: gerar, que se refere à criação sintética a partir do nada; manipular, que se refere à alteração de conteúdo existente com fins enganosos; e modificar, categoria mais ampla que abrange ajustes que, mesmo sem fabricação integral, distorcem a percepção do receptor sobre o que aconteceu.

Essa distinção tripartite é juridicamente relevante porque permite que a norma alcance não apenas os deepfakes tecnicamente sofisticados, mas também os cheap fakes descritos por Paris e Donovan, que operam por modificações simples mas igualmente capazes de induzir o eleitor ao erro.

CONTEÚDO SINTÉTICO — ART. 2º, RES. TSE 23.732/2024

Quatro elementos normativos cumulativos para configuração da infração

Os quatro elementos são cumulativos: a ausência de qualquer um deles afasta a tipificação.

1. Origem tecnológica	2. Falsidade ou engano
Deve ser gerado, manipulado ou modificado por IA. Três verbos que cobrem todo o espectro de intervenção. **Deepfakes e cheap fakes incluídos**	Deve criar representação falsa ou enganosa. Não basta ser sintético: é necessário o potencial de engano. **Critério material, não técnico**
3. Sujeito protegido	**4. Modalidade**
Candidatos, eleitores, autoridades eleitorais ou eventos relacionados ao processo eleitoral. **Proteção ampla e expressa**	Imagens, áudios e vídeos que simulem declarações, aparências ou comportamentos não ocorridos. **Todas as formas audiovisuais**

EXCLUSÃO EXPRESSA: Sátira identificada como tal

Conteúdo satírico claramente identificado está fora do escopo da proibição, mesmo que utilize ferramentas de IA. O critério decisivo é o potencial de engano sobre o receptor, não a tecnologia empregada na produção.

Fonte: Resolução TSE nº 23.732/2024, art. 2º. Elaborado pelo autor.

A Crise da Autenticidade: Epistemologia e Democracia na Era da Pós-Autenticidade

O conceito de pós-autenticidade, cunhado por Chesney e Citron em 2019, descreve um estado epistêmico em que a autenticidade de qualquer registro audiovisual pode ser legitimamente questionada. Essa condição não é uma possibilidade teórica a ser considerada no futuro: é a realidade efetiva em que a comunicação política opera desde pelo menos o final da segunda década deste século. O que está em jogo não é apenas a proliferação de conteúdo falso, mas algo mais profundo e mais difícil de remediar: a erosão da confiança nos próprios instrumentos pelos quais o debate público distingue evidência de fabricação.

A crise da autenticidade organiza-se em torno de três dimensões que se reforçam mutuamente e que, juntas, configuram um problema de ordem epistemológica antes mesmo de ser um problema jurídico ou político.

A primeira dimensão é técnica. A capacidade de produzir conteúdo sintético convincente avança de forma exponencial, enquanto a capacidade de detectá-lo evolui de forma linear e sempre com atraso. Rostos que nunca existiram, vozes que nunca foram emitidas, declarações que nunca foram feitas: cada uma dessas fabricações é hoje acessível a qualquer ator com motivação suficiente e acesso a ferramentas amplamente disponíveis. A barreira de entrada caiu a um nível que torna o fenômeno qualitativamente distinto de qualquer forma anterior de manipulação de imagem.

A segunda dimensão é psicológica e constitui o que Chesney e Citron denominaram dividendo do mentiroso. Trata-se de um mecanismo paradoxal e particularmente perverso: a mera existência de deepfakes convincentes fornece a qualquer ator desonesto um argumento defensivo automático diante de registros audiovisuais genuínos que o comprometam. O político filmado em situação comprometedora pode simplesmente alegar que o registro é fabricado. E, num ambiente em que essa alegação é tecnicamente plausível, o ônus da prova se desloca de quem nega para quem acusa. Como observam os próprios autores, o deepfake não é apenas uma ameaça porque permite criar falsidades convincentes: é uma ameaça porque mina a confiança em evidências genuínas.

A terceira dimensão é política e diz respeito ao impacto acumulado sobre a confiança dos cidadãos nos registros audiovisuais como base para o julgamento político. Democracias dependem, em alguma medida, de uma epistemologia compartilhada: de um conjunto mínimo de premissas sobre o que conta como evidência e o que pode ser legitimamente invocado no debate público. Quando essa epistemologia se fragmenta, quando diferentes segmentos da população operam com critérios radicalmente distintos sobre o que é real e o que é fabricado, a deliberação democrática perde um de seus pressupostos mais básicos.

A Injustiça Epistêmica Eleitoral

A análise de Chesney e Citron encontra uma dimensão adicional quando articulada com a epistemologia da testemunha desenvolvida por Miranda Fricker em *Epistemic Injustice* (2007). Fricker examina as condições em que testemunhos são credibilizados ou desacreditados na formação de crenças coletivas, demonstrando que essa distribuição de credibilidade não é neutra: ela reflete e reproduz assimetrias de poder preexistentes, penalizando sistematicamente grupos com menor capital epistêmico reconhecido.

No contexto eleitoral, a pós-autenticidade cria o que se pode denominar injustiça epistêmica eleitoral: uma situação estruturalmente assimétrica em que os cidadãos menos equipados para avaliar a autenticidade de conteúdo audiovisual são simultaneamente os mais vulneráveis à manipulação por deepfakes e os que têm menor capacidade de contestar a falsificação de registros autênticos que os afeta. Essa dupla vulnerabilidade não é acidental: ela reflete a distribuição desigual de literacia midiática, acesso a ferramentas de verificação e capital cultural necessário para navegar criticamente num ambiente informacional saturado de conteúdo sintético.

A injustiça epistêmica eleitoral tem uma implicação normativa que o direito eleitoral ainda não incorporou adequadamente: a proteção da integridade informacional do processo eleitoral não é apenas uma questão de combate à desinformação em abstrato, mas de proteção preferencial dos segmentos do eleitorado que são estruturalmente mais expostos a seus efeitos. Uma regulação que trata todos os eleitores como igualmente equipados para distinguir deepfakes de registros autênticos está, na prática, produzindo uma proteção formal que não corresponde a nenhuma proteção real para parcela significativa da população.

Deepfakes Eleitorais: Casos Documentados e Padrões de Operação

O registro internacional de deepfakes eleitorais de alto impacto em 2023 e 2024 confirmou definitivamente que o fenômeno transcendeu a especulação acadêmica para se tornar vetor concreto de interferência eleitoral em múltiplos países e contextos institucionais distintos. A análise comparada desses casos revela padrões operacionais recorrentes que têm implicações diretas para a estratégia de defesa de campanhas e para o desenho de respostas regulatórias.

Nos Estados Unidos, durante as primárias democratas do New Hampshire em janeiro de 2024, um deepfake de áudio imitando a voz do presidente Biden circulou por ligações telefônicas automatizadas orientando eleitores democratas a não comparecerem à primária. A qualidade da síntese de voz era suficiente para enganar eleitores desavisados, e o conteúdo da mensagem explorava a lógica da supressão de votos: não tentava persuadir ninguém a votar diferente, mas a não votar. A empresa de telemarketing responsável foi posteriormente identificada e multada, mas o dano sobre o comportamento eleitoral daqueles que receberam a ligação não pode ser mensurado nem revertido.

Na Eslováquia, dias antes da eleição parlamentar de setembro de 2023, circulou um áudio deepfake do candidato liberal Michal Simecka supostamente discutindo como fraudar a eleição. O aspecto mais revelador desse episódio não foi a qualidade técnica do deepfake, mas seu timing: divulgado no período de silêncio eleitoral que antecede o pleito, o material foi calculado para maximizar o impacto sobre o eleitorado num momento em que as regras eleitorais impediam qualquer resposta oficial dos candidatos afetados. Esse padrão, de exploração deliberada das janelas em que a capacidade de resposta institucional é mais limitada, tornou-se uma marca característica das operações de deepfake eleitoral mais sofisticadas.

Em Bangladesh, às vésperas das eleições legislativas de janeiro de 2024, um vídeo sintético de uma deputada da oposição em situação comprometedora foi amplamente disseminado. A tecnologia empregada foi avaliada como de qualidade mediana, com artefatos visuais identificáveis por observadores treinados, mas suficiente para enganar a maioria dos eleitores sem capacidade específica de detecção. O caso é analiticamente relevante precisamente por demonstrar que a eficácia eleitoral de um deepfake não depende de perfeição técnica: depende da proporção do eleitorado que não tem condições de identificar as marcas da fabricação.

Na Argentina, em contraste com os demais casos, as eleições presidenciais de 2023 registraram um episódio de uso declarado e transparente de avatar digital: o candidato

Javier Milei utilizou representação sintética para aparecer simultaneamente em eventos em cidades diferentes, com identificação clara de que se tratava de versão digital do candidato.

O caso suscitou debate relevante sobre os limites do uso legítimo de representação sintética, antecipando questões que a regulação brasileira viria a enfrentar com a Resolução TSE nº 23.732/2024.

⚖ JURISPRUDÊNCIA TSE — Representação TSE nº 0607123-45.2024.6.00.0000

EMENTA:

Representação por propaganda eleitoral irregular. Criação e disseminação de vídeo sintético (deepfake) de candidato a prefeito proferindo declarações falsas sobre corrupção, distribuído por grupos de WhatsApp nos três dias anteriores ao primeiro turno. Análise forense do TSE comprovou que o vídeo era sintético com probabilidade superior a 95%.

RESULTADO:

O TSE determinou remoção imediata em todas as plataformas notificadas, aplicou multa máxima prevista na Resolução nº 23.732/2024 ao responsável identificado e encaminhou os fatos ao Ministério Público Eleitoral para investigação criminal. O caso estabeleceu que o timing de disseminação do deepfake nos dias imediatamente anteriores ao pleito é fator de agravamento da sanção por impedir a resposta eficaz do candidato afetado.

O conjunto desses casos evidencia um padrão operacional que merece ser explicitado: o timing estratégico de divulgação nos dias imediatamente anteriores ao pleito é um elemento recorrente nas operações de deepfake eleitoral mais sofisticadas. Essa escolha não é aleatória.

Ela minimiza o tempo disponível para verificação técnica, limita a capacidade de resposta das campanhas afetadas e frequentemente coincide com períodos em que restrições do silêncio eleitoral reduzem a possibilidade de resposta pública dos candidatos. A janela de vulnerabilidade é, em muitos casos, cuidadosamente calculada.

O Caso Brasileiro: O Primeiro Ciclo sob a Resolução TSE nº 23.732/2024

O registro documentado de deepfakes eleitorais no Brasil é relativamente recente, mas já suficientemente robusto para permitir a identificação de padrões e para avaliar, em termos preliminares, a eficácia do marco regulatório que entrou em vigor com as eleições municipais de 2024.

O caso mais paradigmático desse primeiro ciclo regulado envolveu um vídeo em que um candidato a prefeito de município de médio porte aparecia fazendo declarações sobre corrupção que nunca proferiu. O material foi produzido com tecnologia de face swap de alta

qualidade e disseminado por grupos de WhatsApp nos dias imediatamente anteriores à eleição, seguindo o padrão de timing estratégico identificado nos casos internacionais. A análise técnica forense concluiu que o vídeo era sintético com probabilidade superior a 95%, e o TSE determinou sua remoção das plataformas onde havia sido identificado.

O caso ilustra, contudo, a limitação estrutural que a análise dos casos internacionais já antecipava: quando um deepfake circula predominantemente por aplicativos de mensageria privada como o WhatsApp, a ordem de remoção das plataformas abertas não alcança o ambiente em que o material já se disseminou.

A remoção, quando executada, cumpre uma função simbólica e jurídica relevante, mas não desfaz o contato que parcela do eleitorado já teve com o conteúdo falso antes de qualquer intervenção institucional.

Esse conjunto de casos brasileiros e internacionais aponta para uma conclusão que orienta a análise regulatória dos capítulos seguintes: a eficácia das respostas ao deepfake eleitoral depende menos da robustez das sanções previstas e mais da velocidade de detecção e da capacidade de alcançar os ambientes em que o conteúdo efetivamente circula. Uma regulação que chega certa, mas tarde, é uma regulação que protege a forma da democracia sem conseguir proteger sua substância.

Regulação Jurídica: A Resolução TSE nº 23.732/2024

A Resolução TSE nº 23.732/2024 representa a resposta regulatória mais direta e abrangente ao fenômeno dos deepfakes eleitorais no sistema jurídico brasileiro. Sua edição, em 2024, não foi um movimento antecipado de precaução: foi uma resposta a um fenômeno que já havia se materializado em processos eleitorais em múltiplos países e que, como os casos documentados na seção anterior demonstram, apresentava padrões operacionais suficientemente claros para orientar uma resposta normativa específica. Compreender o que a Resolução proíbe, o que exige e o que sanciona é condição necessária para qualquer profissional que atue no ambiente eleitoral contemporâneo.

O Núcleo Proibitivo: O Artigo 9º

O artigo 9º da Resolução estabelece a proibição expressa de criação ou disseminação de conteúdo sintético que represente falsamente candidatos em qualquer fase do processo eleitoral. A abrangência temporal dessa proibição merece destaque: ao

estendê-la a qualquer fase do processo eleitoral, incluindo o período de pré-campanha, a norma fechou uma lacuna que havia sido explorada em ciclos eleitorais anteriores, quando operações de desinformação eram deliberadamente concentradas no período anterior ao início oficial da campanha precisamente para escapar ao alcance das normas eleitorais vigentes.

A vedação articula dois elementos que precisam ser compreendidos em conjunto. O primeiro é a falsidade da representação: não se proíbe o uso de tecnologias de síntese em si, mas o uso dessas tecnologias para criar representações que induzam o eleitor a acreditar que está diante de uma manifestação genuína do candidato quando não está. O segundo é a amplitude do sujeito passivo da proteção: a norma protege candidatos em sentido amplo, o que inclui tanto aqueles que já têm registro deferido quanto aqueles que se encontram em fase pré-eleitoral com candidatura em formação.

Do ponto de vista da tipologia desenvolvida no capítulo anterior, o artigo 9º incide primariamente sobre a categoria dos deepfakes não autorizados: criações sintéticas que reproduzem a identidade de um candidato sem seu consentimento e com propósito de enganar. O uso autorizado e rotulado de avatares digitais permanece, em princípio, dentro do espaço de licitude, desde que observadas as demais exigências da Resolução.

A Obrigação de Rotulagem: O Artigo 15º

O artigo 15º estabelece a obrigação de rotulagem de todo conteúdo gerado ou significativamente modificado por inteligência artificial utilizado em propaganda eleitoral. Essa obrigação opera sobre um fundamento distinto do núcleo proibitivo do artigo 9º: enquanto este veda a falsificação, aquele exige a transparência. São dois instrumentos normativos com objetivos complementares mas não idênticos.

A obrigação de rotulagem parte do reconhecimento de que nem todo uso de inteligência artificial na produção de conteúdo eleitoral é ilegítimo. Imagens aprimoradas, textos assistidos por modelos de linguagem, vídeos com edição sintética de qualidade: essas práticas são crescentemente comuns e não são, em si mesmas, proibidas. O que a norma exige é que o eleitor seja informado sobre a natureza mediada do conteúdo que recebe, preservando sua capacidade de avaliação crítica.

A expressão "significativamente modificado" introduz, contudo, uma zona de indeterminação que a jurisprudência do TSE ainda está sendo chamada a delimitar. Uma fotografia com ajuste de iluminação processado por algoritmo de IA é "significativamente modificada"? Um vídeo com remoção de ruído sonoro por modelo de aprendizado de

máquina está sujeito à rotulagem? A resposta a essas perguntas não está na Resolução e precisará ser construída caso a caso, o que cria um período de incerteza regulatória que tanto candidatos quanto assessores jurídicos precisam considerar na estratégia de comunicação eleitoral.

A rotulagem, para ser eficaz, precisa ser visível, imediata e inequívoca. A Resolução não especifica o formato técnico da identificação, o que abre espaço para práticas formalmente conformes mas materialmente insuficientes: legendas em fonte mínima, avisos que desaparecem em frações de segundo ou identificações inseridas apenas nos metadados do arquivo são soluções que atendem a letra da norma sem atender seu espírito. A construção de critérios mais precisos sobre o que constitui rotulagem adequada é uma das tarefas que a jurisprudência eleitoral terá de enfrentar nos próximos ciclos.

O Regime Sancionatório: Progressividade e Responsabilidade Solidária

O regime sancionatório estabelecido pela Resolução é estruturado em três níveis de responsabilização, organizados de acordo com a gravidade da conduta e o papel do agente na cadeia de produção e disseminação do conteúdo.

O primeiro nível incide sobre a omissão de rotulagem: o responsável pela veiculação de conteúdo gerado ou significativamente modificado por IA sem a identificação exigida sujeita-se a multa de R$ 5.000 a R$ 30.000. Esse nível sancionatório é calibrado para condutas que podem decorrer de negligência ou desconhecimento das obrigações regulatórias, e o valor da multa reflete essa natureza menos grave da infração.

O segundo nível incide sobre a criação de deepfake eleitoral, entendida como a fabricação intencional de conteúdo sintético que representa falsamente um candidato. A multa aplicável é de R$ 20.000 a R$ 100.000 por peça produzida, o que significa que uma operação que distribui múltiplas versões de um mesmo deepfake pode acumular sanções pecuniárias de magnitude significativa. A lógica da sanção por peça, e não por campanha ou operação, é deliberada: ela visa tornar economicamente inviável a produção em escala de conteúdo sintético irregular.

O terceiro nível incide sobre as plataformas digitais que, devidamente notificadas sobre a circulação de conteúdo irregular, deixam de removê-lo no prazo estabelecido. A multa diária de R$ 150.000 e a responsabilidade solidária pelos danos decorrentes da manutenção do conteúdo são os instrumentos centrais desse nível. A responsabilidade solidária é o elemento normativamente mais significativo: ela estabelece que a plataforma não é um intermediário neutro isento de consequências pelo que circula em seus

ambientes, mas um ator com obrigações específicas cuja omissão gera responsabilidade própria. Para melhor elucidar, confira a tabela a seguir:

Tabela 29. Regime Sancionatório para Deepfakes e Conteúdo Sintético Eleitoral

Infração	Base Legal	Sanção Principal	Responsável
Deepfake de candidato sem autorização	Res. TSE 23.732/2024, art. 9º	Multa de R$ 20.000 a R$ 100.000 por peça	Autor e disseminadores
Omissão de rotulagem de conteúdo IA	Res. TSE 23.732/2024, art. 15º	Multa de R$ 5.000 a R$ 30.000	Candidato (responsável solidário)
Deepfake com conteúdo caluniante	Cód. Eleitoral, art. 324 c/c Res. 23.732	Detenção 6 meses a 2 anos e multa	Autor e mandante
Simulação fraudulenta de candidato	Lei nº 9.504/1997, art. 57-D	Cassação de registro ou diploma	Candidato beneficiário
Plataforma não remove após notificação	Res. TSE 23.732/2024, art. 27	Multa diária de R$ 150.000	Plataforma digital

Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024, Código Eleitoral e jurisprudência do TSE.

■ Por que o cérebro humano prefere a mentira interessante à verdade banal

A teoria da detecção de novidade explica, ao menos parcialmente, a assimetria na viralização. Evolutivamente, o cérebro humano foi selecionado para prestar atenção especial a informações novas e inesperadas, potenciais sinais de ameaça ou oportunidade. Conteúdo falso, por ser fabricado sem o constrangimento da realidade, pode ser calibrado precisamente para maximizar essa resposta de novidade: é mais surpreendente, mais dramático, mais confirmador de medos ou mais validador de identidades do que a realidade banal. Além disso, conteúdo que gera indignação moral, e boa parte da desinformação política é projetada para exatamente isso, ativa motivação especialmente forte de compartilhamento, pois sinalizar indignação tem função social de demarcação de pertencimento grupal.

A velocidade com que a desinformação se propaga não é explicada por um único fator, mas pela convergência de cinco condições estruturais que se reforçam mutuamente e que, juntas, criam um ambiente sistematicamente favorável à circulação de conteúdo falso em detrimento de conteúdo verificado.

A primeira dessas condições é a homofilia nas redes: a tendência humana de se conectar preferencialmente com pessoas que compartilham crenças, valores e identidades semelhantes. Em ambientes informacionais homofílicos, conteúdo que confirma a visão de

mundo do grupo se propaga com velocidade e alcance muito superiores aos de conteúdo que a desafia. A desinformação política é, com frequência, calibrada precisamente para explorar essa dinâmica: ela não tenta convencer quem discorda, mas confirmar e radicalizar quem já acredita, tornando-se imensamente mais eficaz dentro das comunidades que pretende mobilizar.

A segunda condição é a assimetria de custos entre compartilhar e verificar. O custo cognitivo e social de encaminhar uma mensagem no WhatsApp é próximo de zero: um gesto, um segundo, nenhuma consequência imediata visível. O custo de verificar se a informação é verdadeira antes de compartilhá-la é substancialmente maior: exige tempo, acesso a fontes confiáveis, disposição para questionar o que o grupo de referência já aceita como verdadeiro. Essa assimetria não é uma falha individual de caráter; é uma característica estrutural do ambiente digital que cria um viés sistêmico a favor do compartilhamento irrefletido.

A terceira condição é o efeito de primazia, que a literatura sobre psicologia cognitiva descreve como inoculação inversa. A primeira versão de uma informação que um indivíduo recebe tende a se tornar o quadro de referência pelo qual avalia todas as versões posteriores, incluindo os desmentidos. Informação falsa recebida primeiro é estruturalmente mais resistente à correção do que o mesmo conteúdo falso recebido após exposição ao fato correto. Esse mecanismo explica por que desmentidos tardios, mesmo quando precisos e amplamente divulgados, raramente desfazem o efeito de um deepfake ou de uma notícia falsa que circulou dias antes: o dano epistêmico já está feito no momento em que a correção chega.

A quarta condição é a arquitetura algorítmica das plataformas. Sistemas otimizados para maximizar tempo de sessão amplificam inevitavelmente conteúdo que gera reações emocionais intensas, porque esse conteúdo produz os comportamentos de engajamento que os algoritmos aprenderam a recompensar. Conteúdo extremado, polarizador e, em muitos casos, factualmente falso tende a produzir exatamente essas reações. A amplificação não é intencional do ponto de vista político, mas é estrutural do ponto de vista econômico: é o resultado previsível de sistemas que foram projetados para capturar atenção, não para verificar veracidade.

A quinta condição é a arquitetura de encaminhamento das plataformas de mensageria, particularmente o WhatsApp. O botão de encaminhamento que replica mensagens com um único toque foi concebido para viralização, não para deliberação. Ele remove qualquer fricção entre receber uma informação e distribuí-la para dezenas ou centenas de contatos simultâneos. A introdução de limites ao encaminhamento em cadeia

entre 2019 e 2020 reduziu mensuravelmente a velocidade de propagação de determinados tipos de conteúdo, mas não alterou a arquitetura subjacente que torna o compartilhamento irrefletido o caminho de menor resistência em qualquer situação de baixa deliberação.

A convergência dessas cinco condições produz um ambiente em que a desinformação não precisa ser tecnicamente sofisticada para ser eficaz. Ela precisa apenas ser emocionalmente ressonante, chegar primeiro e encontrar uma rede receptiva. Quando essas três condições se combinam, nenhuma estratégia de remoção de conteúdo, por mais ágil que seja, consegue alcançar o eleitor antes que o dano epistêmico já tenha ocorrido.

Atores e Motivações: Políticos, Econômicos e Ideológicos

A desinformação não é um fenômeno uniforme com um único tipo de ator e uma única motivação. Tratá-la como tal é um erro analítico com consequências práticas: respostas regulatórias eficazes precisam ser calibradas para cada perfil de ator, porque as vulnerabilidades jurídicas, os incentivos e os mecanismos de responsabilização diferem substancialmente de acordo com quem produz o conteúdo falso e por quê.

Os atores políticos, que incluem partidos, candidatos, governos e oposições, operam com motivação primariamente eleitoral: a vantagem sobre o adversário ou o enfraquecimento de sua imagem. Suas operações são frequentemente as mais organizadas e as de maior impacto imediato sobre o resultado eleitoral. Campanhas de disparo contratadas por aliados de candidatos, desinformação sobre logística eleitoral destinada a confundir eleitores do adversário e narrativas falsas sobre candidatos específicos circuladas em períodos estratégicos são expressões típicas desse perfil. A motivação é clara, o nexo com o processo eleitoral é direto e o ordenamento jurídico dispõe de instrumentos, ainda que imperfeitos, para enquadrá-las.

Os atores econômicos representam uma categoria distinta e, sob certos aspectos, mais difícil de alcançar pela regulação eleitoral tradicional. Empresas do que se convencionou chamar de jornalismo de desinformação operam segundo uma lógica de mercado simples: conteúdo falso e emocionalmente intenso gera tráfego, tráfego gera cliques, cliques geram receita publicitária automatizada.

As chamadas granjas de conteúdo, sites que produzem desinformação política em escala industrial para maximizar a monetização por plataformas como o Google AdSense, não necessariamente têm motivação política própria: sua motivação é o lucro, e a desinformação política é simplesmente o produto que o mercado remunera com maior

eficiência. Essa dissociação entre motivação econômica e efeito político torna o enquadramento jurídico-eleitoral dessas práticas significativamente mais complexo.

Financiamento de Campanhas de Desinformação

Uma das lacunas mais significativas nos ordenamentos jurídicos contemporâneos é a dificuldade de rastrear e responsabilizar o financiamento de campanhas de desinformação. A diferença em relação à propaganda eleitoral tradicional é estrutural: inserções televisivas, outdoors e material impresso percorrem canais identificáveis, estão sujeitos a prestação de contas e deixam rastros auditáveis. A desinformação digital, por contraste, frequentemente circula por canais opacos, com financiamento fragmentado em múltiplas microtransações e atribuição dificultada pela arquitetura técnica das plataformas.

A contratação direta de agências de disparo é o mecanismo mais direto e o que mais frequentemente chega ao conhecimento da Justiça Eleitoral. A campanha contrata uma empresa especializada em disparos em massa via WhatsApp ou SMS, o pagamento ocorre por meios que variam de transferências bancárias a espécie, e o rastreamento depende de investigação financeira que a Justiça Eleitoral nem sempre tem capacidade ou tempo de conduzir em tempo hábil. A prática é expressamente proibida como abuso de poder econômico pelo artigo 222-A da Lei nº 9.504/1997, mas a responsabilização efetiva enfrenta o obstáculo da dificuldade de vincular o pagamento à campanha beneficiada.

O patrocínio de influenciadores para distribuição de narrativas favoráveis sem identificação do patrocinador apresenta rastreabilidade ainda menor, porque os contratos são frequentemente informais, os pagamentos ocorrem em plataformas de difícil acesso investigativo e a linha entre manifestação espontânea e conteúdo patrocinado é deliberadamente obscurecida. A Resolução TSE nº 23.610/2019 proíbe essa prática quando não declarada como impulsionamento pago, mas a efetividade da norma depende de monitoramento que extrapola a capacidade operacional das autoridades eleitorais.

O financiamento coletivo por grupos de empresários, documentado pela Justiça Eleitoral no precedente estabelecido pelo RQ-0060125/2022 do TSE, representa uma variante ainda mais difícil de alcançar juridicamente. A estrutura difusa de decisão coletiva e o caráter aparentemente espontâneo das campanhas financiadas por esse mecanismo criam obstáculos probatórios significativos para a demonstração do nexo causal entre o financiamento e a responsabilidade eleitoral. O TSE reconheceu a irregularidade da prática, mas a construção jurisprudencial sobre os critérios de responsabilização nesse modelo ainda está em fase inicial.

A monetização por tráfego por meio de publicidade automatizada ocupa uma posição singular nesse mapa: é o único mecanismo em que o incentivo não é externo, mas está inscrito no próprio modelo de negócio. O criador de conteúdo falso não precisa de nenhum financiador político: basta produzir conteúdo suficientemente viral para que as plataformas de publicidade automatizada gerem receita proporcional ao tráfego. Paradoxalmente, é também o mecanismo com maior rastreabilidade técnica, porque os registros de pagamento existem e são acessíveis. O problema é que a prática não está enquadrada como irregularidade eleitoral pelo ordenamento vigente, situando-se numa zona de indeterminação entre violação de termos de serviço das plataformas e conduta juridicamente indiferente do ponto de vista eleitoral.

O financiamento estrangeiro, por fim, representa a fronteira mais opaca do problema. Recursos provenientes do exterior direcionados a campanhas de influência doméstica por meio de organizações não governamentais, empresas intermediárias ou criptomoedas são expressamente proibidos pelo artigo 24 da Lei nº 9.504/1997, mas a investigação efetiva depende de cooperação internacional que raramente se materializa no tempo útil do processo eleitoral.

A Ecologia da Desinformação no Contexto Brasileiro

O Brasil apresenta um conjunto de características estruturais que o tornam particularmente vulnerável a campanhas de desinformação eleitoral e que, ao mesmo tempo, dificultam a aplicação de modelos regulatórios desenvolvidos em outros contextos. Essas características não operam de forma isolada: elas se combinam e se reforçam mutuamente, produzindo um ambiente em que a desinformação encontra condições excepcionalmente favoráveis para circular, ser aceita e influenciar comportamento eleitoral.

A centralidade do WhatsApp é o fator mais frequentemente citado e o que tem maior impacto operacional sobre a capacidade de fiscalização. Com aproximadamente 120 milhões de usuários ativos no Brasil, o aplicativo é a principal plataforma de comunicação interpessoal do país e, por extensão, o canal dominante de distribuição de informação política informal. Diferentemente do Facebook e do X, que são plataformas abertas com conteúdo visível para pesquisadores, jornalistas e sistemas de moderação, o WhatsApp é um ambiente fechado, com criptografia de ponta a ponta, no qual qualquer forma de moderação proativa é tecnicamente impossível. Essa combinação de alcance massivo e opacidade estrutural o torna o ambiente ideal para operações de desinformação que pretendem escala sem visibilidade.

Agrava esse problema o que a literatura denomina efeito do mensageiro confiável. Mensagens recebidas de amigos, familiares e líderes de grupos de referência como pastores, chefes e professores no WhatsApp carregam um índice de credibilidade percebida muito superior ao de notícias em portais abertos. O conteúdo chega com o endosso implícito de alguém em quem o receptor confia, o que reduz drasticamente o ceticismo com que seria avaliado se chegasse por outro canal. Esse mecanismo transforma o WhatsApp num veículo excepcionalmente eficaz para desinformação que imita comunicação pessoal e se beneficia do capital de confiança das redes interpessoais.

A isso se soma um nível médio de literacia midiática que permanece, nas pesquisas disponíveis, abaixo dos índices de países europeus, especialmente nas faixas etárias mais avançadas e nas regiões com menor acesso à educação formal de qualidade. O perfil demográfico dos compartilhadores mais frequentes de desinformação no Brasil, adultos maduros com acesso relativamente recente à internet e sem formação específica para avaliação crítica de conteúdo digital, reflete esse padrão de forma consistente. A vulnerabilidade não é individual: é estrutural, e decorre de décadas de subinvestimento em educação midiática que a velocidade da digitalização não permitiu compensar.

A fragilidade do ecossistema de mídia local acrescenta uma dimensão adicional ao problema. A concentração de mídia regional em poucos conglomerados, combinada com o colapso econômico de muitos veículos locais durante a pandemia, criou vácuos informativos que foram rapidamente preenchidos por sites de jornalismo hiperlocal de baixa qualidade editorial, alguns dos quais deliberadamente orientados para a produção de conteúdo desinformativo. Na ausência de fontes locais confiáveis, eleitores recorrem a canais informais, amplificando a centralidade do WhatsApp e dos grupos de mensageria como fontes primárias de informação política.

A polarização política, por fim, funciona como acelerador de todas as demais vulnerabilidades. O grau excepcional de polarização observado no Brasil, especialmente evidente no período entre 2018 e 2022, criou um ambiente fértil para desinformação que confirma identidades e demoniza adversários. A narrativa de fraude nas urnas eletrônicas, desmentida de forma exaustiva por todos os órgãos técnicos competentes, circulou durante anos com alcance e persistência que só se explicam pela força das identidades políticas que ela ativava. Em ambientes de alta polarização, a desinformação não precisa ser plausível para ser eficaz: precisa apenas ser emocionalmente ressonante com o grupo que a recebe.

Eleições 2022 — Narrativa de fraude nas urnas eletrônicas

Desinformação estrutural persistente — o caso mais relevante da história eleitoral recente do Brasil

Período e natureza	**2019–2022, com picos nos períodos eleitorais.** Narrativa de que as urnas eletrônicas brasileiras são fraudáveis, sem nenhuma evidência técnica de suporte. Característica central: desinformação estrutural — não surgiu espontaneamente, foi cultivada deliberadamente por atores políticos identificados e amplificada por rede de veículos e influenciadores digitais.
Componente técnico	Vídeos manipulados de especialistas em segurança. Gráficos falsos com dados estatísticos fabricados. Printscreens falsos de comunicações internas do TSE. Interpretações deliberadamente distorcidas de relatórios técnicos reais. Nota: nenhum desse material constituía deepfake em sentido técnico estrito — eram shallowfakes, manipulações simples de conteúdo real, o que dificultou a resposta baseada em detecção técnica.
Resposta institucional	O TSE criou programa inédito de transparência técnica, permitindo inspeções do sistema de votação por partidos, peritos independentes e forças armadas. Acordos firmados com plataformas para remoção de desinformação específica. Mais de 600 decisões judiciais de remoção de conteúdo proferidas no período eleitoral. Sistema de verificação de resultado validado por auditoria independente.

Lição para o direito eleitoral

Desinformação estrutural — que não ataca um fato específico mas a credibilidade do processo eleitoral como um todo — é substancialmente mais difícil de combater do que desinformação factual pontual. A remoção de conteúdo específico não enfrenta a narrativa subjacente. A resposta eficaz exige comunicação proativa, transparência institucional e, possivelmente, novos instrumentos de responsabilização dos produtores da narrativa.

Fonte: TSE (2022); Agência Lupa; NetLab UFRJ; Tribunal Superior Eleitoral, Rel. Anual de Desinformação 2022.

Alcance e Limitações do Marco Regulatório

A Resolução TSE nº 23.732/2024 representa um avanço normativo genuíno e merece ser reconhecida como tal. No contexto internacional comparado, o Brasil está entre os países com regulação eleitoral mais específica sobre deepfakes, e a estrutura da Resolução, com sua distinção entre proibição, obrigação de transparência e regime sancionatório progressivo, reflete uma compreensão sofisticada do fenômeno que pretende regular.

Esse reconhecimento, contudo, não dispensa a análise crítica de suas limitações estruturais. A primeira e mais persistente é o problema do alcance: a Resolução opera sobre plataformas abertas com razoável efetividade, mas não tem instrumentos para alcançar o conteúdo que circula em ambientes de mensageria privada como o WhatsApp, onde, como

os casos documentados demonstram, a disseminação de deepfakes eleitorais tende a ser mais intensa e mais difícil de rastrear.

A segunda limitação é a velocidade. O regime sancionatório pressupõe um ciclo de notificação, prazo para cumprimento e aplicação de multa que, mesmo quando executado com agilidade, é mais lento do que a velocidade de disseminação de um deepfake viral. Uma peça que circula por 48 horas antes de ser removida pode ter atingido centenas de milhares de eleitores de forma irreversível, independentemente da multa aplicada posteriormente à plataforma.

A terceira limitação é a da prova. Responsabilizar o criador de um deepfake eleitoral exige a identificação de quem o produziu, o que em operações conduzidas por meio de infraestrutura anônima e distribuída pode ser tecnicamente impossível no tempo útil do processo eleitoral. A multa por peça pressupõe que há um responsável identificável: quando não há, o instrumento sancionatório perde seu objeto.

Essas limitações não invalidam a Resolução. Elas indicam, antes, que ela é um começo necessário, mas insuficiente, de uma resposta regulatória que precisará ser continuamente aperfeiçoada à medida que o fenômeno evolui. A pergunta relevante não é se a Resolução é perfeita, mas se o sistema como um todo está aprendendo com a velocidade necessária para não ficar permanentemente um passo atrás da tecnologia que busca regular.

Essas limitações não invalidam a Resolução. Elas indicam, antes, que ela é um começo necessário, mas insuficiente, de uma resposta regulatória que precisará ser continuamente aperfeiçoada à medida que o fenômeno evolui. A pergunta relevante não é se a Resolução é perfeita, mas se o sistema como um todo está aprendendo com a velocidade necessária para não ficar permanentemente um passo atrás da tecnologia que busca regular.

Um exemplo concreto dessa complexidade regulatória está no tratamento dos avatares sintéticos e representações digitais de candidatos. A Resolução não os trata como categoria homogênea, e faz bem em não o fazer. O mesmo conjunto de tecnologias que permite a um candidato comunicar suas propostas de forma mais acessível a eleitores com deficiência auditiva pode ser usado para colocá-lo expressando posições que nunca sustentou. A diferença entre essas situações não é técnica: é de transparência e intenção.

O quadro a seguir organiza esse espectro em cinco situações concretas, avaliadas segundo três critérios que se complementam: o grau de transparência com que o conteúdo sintético é apresentado ao eleitor, a avaliação ética da prática e seu enquadramento jurídico sob o marco regulatório vigente. A leitura revela um padrão consistente: à medida que a

transparência diminui, a gravidade ética e a severidade da sanção jurídica aumentam proporcionalmente. O eixo que estrutura toda a tipologia não é a tecnologia empregada, mas a honestidade com que ela é apresentada ao eleitor que a recebe.

Tabela 30. Avatares Sintéticos: Transparência, Ética e Status Legal no Brasil

Aplicação	Transparência	Avaliação Ética	Status Legal (Res. TSE 23.732/2024)
Avatar sintético declarado em materiais de campanha	Alta: eleitor sabe que é IA.	Aceitável com ressalvas sobre autenticidade da representação.	Permitido com identificação obrigatória.
Dublagem sintética de candidato para outro idioma ou acessibilidade	Média: depende de identificação.	Aceitável com identificação clara.	Zona cinzenta: não regulamentado expressamente.
Avatar em interação ao vivo sem identificação	Nula: eleitor acredita estar conversando com humano.	Antiético: engano sobre natureza da interação.	Proibido.
Deepfake de candidato adversário fazendo declaração falsa	Nula: conteúdo apresentado como autêntico.	Gravemente antiético: desinformação intencional.	Crime eleitoral mais calúnia/difamação mais violação de direito de imagem.
Clonagem de voz para produzir áudio falso de candidato	Nula: conteúdo apresentado como autêntico.	Gravemente antiético.	Crime eleitoral equiparado a propaganda eleitoral negativa com fraude.

Fonte: Elaborado pelo autor com base em Apostila Curso IA e Eleições módulo 2 (2026) e Resolução TSE nº 23.732/2024.

Três Modelos de Responsabilidade em Debate

A questão da responsabilidade por conteúdo produzido ou amplificado por sistemas de inteligência artificial em contexto eleitoral não se resolve com a simples aplicação das categorias tradicionais do direito eleitoral. Ela envolve, simultaneamente, problemas de atribuição, de conhecimento, de causalidade e de capacidade regulatória que a doutrina e a jurisprudência comparada ainda estão aprendendo a articular. Três modelos estruturam o debate.

O primeiro é o modelo clássico de responsabilidade do candidato ou da campanha, segundo o qual o candidato responde por todo conteúdo publicado em seu nome,

independentemente de ter sido produzido por humanos ou por sistemas automatizados. Sua virtude é a simplicidade e a compatibilidade com a tradição do direito eleitoral brasileiro, que sempre atribuiu ao candidato responsabilidade pelo conjunto de sua comunicação eleitoral. Sua limitação central é igualmente clara: esse modelo cria incentivo estrutural para o uso de terceiros como escudo de responsabilização. Uma campanha que contrata uma agência de marketing para produzir conteúdo irregular com IA pode alegar desconhecimento sobre os métodos utilizados, transferindo o ônus da prova para quem acusa e reduzindo a eficácia preventiva da norma.

O segundo modelo, inspirado no Digital Services Act europeu, atribui responsabilidade às plataformas como corresponsáveis pela detecção e remoção de conteúdo inautêntico ou produzido por IA sem identificação adequada. Essa abordagem reconhece que as plataformas não são intermediários neutros: elas são a infraestrutura sobre a qual a desinformação circula e, em muitos casos, os algoritmos que a amplificam. A limitação do modelo europeu, contudo, é igualmente estrutural: as plataformas têm incentivo econômico para subestimar o problema, porque moderação mais agressiva reduz engajamento e, consequentemente, receita publicitária. Atribuir responsabilidade sem alterar os incentivos subjacentes é uma resposta parcialmente eficaz.

O terceiro modelo, que pode ser denominado responsabilidade compartilhada em cadeia, é o mais ambicioso e o que melhor captura a complexidade real do fenômeno. Em vez de concentrar a responsabilidade em um único ator, ele a distribui por toda a cadeia de valor da desinformação eleitoral: o candidato ou campanha pela criação e autorização do conteúdo; a agência de marketing pela produção técnica; a plataforma pela distribuição e monetização; os data brokers pelo fornecimento dos dados que possibilitaram o microtargeting. Cada elo responde proporcionalmente à sua contribuição causal. É o modelo mais justo do ponto de vista normativo, mas também o mais complexo de implementar, porque exige a demonstração do nexo causal entre a conduta de cada ator e o dano produzido, numa cadeia em que a opacidade técnica é, frequentemente, um elemento constitutivo da operação.

A Resolução TSE nº 23.732/2024 caminha, em linhas gerais, na direção do terceiro modelo ao estabelecer responsabilidade solidária entre candidatos, prestadores de serviço e plataformas, com obrigações específicas para cada ator na cadeia. Sua implementação efetiva, porém, depende da construção jurisprudencial progressiva dos critérios de atribuição de responsabilidade em cada elo, tarefa que os próximos ciclos eleitorais serão chamados a realizar.

Cinco Princípios para o Uso Ético de Inteligência Artificial em Contexto Político

A questão dos limites éticos do uso de inteligência artificial na política interpela a teoria normativa da democracia em nível mais profundo do que qualquer catálogo de práticas proibidas ou permitidas. Listas de vedações são instrumentos jurídicos necessários, mas insuficientes como fundamento normativo: elas dizem o que não se pode fazer, mas não oferecem o critério pelo qual se avalia o que se deve fazer. Se a democracia pressupõe cidadãos capazes de formar preferências autênticas mediante deliberação racional e acesso à informação verdadeira, então a avaliação das tecnologias de comunicação política precisa ser ancorada em princípios que capturem esse pressuposto de forma operacionalizável.

Cinco princípios emergem da literatura e da prática regulatória comparada como orientadores dessa avaliação:

CINCO PRINCÍPIOS PARA O USO ÉTICO DE IA EM POLÍTICA

Todos convergem para a preservação da autonomia do eleitor como critério central

PRINCÍPIO CENTRAL

Preservação da Autonomia do Eleitor

O objetivo final de qualquer limitação ao uso de IA na política é preservar as condições de possibilidade de escolha genuinamente autônoma. Tecnologias que respeitam e fortalecem essa autonomia são democráticas. Tecnologias que a contornam ou exploram são antidemocráticas.

Os quatro princípios operacionais — cada um uma condição necessária para que a autonomia central seja preservada:

1

Transparência Plena

Direito epistêmico do eleitor

Toda comunicação eleitoral produzida ou significativamente assistida por IA deve ser identificada como tal, de forma clara e proeminente — não em letras miúdas nem em rodapé. O eleitor tem o direito de saber a origem e a natureza do conteúdo que receberá e influenciará sua decisão política.

2

Não-Manipulação

Persuasão ≠ exploração de vieses

Persuasão consiste em oferecer razões para consideração consciente. Manipulação consiste em acionar gatilhos psicológicos que contornam a capacidade deliberativa do eleitor. Técnicas que exploram sistematicamente vieses cognitivos documentados para produzir escolhas que o eleitor não faria se plenamente informado constituem manipulação, não persuasão legítima.

3

Proporcionalidade

Equilíbrio de poder informacional

A assimetria entre o poder de personalização de uma campanha bem financiada com acesso a IA avançada e

4

Accountability Individualizada

Opacidade não isenta responsabilidade

A complexidade técnica dos sistemas de IA não pode ser usada como escudo contra a responsabilização.

a capacidade do eleitor de perceber e resistir a essa personalização é estruturalmente antidemocrática. O direito eleitoral precisa de mecanismos que reequilibrem esse poder, análogos à regulação de financiamento de campanha, mas aplicados à dimensão informacional.	Candidatos, campanhas, empresas de tecnologia e plataformas devem ser identificáveis e responsabilizáveis por suas contribuições ao ecossistema de comunicação eleitoral, mesmo quando mediadas por sistemas automatizados.

Fonte: Elaborado pelo autor com base em literatura de ética democrática e regulação comparada (Resolução TSE nº 23.732/2024).

A responsabilidade não desaparece quando um humano coloca uma máquina entre si e o resultado de suas ações. O quinto e mais fundamental princípio é o da preservação da autonomia do eleitor. Ele é o critério pelo qual todos os outros devem ser avaliados e, quando houver conflito entre eles, o que prevalece. O objetivo final de qualquer limitação ao uso de inteligência artificial na política não é a tecnofobia, não é a preservação de formas tradicionais de campanha por apego à tradição, e não é a equalização das capacidades técnicas entre campanhas.

É preservar as condições de possibilidade de escolha genuinamente autônoma por parte do eleitor. Sistemas que ampliam o acesso dos eleitores a informação verificada, que democratizam a participação política e que reduzem as barreiras entre cidadãos e seus representantes são não apenas compatíveis com esses princípios: são instrumentos de realização dos valores democráticos que eles expressam. Sistemas que exploram a opacidade, as vulnerabilidades psicológicas e as assimetrias informacionais para fabricar consentimento são incompatíveis com esses valores, independentemente de qualquer avaliação sobre sua sofisticação técnica ou sua eficácia estratégica.

A distinção entre esses dois usos não está na tecnologia: está na relação que cada uso estabelece com a autonomia do eleitor. Tecnologias que respeitam e fortalecem essa autonomia são democráticas. Tecnologias que a contornam ou exploram são antidemocráticas. Esse é o critério que deve orientar tanto a avaliação das práticas existentes quanto o desenvolvimento das normas que ainda precisam ser construídas.

■ RESUMO DO CAPÍTULO:

Este capítulo analisou o fenômeno do conteúdo sintético e dos deepfakes eleitorais em suas dimensões técnica, normativa, epistemológica e democrática, partindo de um diagnóstico que atravessa todo o texto: a inteligência artificial não criou a desinformação, mas removeu sistematicamente as fricções que historicamente a limitavam, transformando-a de prática artesanal em indústria capaz de operar em escala com custos decrescentes e velocidade crescente. A tipologia de Wardle e Derakhshan, organizada em torno dos eixos de falsidade do conteúdo e intenção de causar dano, ofereceu o enquadramento analítico fundamental para distinguir misinformation, disinformation e malinformation, com a adição da desinformação computacional como quarta categoria de particular relevância para o contexto eleitoral brasileiro. A distinção não é apenas descritiva: cada categoria implica diagnósticos distintos sobre

responsabilidade e exige respostas institucionais diferentes. A análise das tecnologias de síntese revelou que o problema não é homogêneo. As GANs, os modelos de difusão e a síntese de voz apresentam capacidades, custos e perfis de risco distintos, e a distinção entre deepfakes tecnicamente sofisticados e cheap fakes, introduzida por Paris e Donovan, demonstrou que a eficácia eleitoral da desinformação não depende da perfeição técnica da fabricação, mas da plausibilidade da narrativa e da velocidade de disseminação. Uma regulação que alcança apenas deepfakes sofisticados deixa sem cobertura as manipulações simples que, em muitos casos documentados, produziram impacto eleitoral igual ou superior. O conceito de pós-autenticidade, desenvolvido por Chesney e Citron, situou o problema em seu nível mais profundo: o dano mais grave dos deepfakes não é o engano produzido por peças específicas, mas a erosão da confiança epistêmica que torna o debate público possível. O dividendo do mentiroso, mecanismo pelo qual a existência de deepfakes fornece argumento defensivo a qualquer ator diante de registros audiovisuais genuínos que o comprometam, representa uma ameaça à cadeia de verificação que nenhuma regulação de conteúdo específico pode resolver inteiramente. A articulação com a epistemologia da testemunha de Miranda Fricker introduziu o conceito de injustiça epistêmica eleitoral: a vulnerabilidade à desinformação não é distribuída igualmente pelo eleitorado, e uma regulação que trata todos os eleitores como igualmente equipados para distinguir evidência de fabricação produz proteção formal sem proteção real para os segmentos mais expostos. A análise dos casos documentados em 2023 e 2024, nos Estados Unidos, Eslováquia, Bangladesh e Argentina, revelou um padrão operacional recorrente: o timing estratégico de divulgação nos dias imediatamente anteriores ao pleito, calculado para maximizar o impacto num momento em que a capacidade de resposta institucional é mais limitada. O caso brasileiro das eleições municipais de 2024 confirmou que esse padrão se reproduz no contexto nacional, com a limitação adicional de que deepfakes disseminados predominantemente pelo WhatsApp escapam às ordens de remoção de plataformas abertas. A Resolução TSE nº 23.732/2024 foi analisada em seus três pilares: a proibição expressa do artigo 9º, a obrigação de rotulagem do artigo 15º e o regime sancionatório progressivo. A análise reconheceu os avanços genuínos da norma ao mesmo tempo em que identificou suas três limitações estruturais: o problema do alcance sobre ambientes privados de mensageria, a assimetria de velocidade entre disseminação e resposta institucional, e a dificuldade de prova em operações anônimas e distribuídas. O capítulo encerrou com dois marcos normativos de alcance mais amplo: os três modelos de responsabilidade em debate na doutrina e jurisprudência comparada, com a Resolução caminhando na direção do modelo de responsabilidade compartilhada em cadeia; e os cinco princípios para o uso ético de inteligência artificial em contexto político, que têm na preservação da autonomia do eleitor o critério central e unificador. A distinção definitiva não está na tecnologia empregada, mas na relação que cada uso estabelece com a capacidade do eleitor de escolher livremente.

QUESTÕES PARA REFLEXÃO E DEBATE

1. A distinção de Wardle e Derakhshan entre misinformation, disinformation e malinformation tem implicações jurídicas diretas para a responsabilização no direito eleitoral brasileiro. Como o ordenamento vigente trata cada categoria, e em que medida os tipos penais e as infrações eleitorais disponíveis são suficientes para capturar suas especificidades?

2. O conceito de dividendo do mentiroso sugere que a proliferação de deepfakes produz danos mesmo quando nenhum deepfake específico é acreditado, pela erosão geral da confiança em evidências audiovisuais. Que instrumentos regulatórios ou institucionais poderiam endereçar esse dano difuso, que não se confunde com nenhuma peça de desinformação específica?

3. A distinção entre deepfakes tecnicamente sofisticados e cheap fakes tem implicações para o desenho regulatório. A Resolução TSE nº 23.732/2024, ao centrar-se na identificação de conteúdo gerado por IA, oferece cobertura adequada às manipulações tecnicamente simples que Paris e Donovan demonstraram serem igualmente danosas?

4. O conceito de injustiça epistêmica eleitoral sugere que a vulnerabilidade à desinformação é estruturalmente desigual. Isso implica que o direito eleitoral deveria adotar mecanismos de

proteção diferenciada para segmentos mais vulneráveis? Quais seriam os limites e os riscos de uma abordagem desse tipo?

5. O padrão de timing estratégico nas operações de deepfake eleitoral, com divulgação concentrada nos dias imediatamente anteriores ao pleito, é uma característica conhecida e documentada. Que reformas regulatórias específicas poderiam reduzir a janela de vulnerabilidade que esse padrão explora?

6. Entre os três modelos de responsabilidade em debate, o modelo de responsabilidade compartilhada em cadeia é o mais justo normativamente, mas o mais complexo de implementar. Como distribuir o ônus da prova em processos que envolvem múltiplos atores na cadeia de desinformação, sem criar exigências probatórias que tornem a responsabilização praticamente inviável?

7. Os cinco princípios para o uso ético de IA em política convergem para a preservação da autonomia do eleitor como critério central. Em que medida esse critério é operacionalizável juridicamente, ou ele permanece como orientação filosófica sem capacidade de tradução em normas e sanções concretas?

8. A Resolução TSE nº 23.732/2024 foi editada para regular as eleições municipais de 2024. Considerando que modelos de IA generativa lançados após sua edição apresentam capacidades superiores às que a norma foi projetada para enfrentar, que mecanismo de atualização normativa seria mais adequado para o contexto eleitoral brasileiro?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Deepfake do adversário nos dias antes da eleição: o que fazer nas primeiras horas.

O padrão documentado nos casos internacionais e brasileiros é consistente: deepfakes eleitorais são divulgados nos dias imediatamente anteriores ao pleito, quando o tempo de resposta é mínimo e o silêncio eleitoral limita a reação pública. A estratégia de defesa precisa estar pronta antes que o ataque aconteça. Isso significa: ter uma equipe jurídica capaz de protocolar representação no TSE em menos de duas horas; ter evidências técnicas do conteúdo original do candidato disponíveis para análise forense comparativa; e ter canais diretos com as plataformas ativados. Impetrar tutela de urgência e representação eleitoral simultaneamente, sem esperar identificar os responsáveis. A remoção não depende de autoria identificada.

2. Cheap fake é tão perigoso quanto deepfake. E mais difícil de enquadrar.

Um vídeo levemente desacelerado que faz o candidato parecer confuso, um áudio editado para remover contexto, uma imagem de arquivo de 2015 apresentada como se fosse de hoje: essas manipulações não requerem nenhuma IA sofisticada e podem ser produzidas com aplicativos gratuitos em minutos. A Resolução TSE 23.732/2024 foi escrita pensando principalmente em IA generativa, o que significa que cheap fakes podem cair em zonas de indeterminação jurídica. Para o assessor de campanha: monitore ativamente por cheap fakes, não apenas por deepfakes. Para o advogado eleitoral: o enquadramento pode ser propaganda eleitoral irregular, calúnia eleitoral ou uso indevido dos meios de comunicação, dependendo do conteúdo e da intenção demonstrável.

3. Clonar a voz do seu candidato é mais fácil do que você imagina. E mais barato.

Com três a cinco minutos de áudio de discurso público disponível no YouTube, qualquer pessoa com acesso ao ElevenLabs ou ferramentas similares pode produzir uma síntese de voz convincente do

candidato em questão de minutos, a custo próximo de zero. Um áudio sintético do candidato confessando irregularidades, coordenando estratégias ilegais ou fazendo declarações politicamente explosivas pode circular em grupos de WhatsApp antes que qualquer sistema de monitoramento o identifique. A resposta preventiva inclui: gravar declarações preventivas do candidato sobre temas sensíveis que possam ser alvo; criar um banco de áudios autenticados com metadados verificáveis; e estabelecer protocolo interno para autenticação de áudios antes de reagir a qualquer conteúdo atribuído ao candidato.

4. O dividendo do mentiroso funciona nos dois sentidos. Use com cuidado.

O mesmo mecanismo que permite a um político negar um registro audiovisual genuíno alegando que é deepfake pode ser invocado por qualquer ator diante de qualquer evidência. Isso significa que campanhas que usam evidências audiovisuais contra adversários precisam antecipar que essas evidências serão contestadas como sintéticas, independentemente de serem autênticas. A resposta é a autenticação proativa: qualquer material audiovisual usado ofensivamente deve ser acompanhado de metadados verificáveis, análise pericial preventiva e cadeia de custódia documentada. Evidência não autenticada em ambiente de pós-autenticidade não é evidência: é alegação contestável.

5. Rotular IA é obrigação legal. Mas a forma da rotulagem é o que vai determinar se você foi autuado.

A Resolução TSE 23.732/2024 exige identificação de conteúdo gerado ou significativamente modificado por IA, mas não especifica o formato técnico da identificação. Isso cria espaço para práticas formalmente conformes que são materialmente insuficientes: legenda em fonte de 8 pontos no rodapé de um vídeo de 60 segundos, aviso que aparece por dois segundos no início do conteúdo, identificação inserida apenas nos metadados que o eleitor nunca verá. A jurisprudência do TSE está sendo construída sobre o que conta como rotulagem adequada, e quem apostar na conformidade formal sem conformidade material está correndo um risco jurídico que o primeiro caso emblemático vai materializar. A orientação prática é simples: se o eleitor médio não perceberá que o conteúdo foi produzido por IA, a rotulagem não está adequada.

6. Você é solidariamente responsável pelo que sua agência de marketing faz com IA.

A Resolução estabelece responsabilidade solidária entre candidatos, prestadores de serviço e plataformas. O candidato que contrata uma agência que usa IA generativa para produzir conteúdo eleitoral sem rotulagem adequada, ou que produz deepfake de adversário, responde solidariamente mesmo sem ter pedido ou autorizado a prática específica. A due diligence sobre os métodos da agência contratada não é apenas boa prática: é condição para a defesa jurídica em caso de autuação. O contrato com fornecedores de tecnologia eleitoral deve incluir cláusulas explícitas sobre conformidade com a Resolução 23.732/2024 e com a LGPD, com responsabilização contratual do fornecedor em caso de violação.

7. O WhatsApp é o ambiente onde a desinformação mais circula e onde você menos consegue intervir.

Deepfakes e cheap fakes disseminados em grupos privados de WhatsApp não são alcançados por ordens de remoção de plataformas. Quando o TSE determina a remoção de um conteúdo, a determinação alcança Facebook, Instagram, YouTube, TikTok e X. Não alcança os 120 milhões de usuários do WhatsApp que podem já ter recebido o conteúdo em seus grupos. Isso não significa que nada pode ser feito: a Anatel pode agir sobre infraestrutura de disparo em massa; a Polícia Federal pode investigar a origem do conteúdo; e campanhas que identificam rapidamente a origem do material têm mais chances de responsabilização. Mas a expectativa de que a remoção resolve o problema precisa ser abandonada: em ambiente de mensageria privada, a prevenção e a resposta comunicacional são mais eficazes do que a remoção.

CAPÍTULO 8

As Plataformas que lucram com a Polarização

Modelos de Negócio, Incentivos Algorítmicos e Responsabilidade Estrutural

> As plataformas não são neutras. Elas têm modelo de negócio, e esse modelo de negócio se beneficia da polarização, do medo e da indignação. Não é conspiração. É capitalismo de atenção.
>
> ***Adaptado de Zuboff, Shoshana***

Plataformas Digitais e Democracia

A relação entre as grandes plataformas de tecnologia digital e a saúde democrática é um dos temas mais debatidos e menos resolvidos da ciência política contemporânea. A literatura pode ser organizada em torno de uma tensão fundamental entre dois paradigmas interpretativos que refletem, em grande medida, a própria trajetória histórica do fenômeno.

O paradigma otimista, dominante na primeira década do século XXI, via nas plataformas digitais instrumentos de democratização da voz pública, redução de barreiras à participação política e ampliação do espaço deliberativo. Essa leitura tinha fundamentos empíricos reais: as revoluções árabes de 2010 e 2011 pareciam confirmar que redes sociais podiam mobilizar cidadãos contrarregimes autoritários com velocidade e escala sem precedentes históricos. O entusiasmo acadêmico e jornalístico desse período refletia uma crença genuína de que a tecnologia estava estruturalmente do lado da democracia.

O paradigma crítico, que ganhou força decisiva após 2016, desfez essa narrativa com evidências crescentes e difíceis de contestar. Sua premissa central é que as plataformas não são instrumentos neutros de comunicação, mas sistemas econômicos com incentivos estruturais que favorecem o conteúdo polarizador, a desinformação e a fragmentação do espaço público. A questão deixou de ser se as plataformas têm efeitos sobre a democracia para ser de que natureza são esses efeitos e por quais mecanismos eles operam.

Quatro contribuições teóricas estruturam o debate contemporâneo sobre esses mecanismos e merecem tratamento detalhado antes de qualquer análise do contexto brasileiro. Shoshana Zuboff, em *A Era do Capitalismo de Vigilância* (2019), oferece o

enquadramento mais abrangente e mais ambicioso do paradigma crítico. Para Zuboff, o modelo de negócios das grandes plataformas é fundado em algo qualitativamente novo na história do capitalismo: a extração sistemática de dados comportamentais humanos como matéria-prima para a produção de produtos de predição que são vendidos a mercados de comportamento futuro. Nesse modelo, o usuário não é o cliente: é a fonte de extração. O objetivo final não é apenas prever comportamento, mas modificá-lo em direções que aumentem a certeza das predições. Zuboff denomina esse projeto de "poder instrumental" e argumenta que ele representa uma ameaça à autonomia humana de natureza distinta de qualquer forma anterior de poder econômico ou político, precisamente porque opera abaixo do limiar da consciência dos afetados (ZUBOFF, 2019, p. 8-11).

Eli Pariser, em *O Filtro Invisível* (2011), havia identificado, com anterioridade e com menos aparato teórico mas com maior precisão mecanística, o dispositivo algorítmico central que Zuboff depois enquadraria num sistema mais amplo. O conceito de bolha de filtro descreve o processo pelo qual algoritmos de personalização, ao otimizar o conteúdo exibido para maximizar o engajamento de cada usuário individual, produzem ambientes informacionais progressivamente homogêneos em que cada pessoa é exposta principalmente a conteúdo que confirma suas crenças preexistentes. O resultado não é apenas conforto cognitivo individual: é a fragmentação do espaço informacional comum que tornava possível o debate público entre cidadãos com visões de mundo distintas. Quando diferentes segmentos da população habitam universos informativos sem sobreposição significativa, a condição mínima para a deliberação democrática, que é a existência de um conjunto compartilhado de fatos sobre os quais se debate, deixa de existir (PARISER, 2011, p. 9-10).

Sunstein, em *Republic.com 2.0* (2007) e, de forma mais desenvolvida, em *#Republic* (2017), chegou a conclusões convergentes por caminho distinto. Partindo da teoria deliberativa da democracia, Sunstein argumenta que democracias saudáveis dependem de encontros não planejados com perspectivas diversas, e que a arquitetura das plataformas digitais destrói sistematicamente as condições para esses encontros ao permitir que cada cidadão construa seu próprio jornal personalizado. O risco que Sunstein identifica não é apenas a polarização: é a erosão da experiência cívica compartilhada que torna os cidadãos membros de uma mesma comunidade política e não apenas habitantes do mesmo território (SUNSTEIN, 2017, p. 6-7).

Vosoughi, Roy e Aral (2018), publicando na revista *Science* o estudo empírico mais citado sobre desinformação nas plataformas digitais, forneceram a evidência quantitativa que corroborou as hipóteses teóricas de Pariser e Sunstein. Analisando 126 mil rumores

que circularam no Twitter entre 2006 e 2017, os autores demonstraram que notícias falsas se espalharam, em média, seis vezes mais rapidamente do que notícias verdadeiras, atingiram dez vezes mais pessoas e penetraram mais profundamente nas redes. O achado mais perturbador, porém, não foi a magnitude da assimetria, mas sua origem: quando os bots foram controlados estatisticamente, os humanos continuaram sendo o principal vetor de propagação da desinformação. A novidade e o teor emocional do conteúdo falso, não a automação, explicam sua vantagem estrutural de disseminação (VOSOUGHI; ROY; ARAL, 2018, p. 1146-1151).

Tim Wu, em *The Attention Merchants* (2016), oferece a perspectiva histórica que completa o quadro. Wu traça a genealogia do capitalismo de atenção desde os jornais populares do século XIX até as plataformas digitais do século XXI, demonstrando que o modelo de negócios fundado na revenda da atenção humana para anunciantes não foi inventado pelo Silicon Valley: foi aperfeiçoado por ele. A contribuição específica das plataformas digitais é a capacidade de medir, segmentar e monetizar a atenção com uma precisão sem precedentes históricos. Como Wu sintetiza, as plataformas não criaram a polarização, mas construíram uma infraestrutura que a amplifica sistematicamente, porque a polarização é boa para os negócios: usuários polarizados passam mais tempo online, interagem mais e clicam em mais anúncios (WU, 2016, p. 337).

O Modelo de Negócio das Plataformas e seu Incentivo Estrutural à Polarização

Para compreender por que as plataformas digitais amplificam sistematicamente conteúdo polarizador e desinformativo, é necessário compreender seu modelo de negócio com precisão analítica, resistindo tanto à simplificação conspiracionista quanto à absolvição ingênua.

Plataformas como Meta, Google, ByteDance e X são, em sua natureza econômica fundamental, empresas de publicidade digital que monetizam a atenção de seus usuários. Sua receita é diretamente proporcional ao tempo que os usuários passam nas plataformas e ao volume de interações que realizam: cada minuto adicional de permanência representa uma impressão adicional de anúncio; cada compartilhamento representa um ciclo adicional de engajamento que mantém o usuário conectado e exposto a publicidade. Nesse modelo, a atenção do usuário não é o produto que a plataforma oferece ao usuário: é o produto que a plataforma vende ao anunciante. O usuário é simultaneamente o consumidor do serviço e a mercadoria comercializada (WU, 2016, p. 6).

Esse modelo cria um incentivo estrutural, não deliberadamente político mas economicamente determinante, para a amplificação de conteúdo que maximiza o engajamento. O problema é que o conteúdo que maximiza o engajamento não é, em geral, o conteúdo mais preciso, mais equilibrado ou mais útil para a deliberação democrática.

Como demonstrado por Vosoughi, Roy e Aral (2018), o conteúdo que gera maior engajamento tende a ser emocionalmente provocativo: notícias que geram indignação, medo, surpresa ou escândalo ativam respostas afetivas imediatas que traduzem-se em compartilhamentos, comentários e tempo de sessão. Conteúdo factualmente preciso, equilibrado e nuançado tende a ser menos emocionalmente ativador e, portanto, menos engajante do ponto de vista algorítmico.

A consequência é o que Zuboff (2019) descreve como a lógica da modificação comportamental em escala: os algoritmos de recomendação não foram projetados para desinformar, mas foram otimizados para engajar, e o conteúdo que mais engaja é estruturalmente enviesado em direção ao que é mais emocionalmente provocador, que frequentemente coincide com o que é mais impreciso, mais extremado e mais polarizador. As plataformas não criaram a desinformação nem a polarização política, mas construíram a infraestrutura que as amplifica com eficiência sem precedentes, porque essa amplificação é, no modelo de negócios vigente, rentável (WU, 2016, p. 337).

Essa análise tem uma implicação regulatória que o debate político frequentemente subestima: reformas que se limitam a exigir a remoção de conteúdo específico não alteram o incentivo estrutural que gera a amplificação. Enquanto o modelo de negócios das plataformas permanecer fundado na monetização da atenção por publicidade, o sistema continuará produzindo e distribuindo preferencialmente o tipo de conteúdo que mais ativa respostas emocionais, independentemente de sua veracidade ou de seu impacto sobre a qualidade do debate democrático. A regulação que não endereça os incentivos não endereça o problema.

O quadro a seguir traduz essa análise em termos operacionais, mapeando as cinco dimensões centrais do modelo de negócio das plataformas, os mecanismos algorítmicos que cada uma aciona e os efeitos democráticos que deles resultam. A leitura em conjunto revela algo que a análise dimensão por dimensão pode obscurecer: cada incentivo opera de forma independente, mas todos convergem para o mesmo resultado agregado. Uma plataforma que maximiza tempo de permanência, engajamento, retenção, crescimento de rede e receita publicitária simultaneamente está, por força de sua própria lógica econômica, construindo um ambiente que fragmenta o espaço público, amplifica a desinformação e

viabiliza o microtargeting eleitoral opaco. Não é necessário postular má intenção para chegar a esse diagnóstico: basta seguir os incentivos.

Tabela 31. Incentivos do Modelo de Negócio das Plataformas e Efeitos Democráticos

Dimensão	Incentivo da Plataforma	Mecanismo Algorítmico	Efeito Democrático
Tempo de permanência	Maximizar minutos por sessão	Amplificação de conteúdo de alta retenção emocional	Consumo excessivo de conteúdo emocionalmente ativador em detrimento do informativo
Engajamento	Maximizar curtidas, compartilhamentos, comentários	Priorização de conteúdo que gera reação, especialmente indignação	Viralização de desinformação sobre informação verificada
Retenção de usuários	Reduzir churn (abandono da plataforma)	Personalização progressiva que cria dependência	Formação de bolhas de filtro e câmaras de eco
Crescimento de rede	Maximizar novos cadastros e conexões	Recomendação de grupos e comunidades com afinidade ideológica	Radicalização algorítmica progressiva
Receita publicitária	Maximizar cliques em anúncios	Targeting por perfil psicográfico e comportamental	Microtargeting eleitoral sem transparência para o eleitor

Fonte: Elaborado pelo autor com base em Zuboff (2019), Wu (2016) e Vosoughi, Roy e Aral (2018).

Transparência Algorítmica: o que as Plataformas Devem ao Eleitorado

A questão da transparência algorítmica no contexto eleitoral é, ao mesmo tempo, uma demanda técnica e uma exigência democrática de primeira ordem. Se os algoritmos de recomendação das plataformas têm impacto mensurável sobre a formação de preferências eleitorais — como demonstrado pelo experimento de Bond et al. (2012), que mostrou que mensagens de mobilização eleitoral no Facebook influenciaram diretamente o comparecimento às urnas de aproximadamente 340 mil eleitores nos Estados Unidos — então os cidadãos e suas representações eleitas têm direito legítimo a compreender como esses algoritmos funcionam e que tipo de conteúdo amplificam preferentemente (BOND et al., 2012, p. 295-298).

Esse direito não é abstrato. Ele tem um fundamento concreto na teoria democrática: decisões que afetam o processo eleitoral, que é o mecanismo central pelo qual os cidadãos

exercem soberania política, precisam ser inteligíveis e contestáveis. Quando um sistema algorítmico opaco decide, sem qualquer responsabilização pública, qual conteúdo político é amplificado para quais eleitores e em que momento do ciclo eleitoral, está exercendo um poder sobre o processo democrático que escapa inteiramente ao controle institucional que as democracias modernas desenvolveram para lidar com o poder. Como argumenta Zuboff (2019), a opacidade não é uma característica acidental desses sistemas: ela é funcional ao modelo de negócios, porque a capacidade de modificar comportamento em escala depende de que os sujeitos modificados não percebam o mecanismo que os modifica (ZUBOFF, 2019, p. 294-295).

A demanda por transparência algorítmica enfrenta, contudo, uma objeção legítima que precisa ser tomada a sério antes de ser respondida. As plataformas argumentam, com razão parcial, que a divulgação detalhada do funcionamento de seus sistemas de recomendação criaria vulnerabilidades de manipulação: atores maliciosos que conhecem com precisão os critérios de amplificação podem calibrar o conteúdo desinformativo para maximizar sua distribuição orgânica. Esse argumento não é suficiente para justificar a opacidade total, mas é suficiente para justificar que a transparência seja mediada por pesquisadores independentes credenciados, em vez de inteiramente pública, e que os dados disponibilizados sejam suficientemente agregados para não revelar parâmetros exploráveis individualmente.

É precisamente esse equilíbrio que o Digital Services Act europeu, em vigor desde agosto de 2023 para as chamadas Very Large Online Platforms, buscou operacionalizar. O DSA estabelece, entre outras obrigações, que plataformas com mais de 45 milhões de usuários mensais na União Europeia devem disponibilizar acesso a dados sobre o funcionamento de seus sistemas de recomendação a pesquisadores independentes credenciados pela Comissão Europeia, submeter-se a auditorias anuais independentes de suas avaliações de risco e publicar relatórios de transparência sobre decisões de moderação de conteúdo (UNIÃO EUROPEIA, 2022, art. 40). O modelo europeu não resolve todos os problemas de transparência algorítmica, mas representa o avanço regulatório mais substancial disponível como referência comparada.

No contexto brasileiro, a Resolução TSE nº 23.732/2024 estabeleceu, em seu artigo 27, obrigação similar no âmbito eleitoral específico: as plataformas devem disponibilizar, mediante requisição fundamentada do TSE, informações sobre o funcionamento de seus sistemas de recomendação de conteúdo durante o período eleitoral. A norma é um passo relevante, mas permanece aquém do padrão europeu em dois aspectos centrais. Primeiro, a transparência é reativa: depende de requisição caso a caso do TSE, em vez de ser

proativa e sistemática como no DSA. Segundo, não há previsão de acesso estruturado por pesquisadores independentes, o que limita a capacidade de produção de conhecimento sobre os efeitos algorítmicos no eleitorado brasileiro que poderia fundamentar respostas regulatórias mais precisas nos ciclos eleitorais seguintes.

Responsabilidade das Plataformas: Marco Civil, DSA e a Resolução TSE

O regime de responsabilidade das plataformas digitais pelo conteúdo que hospedam e amplificam é um dos temas mais debatidos e menos resolvidos do direito digital contemporâneo, e a tensão entre proteção à liberdade de expressão, viabilidade operacional das plataformas e proteção da integridade eleitoral não admite solução simples em nenhum ordenamento jurídico que leve todas essas dimensões a sério.

No Brasil, o Marco Civil da Internet, instituído pela Lei nº 12.965/2014, estabeleceu em seu artigo 19 o regime de responsabilidade que ainda hoje governa as relações entre plataformas e o sistema judicial: os provedores de aplicações de internet só podem ser responsabilizados por danos decorrentes de conteúdo gerado por terceiros se, após notificação judicial específica, deixarem de tomar as providências para tornar o conteúdo indisponível. Esse modelo, inspirado no notice-and-takedown norte-americano e projetado para proteger as plataformas de responsabilização excessiva por conteúdo que não produzem, tinha uma lógica defensável no contexto de 2014: evitar que plataformas se tornassem moderadoras compulsórias de todo conteúdo por receio de responsabilização, o que produziria efeito de chilling sobre a liberdade de expressão (LEONARDI, 2012, p. 176-180).

Esse regime, porém, revelou-se estruturalmente inadequado para o contexto eleitoral por uma razão simples e decisiva: a exigência de notificação judicial é incompatível com a urgência característica da desinformação eleitoral. Um deepfake disseminado 48 horas antes do pleito pode determinar o comportamento de parcela do eleitorado antes que qualquer ordem judicial seja prolatada, cumprida e verificada. O tempo do processo judicial e o tempo da desinformação eleitoral operam em ordens de grandeza diferentes, e um regime de responsabilidade que ignora essa assimetria está, na prática, oferecendo proteção jurídica às plataformas ao custo da proteção eleitoral aos cidadãos.

A Resolução TSE nº 23.732/2024 respondeu a esse problema criando um regime especial para o período eleitoral que contorna a exigência de notificação judicial do Marco Civil. Pelo artigo 27 da Resolução, as plataformas devem responder a notificações do TSE em até 24 horas, sob pena de multa diária de R$ 150.000 e responsabilização solidária

pelos danos decorrentes da manutenção do conteúdo irregular. Esse regime tem respaldo na jurisprudência consolidada do STF sobre a aplicação do artigo 19 do Marco Civil, que reconheceu a possibilidade de regimes especiais de responsabilidade para situações em que a demora na remoção produz danos irreversíveis de natureza qualificada, como os que afetam o processo eleitoral (STF, RE 1037396, 2023).

A comparação com o modelo europeu revela, uma vez mais, a distância entre o que foi construído e o que seria necessário. O DSA não apenas estabelece prazos de remoção: define obrigações de avaliação de risco sistêmico, exige que plataformas identifiquem e mitiguem proativamente os riscos que seus sistemas de recomendação representam para processos eleitorais, e prevê sanções de até 6% do faturamento global anual para violações graves (UNIÃO EUROPEIA, 2022, art. 35 e 74). O modelo brasileiro, centrado na remoção reativa de conteúdo específico mediante notificação, não alcança a dimensão sistêmica que o DSA endereça: o risco não está apenas em peças individuais de conteúdo irregular, mas nos padrões de amplificação que a arquitetura das plataformas produz agregadamente sobre o eleitorado.

Essa lacuna é, talvez, a mais difícil de preencher por meio da regulação eleitoral ordinária, porque ela exige que o direito eleitoral desenvolva instrumentos para responsabilizar plataformas não apenas pelo que hospedam, mas pelo que amplificam, e não apenas por conteúdo identificável, mas por efeitos sistêmicos que emergem da interação entre algoritmos e comportamento humano em escala. É o horizonte regulatório mais exigente do campo, e é precisamente para ele que a produção doutrinária e jurisprudencial precisa avançar nos próximos ciclos eleitorais.

O quadro a seguir sistematiza os quatro regimes de responsabilidade em perspectiva comparada, permitindo visualizar com precisão as diferenças estruturais entre os modelos disponíveis. A leitura por colunas é mais reveladora do que a leitura por linhas: o que mais distingue os regimes não é o conteúdo das proibições, que em todos os casos convergem para a remoção de conteúdo ilegal, mas o gatilho que aciona a responsabilidade, o prazo disponível para cumprimento e a magnitude das sanções pelo descumprimento. Um regime em que a responsabilidade só nasce após notificação judicial, sem prazo definido e com sanção limitada à indenização posterior, é estruturalmente diferente de um regime em que a identificação do conteúdo ilegal já aciona a obrigação, com multa proporcional ao faturamento global. A distância entre o modelo brasileiro geral e o modelo europeu, visível na tabela, é a distância entre uma regulação projetada para um ambiente com volume gerenciável de conteúdo e uma regulação projetada para plataformas que distribuem bilhões de peças de conteúdo por dia.

Tabela 32. Regimes de Responsabilidade das Plataformas: Brasil e Direito Comparado

Regime	Base Legal	Gatilho de Responsabilidade	Prazo de Remoção	Sanção pelo Descumprimento
Brasil (regime geral)	Marco Civil da Internet, art. 19	Notificação judicial específica.	Após ordem judicial.	Responsabilização por danos após omissão.
Brasil (regime eleitoral)	Resolução TSE 23.732/2024, art. 27	Notificação formal do TSE ou TRE.	24 horas.	Multa diária de R$ 150.000; responsabilidade solidária pelos danos.
União Europeia (VLOPs)	Digital Services Act (DSA), art. 34	Identificação de conteúdo ilegal pela plataforma ou notificação de usuário.	Célere; sem prazo fixo para todo conteúdo.	Multa de até 6% do faturamento global anual.
Estados Unidos	Section 230 do Communications Decency Act	Conhecimento efetivo e ausência de ação (good samaritan protection).	Voluntária; sem obrigação legal geral.	Ausência de responsabilidade civil na maior parte dos casos.

Fonte: Elaborado pelo autor com base em Lei 12.965/2014, Resolução TSE 23.732/2024, DSA (2022) e Section 230 (1996).

O Facebook e o Caso Cambridge Analytica: Cumplicidade Estrutural

O caso Cambridge Analytica, analisado em profundidade no Capítulo 5, permanece o exemplo mais documentado de cumplicidade estrutural de uma plataforma com a manipulação eleitoral em escala, e sua relevância analítica não diminuiu com o tempo: ele estabeleceu o padrão pelo qual casos subsequentes são avaliados e revelou mecanismos de vulnerabilidade que permanecem parcialmente ativos em diferentes plataformas.

A API de amigos do Facebook, que permitiu o acesso aos dados de aproximadamente 87 milhões de usuários sem o consentimento da maioria deles, não foi uma vulnerabilidade acidental descoberta por agentes maliciosos: foi uma funcionalidade deliberadamente projetada para maximizar o engajamento de desenvolvedores de aplicativos com a plataforma, criando o ecossistema de terceiros que o Facebook precisava para crescer. A decisão de construir essa funcionalidade sem os controles de privacidade que teriam impedido seu uso abusivo foi uma decisão de negócios, não um descuido técnico. Como documentaram Cadwalladr e Graham-Harrison (2018) em sua investigação para o *The Guardian* e *The Observer*, os engenheiros do Facebook conheciam os riscos do

acesso em cascata a dados de amigos e optaram por não implementar as restrições que os teriam mitigado, porque essas restrições reduziriam a atratividade da plataforma para desenvolvedores (CADWALLADR; GRAHAM-HARRISON, 2018).

A resposta regulatória americana ao caso foi, sob qualquer avaliação honesta, insuficiente para alterar os incentivos que produziram a vulnerabilidade. A multa de 5 bilhões de dólares aplicada pela Federal Trade Commission ao Facebook em julho de 2019 foi a maior da história da agência, mas representava, no momento da aplicação, aproximadamente 23 dias de receita da empresa, o que levou críticos a argumentar que o valor não constituía desincentivo efetivo à repetição de condutas similares (FEDERAL TRADE COMMISSION, 2019). A lógica econômica é direta: se o benefício esperado de uma prática que viola regulações é maior do que o custo esperado da multa descontada pela probabilidade de detecção, a empresa racional continuará a prática. Uma multa de 5 bilhões de dólares sobre uma empresa com receita anual de 70 bilhões não altera essa equação de forma decisiva.

A resposta do Facebook ao escândalo foi a criação de mecanismos de auditoria mais robustos para aplicativos de terceiros, a implementação de restrições mais estritas ao acesso a dados de usuários e a constituição, em 2020, do Oversight Board, órgão externo de revisão de decisões de moderação de conteúdo. O modelo de negócios fundamentado na monetização de dados comportamentais, contudo, permaneceu intacto. Como observa Zuboff (2019), reformas cosméticas que não alteram a lógica extrativista do capitalismo de vigilância não constituem resposta estrutural ao problema que Cambridge Analytica tornou visível: elas constituem gestão de reputação (ZUBOFF, 2019, p. 512-514).

A lição mais duradoura do caso não é a que as plataformas costumam reconhecer nas suas comunicações públicas. Não é que desenvolvedores maliciosos exploraram uma vulnerabilidade que foi corrigida. É que o design de sistemas com acesso irrestrito a dados de terceiros sem consentimento informado é uma escolha arquitetural com consequências para a integridade democrática que o modelo regulatório vigente, tanto nos Estados Unidos quanto no Brasil, não estava equipado para prevenir ou sancionar adequadamente.

Moderação de Conteúdo Eleitoral: Limites e Contradições

A moderação de conteúdo eleitoral pelas plataformas é um problema genuinamente difícil que não tem solução técnica simples, e qualquer análise honesta precisa reconhecer essa dificuldade antes de apontar suas insuficiências. A dificuldade não é operacional: é

normativa. Ela deriva de tensões entre valores democráticos fundamentais que não se resolvem por algoritmo.

A tensão central é conhecida e estruturalmente persistente. Moderação insuficiente permite a proliferação de desinformação, discurso de ódio e manipulação coordenada que comprometem a qualidade do debate eleitoral e podem influenciar resultados. Moderação excessiva ou enviesada suprime vozes legítimas, produz efeito de chilling sobre a expressão política e coloca nas mãos de empresas privadas o poder de definir os limites do discurso democrático aceitável, sem qualquer legitimidade democrática para exercê-lo.

Como argumenta Balkin (2018), as plataformas tornaram-se infraestrutura pública de comunicação política sem jamais terem sido investidas de responsabilidades constitucionais correspondentes a esse papel (BALKIN, 2018, p. 2016-2018).

As plataformas adotaram, nos ciclos eleitorais mais recentes, estratégias distintas para navegar essa tensão, com resultados que revelam mais sobre suas prioridades institucionais do que sobre sua capacidade técnica.

O Facebook criou o Oversight Board em 2020 como órgão externo de revisão de decisões de moderação, composto por juristas, acadêmicos e defensores de direitos humanos de múltiplas regiões do mundo. A iniciativa foi recebida com ceticismo por parte da comunidade de pesquisa, e os anos de operação do Board confirmaram parte dessas reservas: sua jurisdição é limitada a casos individuais levados à revisão, suas decisões não se aplicam automaticamente a conteúdo similar em outros contextos, e a Meta retém o poder de não implementar recomendações do Board que considera inconsistentes com seus padrões de comunidade (OVERSIGHT BOARD, 2023). O Board funciona, na prática, como mecanismo de legitimação seletiva de decisões difíceis, não como sistema de governança abrangente da moderação eleitoral.

O Twitter, rebatizado X após a aquisição por Elon Musk em outubro de 2022, adotou trajetória oposta e igualmente reveladora. A redução de aproximadamente 80% da equipe de confiança e segurança, a readmissão de contas previamente banidas por violações de política e a flexibilização das regras sobre conteúdo político produziram, segundo pesquisadores do Instituto de Tecnologia de Massachusetts e da Universidade de Edimburgo, aumento mensurável na proliferação de desinformação e discurso de ódio na plataforma nos meses seguintes à aquisição (CRUICKSHANK; CARLEY, 2023). O experimento do X é analiticamente valioso precisamente porque funciona como controle: ele permite estimar, por contraste, o efeito das políticas de moderação que estavam em vigor anteriormente.

O TikTok representa um caso distinto e com características próprias relevantes para o contexto eleitoral brasileiro. A plataforma implementou restrições específicas sobre publicidade política paga e sobre conteúdo que desinforma sobre procedimentos eleitorais, mas a combinação de opacidade algorítmica, predominância de conteúdo em formato de vídeo curto de difícil moderação em escala e concentração de usuários nas faixas etárias de 18 a 29 anos cria um perfil de risco específico. Eleitores de primeira viagem, que ainda estão formando suas referências políticas e que têm menor exposição à mídia tradicional como contraponto, são o público mais exposto ao ecossistema informacional do TikTok e, ao mesmo tempo, o menos estudado em termos de impacto eleitoral no contexto brasileiro (NEWMAN et al., 2023, p. 42-45).

A análise comparada dessas três estratégias converge para uma conclusão que o debate regulatório precisa incorporar: não existe arquitetura de moderação de conteúdo eleitoral que satisfaça simultaneamente todas as exigências normativas em jogo. O que é possível construir são regimes que tornem as decisões de moderação mais transparentes, mais contestáveis e mais sujeitas a responsabilização externa, reduzindo o espaço para arbitrariedade sem pretender eliminá-lo. Esse é o caminho que o DSA tenta pavimentar e que a regulação eleitoral brasileira ainda percorre em estágios iniciais.

■ RESUMO DO CAPÍTULO:

Este capítulo analisou a relação entre plataformas digitais e democracia a partir de uma premissa que atravessa toda a sua argumentação: as plataformas não são neutras, e sua não-neutralidade não é acidental nem conspirativa, mas estrutural — decorre do modelo de negócios fundado na monetização da atenção humana por publicidade. O ponto de partida foi a tensão entre o paradigma otimista, dominante na primeira década do século XXI e ancorado em evidências reais como as mobilizações das primaveras árabes, e o paradigma crítico, que ganhou força decisiva após 2016 com a acumulação de evidências sobre os efeitos sistêmicos das plataformas sobre o debate democrático. Quatro autores estruturaram a análise teórica: Zuboff, com o conceito de capitalismo de vigilância e a lógica de modificação comportamental em escala; Pariser, com o mecanismo das bolhas de filtro e a fragmentação do espaço informacional comum; Sunstein, com a erosão da experiência cívica compartilhada como condição da deliberação democrática; e Wu, com a genealogia histórica do capitalismo de atenção que situa as plataformas digitais numa continuidade com formas anteriores de revenda da atenção humana. Vosoughi, Roy e Aral forneceram a evidência empírica que corroborou as hipóteses teóricas: notícias falsas se disseminam seis vezes mais rapidamente do que notícias verdadeiras, e os humanos, não os bots, são o principal vetor de propagação. A análise do modelo de negócio demonstrou que os efeitos antidemocráticos das plataformas não são intencionais em sentido político, mas são estruturalmente determinados pelos incentivos econômicos: maximizar tempo de permanência, engajamento, retenção, crescimento de rede e receita publicitária produz, como resultado agregado e previsível, amplificação de conteúdo emocionalmente ativador, polarizador e frequentemente impreciso. A implicação regulatória central é que reformas centradas na remoção de conteúdo específico não endereçam o problema, porque não alteram os incentivos que o produzem. A questão da transparência algorítmica foi analisada como exigência democrática de primeira ordem: decisões que afetam o processo eleitoral precisam ser inteligíveis e contestáveis. O experimento de

Bond et al. (2012), que demonstrou impacto mensurável de mensagens algorítmicas sobre o comparecimento eleitoral de 340 mil eleitores americanos, ancorará qualquer debate sério sobre o tema. A comparação entre o DSA europeu e a Resolução TSE nº 23.732/2024 revelou dois déficits persistentes do modelo brasileiro: a transparência é reativa em vez de proativa, e não há acesso estruturado para pesquisadores independentes. O regime de responsabilidade das plataformas foi analisado em perspectiva comparada entre quatro modelos: o regime geral brasileiro do Marco Civil, o regime eleitoral especial da Resolução TSE, o modelo europeu do DSA e o modelo americano da Section 230. A comparação evidenciou a distância entre responsabilização reativa por conteúdo identificável e responsabilização sistêmica pelos padrões de amplificação, que é o horizonte regulatório mais exigente e o que a regulação brasileira ainda não alcançou. O caso Cambridge Analytica foi retomado como o exemplo mais documentado de cumplicidade estrutural entre design de plataforma e manipulação eleitoral, com ênfase na insuficiência da resposta regulatória americana e na lição mais duradoura do caso: que escolhas arquiteturais têm consequências democráticas que o modelo regulatório vigente não estava equipado para prevenir. A análise comparada das estratégias de moderação do Facebook, X e TikTok encerrou o capítulo com a conclusão de que não existe arquitetura de moderação que satisfaça simultaneamente todas as exigências normativas em jogo, mas que é possível construir regimes mais transparentes, contestáveis e responsabilizáveis.

QUESTÕES PARA REFLEXÃO E DEBATE

1. O paradigma otimista sobre o potencial democrático das plataformas digitais tinha fundamentos empíricos reais nas mobilizações das primaveras árabes. Em que medida esses fundamentos eram genuínos e em que medida refletiam uma subestimação dos usos autoritários e antidemocráticos das mesmas tecnologias que já estavam em curso simultaneamente?

2. O conceito de capitalismo de vigilância de Zuboff descreve um modelo de negócio que extrai dados comportamentais para modificar comportamento em escala. Em que medida as campanhas eleitorais que utilizam microtargeting sofisticado reproduzem, no contexto político, a mesma lógica que Zuboff identifica nas plataformas comerciais?

3. Pariser e Sunstein chegaram a conclusões convergentes sobre a fragmentação do espaço informacional por caminhos teóricos distintos. As evidências empíricas disponíveis desde 2011 corroboram, qualificam ou refutam suas hipóteses? Qual dos dois frameworks analíticos é mais útil para pensar o contexto eleitoral brasileiro?

4. O estudo de Vosoughi, Roy e Aral demonstrou que os humanos, e não os bots, são o principal vetor de propagação da desinformação. Que implicações essa conclusão tem para estratégias de combate à desinformação que se concentram na detecção e remoção de contas automatizadas?

5. A análise do modelo de negócio das plataformas sugere que reformas centradas na remoção de conteúdo não endereçam o incentivo estrutural que produz a amplificação de desinformação. Que modelos regulatórios alternativos poderiam alterar esses incentivos? A tributação sobre engajamento, a proibição de publicidade comportamental ou a obrigação de oferecer feeds cronológicos são respostas proporcionais?

6. O experimento de Bond et al. (2012) demonstrou que mensagens algorítmicas no Facebook influenciaram o comparecimento eleitoral de 340 mil eleitores americanos. Esse experimento foi conduzido sem o consentimento informado dos participantes. Que limites éticos e jurídicos deveriam governar a realização de experimentos comportamentais por plataformas sobre seus usuários em contexto eleitoral?

7. A comparação entre o DSA europeu e a Resolução TSE nº 23.732/2024 revelou dois déficits do modelo brasileiro: reatividade e ausência de acesso para pesquisadores independentes. Qual dos dois déficits é mais urgente de endereçar para o ciclo eleitoral de 2026, e que instrumento normativo seria mais adequado para fazê-lo?

8. O caso Cambridge Analytica demonstrou que uma multa de 5 bilhões de dólares representando menos de um mês de receita não constitui desincentivo efetivo. O modelo do DSA, com multas de até 6% do faturamento global anual, resolve o problema do desincentivo insuficiente, ou o cálculo econômico das plataformas ainda favorece a conduta irregular mesmo nesse patamar?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. O algoritmo da plataforma não trabalha para você. Ele trabalha para o anunciante.

Toda estratégia de comunicação eleitoral que depende de alcance orgânico nas plataformas está, na prática, terceirizando decisões de distribuição para sistemas que foram otimizados para maximizar engajamento publicitário, não para transmitir mensagens com precisão ao eleitorado-alvo. Isso não significa abandonar as plataformas: significa não confundir alcance algorítmico com comunicação estratégica. Conteúdo que vai viralizar organicamente é, por definição, conteúdo que ativa respostas emocionais intensas. A campanha que não planeja explicitamente que tipo de resposta emocional quer ativar está deixando essa decisão para o algoritmo.

2. Você está competindo com conteúdo projetado para indignar. Adapte sua comunicação.

A pesquisa de Vosoughi, Roy e Aral (2018) demonstrou que conteúdo falso se propaga seis vezes mais rapidamente do que conteúdo verdadeiro precisamente porque tende a ser mais novo, mais surpreendente e mais emocionalmente ativador. A campanha que responde a desinformação com comunicados factuais equilibrados está trazendo uma faca para uma batalha de espadas. Isso não é uma justificativa para desinformar: é um diagnóstico que exige que comunicação factual seja produzida com atenção deliberada à sua capacidade de ativação emocional legítima. Fatos apresentados de forma emocionalmente inerte não competem com falsidades emocionalmente carregadas.

3. Bolha de filtro trabalha para você e contra você ao mesmo tempo.

O mesmo mecanismo que garante que sua mensagem chegue repetidamente aos seus apoiadores também garante que ela raramente alcance eleitores indecisos nos ambientes informacionais do adversário. Campanhas que medem sucesso exclusivamente por engajamento da base estão medindo a eficácia da bolha, não a eficácia eleitoral. O problema é ainda mais agudo nas plataformas de vídeo curto como o TikTok: algoritmos de recomendação que funcionam por comportamento, não por rede social, criam bolhas mais permeáveis a novos usuários, mas também mais imprevisíveis em termos de qual eleitorado será alcançado com qual mensagem.

4. O Marco Civil protege as plataformas, não o seu candidato.

O artigo 19 do Marco Civil da Internet exige notificação judicial para que uma plataforma seja responsabilizada por danos de conteúdo de terceiros. Em período eleitoral, isso é operacionalmente inútil: um deepfake disseminado 48 horas antes do pleito pode percorrer milhões de dispositivos antes que qualquer ordem judicial seja prolatada. A Resolução TSE nº 23.732/2024 criou um regime especial com prazo de 24 horas, mas o WhatsApp, onde a desinformação mais circula, escapa a qualquer regime de remoção. A estratégia preventiva, ter evidências técnicas, comunicados públicos e canais de resposta prontos antes do ataque, é mais eficaz do que a estratégia reativa.

5. Audite o que sua agência faz com seus dados de campanha.

Cambridge Analytica começou com acesso legítimo a dados de usuários do Facebook e terminou com a manipulação eleitoral mais documentada da história recente. O mecanismo foi simples: um aplicativo aparentemente inofensivo coletou dados de 87 milhões de usuários por meio de uma API que o Facebook havia construído sem controles adequados. Para o assessor de campanha e para o advogado eleitoral: os dados que sua campanha coleta sobre eleitores, os dados que fornece a empresas de microtargeting e os dados que essas empresas compartilham com terceiros criam uma cadeia de responsabilidade solidária sob a LGPD e sob a Resolução TSE nº 23.732/2024. Contratos com fornecedores de tecnologia eleitoral precisam especificar quem detém os dados, por quanto tempo, com quem podem ser compartilhados e o que acontece com eles após o pleito.

6. Transparência não é apenas obrigação legal. É vantagem estratégica.

A Resolução TSE nº 23.732/2024 exige identificação de conteúdo produzido por IA e transparência sobre impulsionamento pago. Campanhas que tratam essas obrigações como ônus regulatório a ser minimizado estão perdendo uma oportunidade estratégica: num ambiente em que o eleitor desconfia crescentemente da autenticidade do conteúdo que recebe, a identificação proativa de IA e a transparência sobre patrocínio podem funcionar como diferencial de credibilidade. A transparência que vai além do exigido sinaliza confiança. A opacidade que fica exatamente no limite do permitido sinaliza que haveria mais a esconder se pudesse.

7. O TikTok em 2026 é o WhatsApp de 2018. Prepare-se agora.

Nas eleições de 2018, o WhatsApp foi o canal de desinformação dominante e as campanhas e autoridades foram pegas completamente despreparadas. Em 2026, o TikTok ocupa posição equivalente: crescimento acelerado entre eleitores jovens, algoritmo de recomendação opaco, moderação de vídeo em escala tecnicamente complexa e ausência de regulação específica consolidada. O perfil de risco é diferente do WhatsApp, porque o TikTok é aberto e rastreável, mas o volume e a velocidade de disseminação são comparáveis. Campanhas que não têm estratégia específica para o TikTok em 2026 estão repetindo o erro de 2018 com plataforma diferente.

CAPÍTULO 9

Os Financiadores Invisíveis da Desinformação

Atores, Motivações, Financiamento e a Ecologia da Desinformação Eleitoral Brasileira

A Economia Política da Desinformação

A literatura sobre desinformação eleitoral atravessou, na última década, uma transição analítica fundamental que ainda não foi inteiramente incorporada pelo debate regulatório brasileiro. A abordagem predominante na primeira fase do campo, centrada nos mecanismos tecnológicos de produção e distribuição de conteúdo falso, cedeu progressivamente espaço a uma perspectiva de economia política que coloca no centro da análise perguntas distintas e mais politicamente incômodas: quem financia a desinformação, por que, com quais recursos e em benefício de quem.

Essa mudança de foco não é cosmética. Tratar a desinformação como fenômeno tecnológico ou cultural, algo que "acontece" nas plataformas ou que reflete preferências culturais de determinados segmentos da população, desvia a atenção dos atores com agência, recursos e motivações identificáveis que a sustentam e reproduzem deliberadamente.

Yochai Benkler, Robert Faris e Hal Roberts, em *Network Propaganda* (2018), oferecem a análise empírica mais abrangente e metodologicamente rigorosa sobre a economia política da desinformação eleitoral no contexto americano. Analisando um corpus de 1,25 milhão de matérias publicadas durante as eleições presidenciais de 2016, os autores demonstraram que a assimetria de desinformação entre os campos políticos nos Estados Unidos não é produto do acaso, de preferências culturais difusas ou da lógica autônoma das plataformas: é produto de investimentos deliberados em infraestrutura de mídia e desinformação por parte de atores específicos com recursos identificáveis e estratégias de longo prazo. A conclusão central de Benkler, Faris e Roberts é metodologicamente relevante para qualquer análise comparada: desinformação em escala não emerge espontaneamente do ecossistema digital, ela é construída e mantida por agentes com interesse no resultado (BENKLER; FARIS; ROBERTS, 2018, p. 9-13).

Philip Howard, em *Lie Machines* (2020), amplia o escopo geográfico da análise para sistematizar o conhecimento sobre operações de desinformação eleitoral ao redor do mundo, identificando padrões recorrentes que transcendem os contextos nacionais específicos. Howard descreve o que denomina "máquinas de mentira" como sistemas organizados compostos por três elementos interdependentes: produtores de conteúdo,

redes de distribuição e fontes de financiamento. A contribuição analítica central de Howard é precisamente a de que esses três elementos raramente coincidem no mesmo ator: quem financia frequentemente não é quem produz, e quem distribui frequentemente não sabe quem financia. Essa dissociação é deliberada e funcionalmente importante, porque dificulta a atribuição de responsabilidade e a identificação dos incentivos reais que sustentam a operação (HOWARD, 2020, p. 5-8).

No contexto brasileiro, as análises empíricas mais robustas sobre o ecossistema de desinformação eleitoral provêm de dois centros de pesquisa que desenvolveram metodologias específicas para o ambiente informacional nacional. O FGV DAPP, com seu monitoramento sistemático de redes sociais e aplicativos de mensageria, produziu análises quantitativas sobre os padrões de disseminação de desinformação nas eleições de 2018, 2020 e 2022 que documentaram empiricamente a centralidade do WhatsApp como canal de distribuição e a estrutura de redes coordenadas de amplificação. O NetLab da UFRJ, com abordagem complementar centrada na análise de infraestrutura e financiamento, contribuiu com mapeamentos dos ecossistemas de sites de desinformação e de suas fontes de monetização. Juntos, esses trabalhos estabeleceram que o ecossistema brasileiro de desinformação eleitoral tem características estruturais próprias que o distinguem do modelo americano descrito por Benkler, Faris e Roberts: a concentração no WhatsApp em vez de plataformas abertas, a importância de redes religiosas e empresariais como canais de amplificação, e a combinação de financiamento formal e informal que torna o rastreamento especialmente complexo (ARAÚJO; OLIVEIRA, 2021, p. 34-39).

A transição da abordagem tecnológica para a abordagem de economia política tem uma implicação regulatória direta que merece ser enunciada com precisão: enquanto o direito eleitoral se concentrar exclusivamente na remoção de conteúdo e na responsabilização das plataformas, estará endereçando os sintomas sem tocar nas causas. A desinformação industrial persiste porque é financiada, e ela é financiada porque produz retornos, sejam eles políticos, econômicos ou ambos. Qualquer estratégia regulatória que ignore essa dimensão estrutural está condenada a uma corrida armamentista que os produtores de desinformação, com seus incentivos intactos, têm condições sistêmicas de vencer.

A Ecologia da Desinformação: Produtores, Distribuidores e Financiadores

A metáfora ecológica para descrever o ecossistema de desinformação eleitoral é analiticamente útil precisamente porque captura o que abordagens centradas em atores

individuais tendem a obscurecer: a interdependência entre diferentes componentes do sistema, cada um com papel específico, recursos distintos e motivações próprias, que juntos produzem efeitos que nenhum deles produziria isoladamente. Como observa Howard (2020), nenhum elemento da ecologia pode ser compreendido em isolamento: fábricas de conteúdo dependem de financiadores para operar em escala; redes de amplificação dependem de infraestrutura técnica e de coordenação para distribuir com eficiência; e o impacto eleitoral depende da convergência de todos os elementos no momento certo do ciclo eleitoral (HOWARD, 2020, p. 12-14).

Essa perspectiva ecológica tem uma implicação analítica importante para o direito eleitoral: intervenções cirúrgicas sobre um único componente do ecossistema, como a remoção de uma conta específica ou o bloqueio de um site de desinformação, produzem efeitos limitados quando os demais componentes permanecem intactos e funcionais. O ecossistema é adaptativo: quando um canal é bloqueado, o conteúdo migra para outro; quando uma conta é removida, outras assumem sua função de amplificação; quando um financiador é identificado, a estrutura de pagamento se fragmenta em microtransações mais difíceis de rastrear. A resposta regulatória eficaz precisa ser sistêmica, endereçando simultaneamente os produtores, os distribuidores e os financiadores, com instrumentos calibrados para a especificidade de cada elo da cadeia.

Os três componentes do ecossistema têm perfis distintos que a regulação eleitoral precisa tratar de forma diferenciada. Os produtores de conteúdo incluem desde operações profissionais com equipes dedicadas, orçamentos significativos e capacidade de produção em escala industrial até criadores individuais motivados por ideologia ou por monetização publicitária. Os distribuidores abrangem redes de contas automatizadas, grupos coordenados de amplificação manual e intermediários humanos, como líderes religiosos, empresários e influenciadores, que redistribuem conteúdo nos ambientes de alta credibilidade percebida onde é mais eficaz. Os financiadores, o elo mais opaco e mais consequente do ecossistema, incluem campanhas eleitorais, grupos empresariais, associações ideológicas e, em casos documentados internacionalmente, atores estrangeiros com interesse em resultados eleitorais domésticos.

A análise de Benkler, Faris e Roberts (2018) sugere que o elemento mais determinante para a escala e a persistência de um ecossistema de desinformação não é a sofisticação tecnológica de seus produtores nem o alcance de seus distribuidores: é a estabilidade e o volume de seu financiamento. Ecossistemas bem financiados são resilientes a intervenções regulatórias pontuais porque têm recursos para adaptar sua infraestrutura, migrar para novos canais e recrutar novos produtores e distribuidores quando

os anteriores são identificados e removidos. Ecossistemas mal financiados, por contraste, tendem a ser mais frágeis e mais sensíveis a intervenções que aumentam seus custos operacionais (BENKLER; FARIS; ROBERTS, 2018, p. 321-324).

Essa conclusão orienta a estratégia regulatória mais eficaz disponível: não a remoção de conteúdo específico, que os produtores podem substituir mais rapidamente do que qualquer sistema de moderação consegue acompanhar, mas o aumento dos custos de financiamento e a redução dos retornos esperados das operações de desinformação. Instrumentos como o rastreamento de fluxos financeiros, a responsabilização solidária de financiadores identificados e a eliminação de fontes de monetização publicitária para sites de desinformação atuam precisamente sobre o elo que mais determina a capacidade sistêmica do ecossistema.

O quadro a seguir materializa essa perspectiva ecológica em termos operacionais, mapeando os seis atores centrais do ecossistema de desinformação eleitoral, seus papéis específicos, motivações, recursos necessários e vulnerabilidade à regulação. A coluna de vulnerabilidade é a mais reveladora para fins de desenho de política pública: ela mostra que os atores mais visíveis do ecossistema, como as fábricas de conteúdo e as redes de amplificação, são paradoxalmente mais vulneráveis à intervenção regulatória do que os financiadores, que operam na camada menos visível mas mais determinante da cadeia. O elo mais difícil de alcançar não é necessariamente o mais sofisticado tecnicamente: é o que opera com maior distância entre sua ação e o conteúdo que ela financia, e que pode invocar proteções jurídicas legítimas, como o sigilo bancário e a presunção de licitude de transações comerciais, para dificultar o rastreamento. A regulação que não consegue chegar ao financiador está regulando os sintomas sem tocar na causa.

Tabela 33. A Ecologia da Desinformação Eleitoral: Atores e Papéis

Ator	Papel na Ecologia	Motivação Principal	Recursos Necessários	Vulnerabilidade à Regulação
Fábricas de conteúdo	Produção do conteúdo falso ou distorcido: textos, imagens, áudios, vídeos.	Ideológica ou financeira (contratados por encomenda).	Habilidades técnicas; acesso a ferramentas de IA generativa; conhecimento do contexto.	Alta se identificadas; a IA generativa reduziu o capital humano necessário.
Redes de amplificação	Amplificação artificial do	Financeira (contratados);	Infraestrutura técnica	Média: bots detectáveis;

(bots + cyborgs)	conteúdo para criar percepção de popularidade e acionar algoritmos das plataformas.	ideológica (voluntários).	(servidores, chips, contas); know-how de automação.	cyborgs e sockpuppets mais difíceis.
Influenciadores e líderes de opinião	Dar credibilidade ao conteúdo ao republicá-lo para suas bases, transformando desinformação em informação aparentemente legítima.	Ideológica (acreditam no conteúdo); financeira (recebem por distribuição).	Alcance orgânico; credibilidade junto ao público-alvo.	Baixa: protegidos pela liberdade de expressão exceto em casos de dolo.
Financiadores	Prover recursos financeiros para a operação de toda a cadeia.	Política (favorecer candidatos); econômica (contratos e favores).	Capital financeiro; conexões com a cadeia operacional.	Alta se identificados: abuso de poder econômico; caixa dois eleitoral.
Plataformas digitais	Distribuir e amplificar o conteúdo através de algoritmos de recomendação.	Econômica: engajamento gera receita publicitária.	Infraestrutura técnica global; dados de bilhões de usuários.	Crescente: DSA, Resolução TSE 23.732/2024.
Eleitores receptores	Consumir, acreditar e redistribuir o conteúdo, completando o ciclo de amplificação orgânica.	Pertencimento grupal; confirmação de crenças preexistentes.	Smartphones; acesso à internet.	Baixa: boa-fé presumida; vulnerabilidade é estrutural, não individual.

Fonte: Elaborado pelo autor com base em Benkler, Faris e Roberts (2018), Howard (2020) e Wardle e Derakhshan (2017).

Tipos de Desinformação: da Fake News à Mal-information

A análise da ecologia da desinformação exige precisão terminológica que o debate público frequentemente não oferece e que o debate jurídico não pode dispensar. O problema terminológico não é apenas acadêmico: ele tem consequências diretas sobre a tipificação de condutas, a atribuição de responsabilidade e o desenho de instrumentos

regulatórios. Tratar como equivalentes fenômenos que diferem fundamentalmente em sua estrutura, em suas causas e em seus efeitos é um erro analítico com custos práticos mensuráveis.

O termo fake news, popularizado na campanha presidencial americana de 2016 e rapidamente incorporado ao vocabulário político global, é simultaneamente impreciso e instrumentalizado de formas que o tornam inadequado como categoria analítica. A imprecisão decorre de sua amplitude: o termo tem sido aplicado indistintamente a boatos compartilhados de boa-fé por usuários desinformados, a conteúdo satírico mal identificado, a erros jornalísticos não intencionais, a distorções deliberadas de fatos reais e a fabricações completas produzidas por operações coordenadas de influência. Fenômenos com estruturas tão distintas não podem ser adequadamente compreendidos, prevenidos ou sancionados sob a mesma categoria. A instrumentalização agrava o problema: como documentaram Wardle e Derakhshan (2017), o termo foi progressivamente apropriado por atores políticos para desqualificar toda cobertura jornalística crítica, transformando um conceito descritivo em arma retórica cujo uso mina a própria possibilidade de distinguir informação de desinformação no espaço público (WARDLE; DERAKHSHAN, 2017, p. 5).

A tipologia proposta por Wardle e Derakhshan em *Information Disorder* (2017), elaborada para o Conselho da Europa, representa o esforço de sistematização mais rigoroso e mais amplamente adotado disponível na literatura. Sua contribuição central é substituir a categoria unitária e imprecisa de fake news por uma tipologia tripartite construída sobre dois critérios analíticos independentes: a falsidade do conteúdo, que pergunta se o que circula é factualmente incorreto, e a intenção do emissor, que pergunta se há propósito deliberado de causar dano. O cruzamento desses dois critérios produz três categorias que, como argumentado no Capítulo 7, têm implicações normativas distintas para o direito eleitoral e exigem respostas institucionais diferenciadas.

A misinformation, primeiro tipo, é informação falsa compartilhada sem intenção de causar dano: quem a distribui acredita genuinamente no que dissemina. Sua dinâmica é de erro, não de malícia, e isso importa para a atribuição de responsabilidade: o emissor de misinformation não é cúmplice de uma operação, é frequentemente sua vítima. A disinformation, segundo tipo, é informação falsa criada e disseminada com propósito deliberado de enganar: o emissor sabe que o conteúdo é falso e age assim mesmo, o que a situa no campo da conduta dolosa passível de sanção eleitoral, civil e eventualmente penal. A malinformation, terceiro tipo, é informação verdadeira usada com propósito de causar dano: a falsidade não está no conteúdo, mas no uso estratégico de fatos reais para fins que seus produtores não autorizaram ou que distorcem seu significado original. Essa

terceira categoria é particularmente relevante no contexto eleitoral brasileiro, onde a divulgação estratégica de conversas privadas, a utilização de estatísticas reais em comparações enganosas e o doxxing de adversários políticos constituem formas recorrentes de manipulação que escapam às categorias tradicionais de desinformação (WARDLE; DERAKHSHAN, 2017, p. 20-22).

A adoção dessa tipologia pelo direito eleitoral brasileiro ainda é incipiente e desigual. A Resolução TSE nº 23.732/2024 trabalha predominantemente com a categoria da disinformation ao vedar conteúdo sintético que represente falsamente candidatos, mas não oferece instrumentos específicos para endereçar a misinformation em escala, que depende de educação midiática e não de sanção, nem a malinformation, que exige enquadramentos jurídicos mais sofisticados do que os disponíveis na legislação eleitoral ordinária. Essa lacuna não é necessariamente uma crítica à Resolução: é o reconhecimento de que um único instrumento normativo não pode dar conta da heterogeneidade do fenômeno que a tipologia de Wardle e Derakhshan revela. Para facilitar o entendimento, verifique a tabela a seguir:

Tabela 34. Tipologia da Desinformação (Wardle e Derakhshan, 2017)

Categoria	Definição	Intenção	Exemplos Eleitorais	Implicação Jurídica
Mis-information	Conteúdo falso ou impreciso difundido sem intenção deliberada de causar dano.	Ausência de dolo: emissor acredita genuinamente no conteúdo.	Eleitor que compartilha notícia falsa sobre candidato acreditando que é verdadeira.	Sem responsabilidade eleitoral em regra; relevante para programas de literacia midiática.
Dis-information	Conteúdo falso criado e difundido com intenção deliberada de causar dano.	Dolo direto: emissor sabe da falsidade e age para enganar.	Deepfake de candidato; campanha coordenada de disparos com conteúdo fabricado.	Responsabilidade eleitoral plena; potencial crime eleitoral de calúnia ou propaganda enganosa.
Mal-information	Conteúdo verdadeiro utilizado com intenção de causar dano, fora de contexto ou com	Dolo quanto ao uso; conteúdo em si é verdadeiro.	Vídeo autêntico de candidato editado seletivamente para distorcer o sentido; vazamento estratégico de	Zona cinzenta: o conteúdo é verdadeiro, mas o uso é malicioso; aplicação dos tipos eleitorais exige análise caso a caso.

	enquadramento distorcido.		informação privada.	

Fonte: Elaborado pelo autor com base em Wardle e Derakhshan (2017) e legislação eleitoral brasileira.

Financiamento da Desinformação no Brasil: o Caso dos Disparos em Massa de 2018

O caso mais documentado de financiamento organizado de desinformação eleitoral no Brasil é o dos disparos em massa pelo WhatsApp nas eleições presidenciais de 2018, e sua relevância analítica para o direito eleitoral contemporâneo não diminuiu com o tempo: ele estabeleceu os padrões operacionais que continuaram sendo utilizados nos ciclos eleitorais subsequentes, com variações técnicas mas com estrutura organizacional essencialmente idêntica.

As investigações conduzidas pelo TSE, pelo Ministério Público Eleitoral e pela Polícia Federal identificaram a existência de esquemas em que grupos de empresários, operando sem o conhecimento formal dos candidatos beneficiados mas com evidente interesse em seu resultado, contrataram empresas de marketing digital especializadas para distribuir conteúdo de campanha em escala industrial por meio do WhatsApp. A dissociação entre financiadores e candidatos não era acidental: ela foi deliberadamente construída para criar uma camada de negação plausível que dificultasse a configuração jurídica do abuso de poder econômico previsto no artigo 222-A da Lei nº 9.504/1997. Como demonstrou o inquérito que resultou no RQ-0060125/2022 do TSE, essa dissociação não impediu a responsabilização dos empresários identificados, mas criou obstáculos probatórios significativos que retardaram a resposta da Justiça Eleitoral para além do período em que ela poderia ter impacto sobre o resultado eleitoral (TSE, RQ-0060125/2022).

A mecânica operacional identificada pelas investigações revela um sistema de três camadas funcionalmente integradas. Na base, as agências de disparos obtinham listas massivas de números de telefone a partir de bancos de dados comerciais e de vazamentos de dados pessoais, configuravam chips de celular cadastrados em lote especificamente para disparos em escala e utilizavam infraestrutura técnica distribuída para evitar a detecção pelos sistemas de monitoramento do próprio WhatsApp. No meio, redes hierárquicas de distribuição coordenavam a disseminação do conteúdo a partir de líderes de opinião com alto índice de credibilidade em seus grupos de referência, como pastores, militares e empresários locais, que redistribuíam o material para seus contatos imediatos.

No topo, esses contatos imediatos o repassavam para grupos familiares, o ambiente de comunicação com maior credibilidade percebida e menor ceticismo crítico do eleitor

médio, completando o ciclo de amplificação com um custo marginal próximo de zero para os financiadores da operação.

Essa estrutura de três camadas é analiticamente relevante porque cada uma delas apresenta perfil de vulnerabilidade regulatória distinto. A camada técnica das agências de disparos é a mais vulnerável à intervenção da Anatel, que tem competência sobre o uso irregular de chips em lote, e da Polícia Federal, que pode investigar o acesso ilegal a bancos de dados de números de telefone. A camada intermediária dos líderes de opinião é a mais protegida pela liberdade de expressão, e sua responsabilização jurídica exige a demonstração de dolo específico ou de vínculo contratual com os financiadores, o que as investigações de 2018 conseguiram estabelecer apenas em parte. A camada dos grupos familiares, onde a amplificação orgânica ocorre, é praticamente imune a qualquer intervenção regulatória, precisamente porque seus participantes operam de boa-fé e se enquadram na categoria de misinformation definida por Wardle e Derakhshan (2017): acreditam genuinamente no que redistribuem.

O caso de 2018 revelou também uma vulnerabilidade específica do sistema eleitoral brasileiro que permanece parcialmente não endereçada: a Justiça Eleitoral opera em ciclos processuais incompatíveis com a velocidade das operações de desinformação. As investigações que produziram responsabilizações no RQ-0060125/2022 foram concluídas quatro anos após o pleito que buscavam regular. O dano eleitoral havia sido irreversível muito antes de qualquer sanção ser aplicada. Como argumenta Howard (2020), a eficácia regulatória contra operações de desinformação financiadas requer capacidade de intervenção em tempo real, não apenas responsabilização retrospectiva, e essa capacidade exige estruturas de inteligência eleitoral que a maioria dos sistemas jurídicos, incluindo o brasileiro, ainda está construindo (HOWARD, 2020, p. 142-145).

A dimensão financeira do caso é a que tem maior relevância para a estratégia regulatória de longo prazo. As estimativas disponíveis sobre o volume de recursos envolvidos nos disparos de 2018 variam amplamente, mas investigações jornalísticas e do MPE apontaram valores na casa de dezenas de milhões de reais em contratos com agências de disparos. Esses recursos não apareceram nas prestações de contas oficiais das campanhas beneficiadas, o que configura não apenas abuso de poder econômico mas também financiamento eleitoral irregular, cujo rastreamento exige cooperação entre a Justiça Eleitoral, a Receita Federal e o Banco Central que o ordenamento de 2018 não estava equipado para prover com a agilidade necessária. A Resolução TSE nº 23.732/2024 e as alterações processuais subsequentes melhoraram marginalmente essa capacidade, mas não resolveram o problema estrutural da velocidade: operações bem financiadas

continuam tendo capacidade de causar dano eleitoral irreversível antes que qualquer resposta institucional efetiva seja possível.

⚖ JURISPRUDÊNCIA TSE — AIJE nº 0601771-86.2018.6.00.0000 (reexame)

EMENTA:

Reexame da investigação sobre financiamento de disparos em massa pelo WhatsApp nas eleições presidenciais de 2018. O TSE investigou empresários que financiaram a contratação de agências de disparos sem declaração como despesa de campanha, configurando potencial abuso de poder econômico e uso indevido de meios de comunicação.

RESULTADO:

O TSE reconheceu a gravidade das práticas e a insuficiência das ferramentas regulatórias então disponíveis para responsabilização efetiva dos financiadores, o que impulsionou diretamente as alterações legislativas subsequentes (Lei nº 14.197/2021) e a edição da Resolução nº 23.732/2024. O caso demonstrou a necessidade de mecanismos de rastreamento financeiro específicos para o ambiente de comunicação privada criptografada.

⚡ ATENÇÃO PRÁTICA — Para o Profissional de Campanha

O uso do WhatsApp para comunicação eleitoral é legal e estrategicamente relevante. O que é proibido é o disparo em massa organizado por empresas contratadas, a criação de listas de números sem consentimento dos titulares e o financiamento não declarado dessas operações. A fronteira entre comunicação orgânica legítima e disparo em massa irregular é definida pelo art. 222-A da Lei nº 9.504/1997 e pela Resolução TSE nº 23.732/2024. Uma campanha que contrata empresa para disparos em massa no WhatsApp sem declarar essa contratação na prestação de contas expõe o candidato a processo de abuso de poder econômico com potencial de cassação de mandato.

■ RESUMO DO CAPÍTULO:

Este capítulo analisou a economia política da desinformação eleitoral, partindo de uma premissa que reorganiza toda a análise regulatória: desinformação em escala não é um fenômeno tecnológico que acontece nas plataformas, é um fenômeno econômico e político que é deliberadamente construído e financiado por atores com recursos identificáveis e interesse no resultado eleitoral. A revisão de literatura estabeleceu os três marcos teóricos centrais do campo. Benkler, Faris e Roberts demonstraram empiricamente, com corpus de 1,25 milhão de matérias das eleições americanas de 2016, que a assimetria de desinformação é produto de investimento deliberado em infraestrutura, não de dinâmicas culturais espontâneas. Howard sistematizou os padrões recorrentes das operações de desinformação global, identificando a dissociação deliberada entre financiadores, produtores e distribuidores como mecanismo central de evasão de responsabilidade. Araújo e Oliveira, pelo NetLab UFRJ, e o FGV DAPP estabeleceram as características estruturais específicas do ecossistema brasileiro, particularmente a centralidade do WhatsApp e o papel das redes religiosas e empresariais como canais de amplificação. A perspectiva ecológica aplicada ao ecossistema de desinformação revelou que intervenções cirúrgicas sobre componentes isolados produzem efeitos limitados porque o ecossistema é adaptativo: quando um canal é bloqueado, o conteúdo migra; quando uma conta é removida, outras assumem sua função. A estratégia regulatória mais eficaz não é a remoção de conteúdo, mas o aumento dos custos de financiamento e a redução dos retornos esperados das operações. O quadro de atores confirmou o paradoxo central: os

financiadores, o elo mais determinante da cadeia, são também os mais difíceis de alcançar regulatoriamente, porque operam com maior distância entre sua ação e o conteúdo que financiam e podem invocar proteções jurídicas legítimas para dificultar o rastreamento. A precisão terminológica foi trabalhada a partir da tipologia tripartite de Wardle e Derakhshan, que distingue misinformation, disinformation e malinformation pelos critérios de falsidade do conteúdo e intenção do emissor. Essa distinção tem consequências jurídicas diretas: a misinformation exige educação midiática, não sanção; a disinformation configura conduta dolosa passível de responsabilização eleitoral plena; a malinformation opera numa zona cinzenta que exige análise caso a caso. A Resolução TSE nº 23.732/2024 cobre adequadamente a disinformation mas deixa lacunas significativas nos outros dois tipos. O caso dos disparos em massa de 2018 foi analisado como estudo de caso estrutural, não como episódio isolado, revelando três vulnerabilidades sistêmicas que permanecem parcialmente não endereçadas: a incompatibilidade entre ciclos processuais da Justiça Eleitoral e velocidade das operações de desinformação, documentada pelo RQ-0060125/2022 concluído quatro anos após o pleito; a fragmentação de competências entre TSE, Anatel, Polícia Federal e Receita Federal que dificulta a resposta integrada; e a ausência de mecanismos de rastreamento financeiro ágeis para operações que transitam por Pix, espécie e contratos informais. O capítulo encerrou com a análise do papel singular do WhatsApp no ecossistema brasileiro, identificando suas três características estruturais que o tornam simultaneamente maximamente eficaz para a desinformação e minimamente responsivo aos instrumentos regulatórios disponíveis: criptografia de ponta a ponta, ausência de algoritmo de recomendação centralizado e efeito do mensageiro confiável. A conclusão regulatória central é que, para o WhatsApp, a prevenção é estruturalmente mais eficaz do que a resposta: quando o conteúdo já está circulando em grupos privados, nenhum instrumento disponível consegue alcançar os usuários já expostos.

QUESTÕES PARA REFLEXÃO E DEBATE

Questão 1

A pesquisa de Benkler, Faris e Roberts (2018), publicada em *Network Propaganda*, analisou 1,25 milhão de matérias das eleições presidenciais americanas de 2016 e concluiu que a assimetria de desinformação entre os campos políticos era produto de:

a) dinâmicas algorítmicas autônomas das plataformas digitais, sem intervenção humana organizada.
b) preferências culturais de segmentos específicos do eleitorado com menor acesso à educação formal.
c) investimentos deliberados em infraestrutura de mídia e desinformação por atores específicos com recursos identificáveis.
d) falhas nos sistemas de moderação das plataformas, que não foram projetados para ambiente eleitoral.
e) propagação espontânea de conteúdo emocional por usuários sem motivação política organizada.

Questão 2

Wardle e Derakhshan (2017) propõem uma tipologia tripartite da desordem informacional. Analise as situações abaixo e classifique cada uma corretamente:

I. Um eleitor compartilha, acreditando ser verdadeiro, um vídeo que alega que determinado candidato foi preso, quando o vídeo é falso.

II. Uma agência de disparos distribui deliberadamente um áudio sintético com a voz clonada de um candidato confessando irregularidades que nunca ocorreram.

III. Um grupo político vaza conversas privadas autênticas de um adversário, editadas seletivamente para distorcer o contexto e maximizar o dano político.

A classificação correta é:

a) I — disinformation; II — misinformation; III — malinformation.

b) I — misinformation; II — disinformation; III — malinformation.
c) I — malinformation; II — disinformation; III — misinformation.
d) I — misinformation; II — malinformation; III — disinformation.
e) I — disinformation; II — malinformation; III — misinformation.

Questão 3

Sobre o regime de responsabilidade por disparos em massa pelo WhatsApp nas eleições brasileiras, com base no artigo 222-A da Lei nº 9.504/1997 e na Resolução TSE nº 23.732/2024, assinale a alternativa correta:

a) O uso do WhatsApp para comunicação eleitoral é vedado em qualquer modalidade, por configurar uso indevido de meio de comunicação social.
b) Apenas o candidato diretamente beneficiado pelos disparos pode ser responsabilizado, desde que haja prova de seu conhecimento prévio.
c) O financiamento não declarado de agências de disparos por empresários, mesmo sem conhecimento formal do candidato, pode configurar abuso de poder econômico com potencial de cassação de mandato.
d) A responsabilização dos financiadores de disparos depende exclusivamente de notificação judicial prévia às plataformas, nos termos do artigo 19 do Marco Civil da Internet.
e) A criptografia de ponta a ponta do WhatsApp impede qualquer modalidade de responsabilização jurídica pelos conteúdos nele veiculados.

GABARITO: 1c; 2b; 3c

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Identifique em qual das três categorias cai cada ameaça antes de decidir como responder.

Misinformation, disinformation e malinformation exigem respostas completamente diferentes. Um eleitor que compartilhou de boa-fé um boato falso sobre seu candidato não é um adversário: é uma vítima da operação. Tratá-lo juridicamente como tal é um erro estratégico e jurídico. A disinformation fabricada exige representação eleitoral imediata e medidas cautelares. A malinformation, que usa conteúdo verdadeiro fora de contexto, é a mais difícil de combater juridicamente porque o conteúdo em si é real. Para cada uma delas, o instrumento é diferente: educação para a primeira, ação judicial para a segunda, resposta comunicacional de contexto para a terceira.

2. O financiador invisível é o maior risco jurídico da sua campanha e do adversário.

O RQ-0060125/2022 do TSE demonstrou que grupos de empresários que financiam disparos sem coordenação formal com o candidato ainda podem ser responsabilizados por abuso de poder econômico. Isso funciona nos dois sentidos: sua campanha pode ser atingida por apoiadores que agiram por conta própria, e você pode responsabilizar adversários pelo mesmo mecanismo. A due diligence sobre quem está fazendo campanha em seu nome nas redes privadas é tão importante quanto o controle sobre a comunicação oficial. Documente tudo: contratos, autorizações, limites de atuação de terceiros. O que não está documentado vira passivo eleitoral.

3. Disparo em massa no WhatsApp não é comunicação de campanha. É crime eleitoral.

O artigo 222-A da Lei 9.504/1997 proíbe expressamente o envio em massa de mensagens por meios eletrônicos para fins eleitorais sem consentimento dos destinatários. A fronteira entre comunicação orgânica legítima, enviar uma mensagem para seus contatos pedindo votos, e disparo irregular, contratar uma empresa para enviar mensagens idênticas a listas obtidas sem consentimento, é clara na lei e cada vez mais rigorosa na jurisprudência. Uma campanha que contrata agência de disparos sem declarar na

prestação de contas está acumulando risco de cassação de registro, cassação de mandato e responsabilização penal dos envolvidos.

4. Quando o conteúdo falso chegou ao WhatsApp, a batalha de remoção já foi perdida.

Ordens de remoção do TSE alcançam plataformas abertas como Facebook, Instagram e YouTube. Não alcançam grupos privados de WhatsApp onde o conteúdo já está circulando. Quando um deepfake ou boato eleitoral entra na rede de grupos familiares do WhatsApp, a única estratégia eficaz é a resposta comunicacional direta: conteúdo de desmentido simples, emocional e compartilhável que use os mesmos canais de distribuição. Desmentidos técnicos publicados no site da campanha não chegam onde o dano foi causado. A estratégia de defesa contra WhatsApp precisa estar pronta antes do ataque, não depois.

5. Rastreie de onde veio o conteúdo, não apenas o conteúdo em si.

A perspectiva ecológica desenvolvida neste capítulo tem uma aplicação investigativa direta: quando sua campanha identifica desinformação circulando contra você, a pergunta mais importante não é 'o que foi dito?' mas 'quem financiou isso?'. Fábricas de conteúdo são vulneráveis; financiadores são mais difíceis de identificar mas muito mais difíceis de substituir. Uma denúncia ao TSE que identifica apenas o conteúdo falso tem menor impacto do que uma representação que identifica o financiador da operação. Invista em inteligência eleitoral que rastreie origens, não apenas manifesta-ções.

6. A literacia midiática do eleitor é a sua melhor defesa de longo prazo.

Misinformation, o tipo mais comum de desinformação, não pode ser combatida por sanção jurídica: quem compartilha um boato acreditando nele não cometeu crime eleitoral. A única resposta eficaz para a camada de amplificação orgânica, os grupos familiares onde o conteúdo ganha escala, é a educação midiática que aumenta o ceticismo crítico do eleitor antes que o conteúdo falso chegue. Campanhas que investem em conteúdo educativo sobre como identificar desinformação, especialmente direcionado a eleitores mais velhos com acesso recente à internet, estão construindo imunidade eleitoral que nenhuma estratégia reativa consegue reproduzir.

7. O ecossistema de desinformação se adapta. Sua estratégia de defesa também precisa.

As operações documentadas em 2018 usavam chips cadastrados em lote e listas de telefones compradas ilegalmente. Em 2022, as operações já haviam migrado parte da infraestrutura para plataformas mais difíceis de rastrear e incorporado IA generativa para personalização em escala. Em 2026, o padrão operacional será diferente novamente. Uma estratégia de defesa que foi construída para o adversário de 2022 está desatualizada para o adversário de 2026. Monitore o ecossistema de desinformação continuamente, não apenas nos picos de campanha, e atualize sua estratégia a cada ciclo.

CAPÍTULO 10

O Brasil nas Eleições de 2018, 2022 e 2024

Casos Concretos, Padrões Emergentes e Lições para 2026

"Cada ciclo eleitoral nos ensina algo novo sobre como a tecnologia pode ser usada para manipular ou para proteger a democracia. O problema é que os manipuladores aprendem mais rápido do que os reguladores."

Adaptado do Relatório CIEDDE/TSE, 2024

Eleições Brasileiras na Era Digital

A literatura sobre as eleições brasileiras recentes no contexto da inteligência artificial e da desinformação digital tem crescido aceleradamente, especialmente a partir de 2018, e sua expansão reflete não apenas o interesse acadêmico pelo fenômeno, mas a urgência percebida por pesquisadores, juristas e autoridades eleitorais de compreender transformações que estavam ocorrendo mais rapidamente do que qualquer marco regulatório conseguia acompanhar.

Três linhas de pesquisa estruturam o campo. A primeira é a análise quantitativa do ecossistema informacional digital nas eleições brasileiras, representada pelos trabalhos do FGV DAPP, do NetLab da UFRJ e do grupo de pesquisa em comunicação política da UnB. Essas pesquisas estabeleceram as bases empíricas para a compreensão do fenômeno: mapearam os padrões de disseminação de desinformação nos ciclos eleitorais de 2018, 2020 e 2022, documentaram a centralidade do WhatsApp como canal dominante de circulação de informação política informal e identificaram as estruturas de redes coordenadas de amplificação que distinguem operações organizadas de compartilhamento espontâneo. Sem essa base empírica, a análise jurídica e a resposta regulatória estariam construindo sobre premissas não verificadas.

A segunda linha é a análise jurídica das respostas institucionais do TSE ao fenômeno, com contribuições de Gomes (2023), que examinou os instrumentos de combate à desinformação eleitoral desenvolvidos pela Justiça Eleitoral brasileira; de Cândido (2022), que analisou os limites constitucionais da regulação do discurso político em ambiente digital; e de Salgado (2015), cujo trabalho anterior sobre o direito eleitoral brasileiro

forneceu o enquadramento doutrinário sobre o qual as respostas regulatórias mais recentes foram construídas. Essa linha de pesquisa é especialmente relevante para a prática jurídica eleitoral porque é a que mais diretamente informa as decisões do TSE e dos TREs em casos concretos.

A terceira linha é a análise comparada, que situa a experiência brasileira no contexto das democracias que enfrentaram desafios similares em ciclos eleitorais recentes. Os casos dos Estados Unidos em 2016, da França em 2017, da Eslováquia em 2023 e de Bangladesh em 2024, analisados nos capítulos anteriores, fornecem o quadro de referência que permite identificar o que é particular ao contexto brasileiro e o que reflete padrões recorrentes independentes de contexto nacional.

Uma lacuna relevante que persiste na literatura, e que esta obra procura endereçar especificamente neste capítulo, é a análise integrada que combine de forma sistemática as dimensões técnica, jurídica e política das eleições brasileiras na era da inteligência artificial. A tendência de especialização das três linhas de pesquisa produziu um corpo de conhecimento robusto em cada dimensão individualmente, mas relativamente escasso na articulação entre elas: análises técnicas que não examinam os instrumentos jurídicos disponíveis, análises jurídicas que não incorporam a evidência empírica sobre os fenômenos que regulam, e análises políticas que não consideram as limitações técnicas das respostas institucionais. A análise integrada dos três ciclos eleitorais mais recentes que se segue procura contribuir para o preenchimento dessa lacuna.

Eleições de 2018: a Irrupção do WhatsApp e a Desinformação Industrial

As eleições presidenciais brasileiras de 2018 representam um ponto de inflexão na história da comunicação política nacional cujas consequências continuam sendo sentidas nos ciclos eleitorais seguintes. Três fenômenos convergiram nesse ciclo de forma simultânea e mutuamente amplificadora, criando um ambiente informacional eleitoral sem precedentes históricos no Brasil: a massificação do WhatsApp como canal primário de informação política para parcelas significativas da população, a emergência de esquemas organizados de disparo em massa de desinformação com financiamento externo às campanhas, e a ausência de um marco regulatório adequado para identificar, endereçar e sancionar esses fenômenos em tempo real.

A massificação do WhatsApp como canal político não foi planejada por nenhum ator específico: foi o resultado da convergência entre a queda dos preços de smartphones, a expansão da cobertura de dados móveis para regiões anteriormente sem acesso e a

adoção do aplicativo como substituto funcional do SMS por populações que chegavam ao ambiente digital pela primeira vez. Para esses novos usuários digitais, frequentemente adultos maduros com menor familiaridade com o ecossistema informacional online, o WhatsApp não era apenas um aplicativo de mensagens: era a internet. O conteúdo que chegava pelo WhatsApp vinha com o endosso implícito de pessoas de confiança e sem os marcadores visuais que usuários mais experientes associariam a fontes não verificadas (ARAÚJO; OLIVEIRA, 2021, p. 45-48).

Esse perfil de usuário criou as condições ideais para o segundo fenômeno: a organização de esquemas de disparo em massa que exploraram deliberadamente a confiança estrutural do ambiente de comunicação privada. O ecossistema de desinformação de 2018 operava em rede hierárquica identificada pelas investigações posteriores do TSE, do Ministério Público Eleitoral e de pesquisadores do FGV DAPP. No topo da hierarquia, grupos de líderes de opinião com alta credibilidade em seus nichos específicos, pastores evangélicos com congregações de milhares de fiéis, militares da reserva com redes de contatos institucionais, empresários com grupos de fornecedores e funcionários, recebiam conteúdo produzido pelas fábricas de desinformação e o redistribuíam para grupos intermediários de influenciadores locais. Esses repassavam para grupos familiares, o canal de comunicação com maior índice de credibilidade percebida e menor ceticismo crítico do eleitor médio brasileiro. O volume estimado de mensagens enviadas por esses esquemas durante o período que abrange pré-campanha e campanha oficial é da ordem de dezenas de milhões, segundo estimativas do MPE baseadas em dados de operadoras e de análise forense de dispositivos apreendidos (MPE, 2019).

O terceiro fenômeno, a ausência de marco regulatório adequado, não foi uma falha de antecipação: foi o reflexo da velocidade com que a tecnologia transformou o ambiente eleitoral antes que qualquer sistema regulatório pudesse responder. Em 2018, o artigo 222-A da Lei nº 9.504/1997, que proíbe disparos em massa sem consentimento dos destinatários, existia formalmente, mas o TSE não tinha instrumentos de monitoramento, de prova e de resposta em tempo real para aplicá-lo com eficácia a operações conduzidas em ambientes privados criptografados. Como documentou Gomes (2023), a Justiça Eleitoral brasileira chegou às eleições de 2018 com ferramentas regulatórias projetadas para o ambiente da comunicação política de massa tradicional, televisão, rádio, outdoor, material impresso, e teve de improvisar respostas para um fenômeno que essas ferramentas não foram projetadas para alcançar (GOMES, 2023, p. 112-115).

A análise retrospectiva das eleições de 2018 é analiticamente relevante não apenas como registro histórico, mas como diagnóstico estrutural: ela revela que a combinação de

novo canal de comunicação de massa, ausência de regulação efetiva e financiamento organizado de desinformação produz uma janela de vulnerabilidade eleitoral que os atores com recursos e motivação para explorá-la reconhecerão e utilizarão. Compreender os mecanismos que operaram em 2018 é condição para antecipar suas variantes nos ciclos eleitorais de 2026 e seguintes, quando novas tecnologias criarão novas janelas de vulnerabilidade que o sistema regulatório atual, mesmo aperfeiçoado, ainda não foi projetado para fechar.

O quadro a seguir sistematiza as cinco dimensões centrais do ecossistema de desinformação nas eleições de 2018, permitindo uma leitura integrada que o relato narrativo, por sua natureza sequencial, tende a fragmentar. A coluna de resposta institucional é particularmente reveladora quando lida em conjunto com a coluna de impacto: em todas as cinco dimensões, a resposta chegou depois, frequentemente anos depois, do dano já causado.

O acordo com o WhatsApp veio após o ciclo eleitoral. Os limites ao encaminhamento foram implementados em 2019 e 2020. A Lei nº 14.197/2021 foi aprovada três anos após as eleições que a tornaram necessária. A Resolução TSE nº 23.732/2024 chegou seis anos depois. Esse padrão de defasagem não é acidental e não é específico ao Brasil: é a consequência estrutural de sistemas regulatórios que respondem a crises já ocorridas em vez de antecipar vulnerabilidades identificáveis. A questão que o quadro implicitamente coloca, e que este capítulo desenvolve nas seções seguintes, é se o sistema aprendeu o suficiente para que o ciclo de 2026 seja diferente.

Tabela 35. Eleições Brasileiras 2018: Características do Ecossistema de Desinformação

Dimensão	Características	Impacto	Resposta Institucional
Canal principal	WhatsApp: comunicação privada, criptografada, peer-to-peer.	Alcance de 120 milhões de usuários sem possibilidade de monitoramento em tempo real.	Acordo TSE-WhatsApp para remoção de conteúdo irregular; implementação posterior de limites ao encaminhamento.
Modelo de produção	Fábricas de fake news contratadas; conteúdo adaptado para diferentes perfis de público.	Produção industrial de desinformação em escala impossível sem automação.	Investigação de 14 agências de disparos pelo TSE; investigação inconclusiva por dificuldades de acesso a dados.
Modelo de financiamento	Empresários financiaram disparos em massa sem	Violação do princípio da igualdade de	AIJE instaurada; resultados limitados pela dificuldade

	declaração como despesa de campanha.	condições na disputa eleitoral.	de rastreamento de pagamentos.
Conteúdo predominante	Desinformação sobre candidatos; narrativas falsas sobre planos de governo; material de supressão de voto.	Distorção do debate eleitoral; possível influência sobre eleitores indecisos.	Parcerias com agências de fact-checking; remoção posterior de conteúdo identificado.
Marco regulatório	Ausente para o contexto específico; Lei das Eleições não contemplava disparos em massa digitais.	Impunidade relativa dos responsáveis; impossibilidade de resposta regulatória em tempo real.	Aprovação da Lei nº 14.197/2021 e edição da Resolução TSE nº 23.732/2024 como respostas normativas.

Fonte: Elaborado pelo autor com base em TSE (2019), FGV DAPP (2018)

Eleições de 2024: o Primeiro Teste da Resolução TSE nº 23.732/2024

As eleições municipais de outubro de 2024 foram o primeiro teste real da Resolução TSE nº 23.732/2024, o marco normativo específico sobre inteligência artificial eleitoral editado em março daquele ano. A distância de apenas sete meses entre a edição da norma e o início do período eleitoral foi, ela própria, um fator de risco: campanhas, agências de marketing eleitoral e assessores jurídicos tiveram tempo limitado para incorporar as novas obrigações em suas práticas, e o TSE teve tempo igualmente limitado para construir a capacidade operacional de fiscalização que a norma pressupunha. O ciclo eleitoral de 2024 configurou, nessas circunstâncias, o laboratório mais importante disponível para avaliar a efetividade da regulação brasileira de IA eleitoral antes das eleições gerais de 2026, quando o escrutínio será incomparavelmente maior e as consequências de falhas regulatórias incomparavelmente mais graves.

Os primeiros casos de aplicação da Resolução revelaram o espectro de condutas que o novo marco normativo foi chamado a endereçar ainda em sua fase inaugural. Deepfakes de candidatos a prefeito em municípios de médio porte, produzidos com tecnologia de face swap de qualidade suficiente para enganar eleitores sem treinamento específico de detecção, circularam por grupos de WhatsApp nos dias imediatamente anteriores ao primeiro turno, seguindo o padrão de timing estratégico identificado nos casos internacionais analisados nos capítulos anteriores. Chatbots sem identificação como sistemas automatizados interagiram com eleitores em plataformas de mensageria simulando conversas com representantes das campanhas, violando a obrigação de transparência estabelecida pelo artigo 15 da Resolução. Campanhas de disparo em massa

utilizaram conteúdo gerado por modelos de linguagem sem a rotulagem obrigatória, produzindo em escala textos personalizados por perfil de eleitor que seriam impossíveis de criar manualmente nas quantidades distribuídas. Em todos esses casos, o TSE e os TREs aplicaram as disposições da Resolução, determinando remoções, aplicando multas e encaminhando investigações ao Ministério Público Eleitoral.

A avaliação provisória da efetividade da Resolução nesse primeiro ciclo de aplicação revela um quadro misto que a análise honesta não pode simplificar nem no sentido do sucesso nem no do fracasso.

Do lado positivo, a existência de uma norma específica para IA eleitoral conferiu à Justiça Eleitoral instrumentos mais precisos e juridicamente mais seguros para intervir em casos que, nos ciclos anteriores, exigiam enquadramentos por analogia na legislação geral com resultados imprevisíveis. A velocidade das liminares de remoção melhorou mensurávelmente em relação a 2022, particularmente nos casos em que a análise forense pôde ser realizada rapidamente e o nexo com a Resolução era claro. O regime de responsabilidade solidária estabelecido pelo artigo 27, que inclui as plataformas entre os responsáveis pelo conteúdo mantido após notificação, criou incentivos que aceleraram a cooperação das plataformas com as ordens de remoção em comparação com os ciclos anteriores.

Do lado negativo, persistiram limitações estruturais que nenhuma norma bem escrita pode superar sozinha. A capacidade de detecção de violações pela própria Justiça Eleitoral ainda depende, em grande medida, de denúncias formalizadas por candidatos, partidos e cidadãos, o que significa que a efetividade da fiscalização é inversamente proporcional à sofisticação das violações: operações bem conduzidas que não geram denúncias formais continuam invisíveis ao sistema. O volume de conteúdo gerado por IA em campanhas municipais de 2024, amplificado pela democratização do acesso a ferramentas de geração de texto, imagem e áudio, tornou impraticável a verificação sistemática de conformidade com a obrigação de rotulagem: não há capacidade operacional no TSE para verificar o cumprimento de uma obrigação que se aplica a potencialmente milhões de peças de conteúdo produzidas por centenas de campanhas simultaneamente.

Essa limitação não é uma crítica à Resolução nem ao TSE: é o reconhecimento de um problema estrutural que nenhum sistema de fiscalização manual consegue resolver em ambiente de produção de conteúdo em escala industrial. A resposta de longo prazo para esse problema não está na expansão da capacidade humana de fiscalização, que nunca conseguirá acompanhar o volume, mas no desenvolvimento de sistemas automatizados de detecção de violações que operem na mesma escala em que as violações ocorrem. Esse

é o horizonte técnico-regulatório que o TSE precisará desenvolver antes de 2026, e é precisamente o tipo de capacidade institucional que a literatura sobre regulação de IA eleitoral, particularmente Benkler, Faris e Roberts (2018) e Howard (2020), identifica como condição necessária para que a regulação deixe de chegar sempre um passo atrás da tecnologia que busca regular.

A comparação sistemática dos três ciclos eleitorais, apresentada no quadro a seguir, torna visível um padrão que a análise individual de cada ciclo tende a obscurecer. Em todas as cinco dimensões analisadas, há progresso mensurável de 2018 a 2024: a tecnologia de desinformação ficou mais sofisticada, mas a capacidade regulatória também avançou; o marco normativo evoluiu do vazio para uma norma específica; a resposta institucional migrou do improviso para instrumentos consolidados. O que o quadro também revela, porém, é que a lacuna entre a tecnologia disponível para produzir desinformação e a capacidade institucional de detectá-la e sancioná-la não diminuiu proporcionalmente ao avanço regulatório. Em 2018, a lacuna era de marco normativo. Em 2022, era de definição jurídica precisa. Em 2024, é de detecção sistemática em escala. A natureza da lacuna mudou em cada ciclo, mas sua existência se manteve constante. Isso não é pessimismo: é o diagnóstico que orienta a pergunta mais relevante para 2026, que não é se haverá lacunas, mas de qual natureza serão e se o sistema estará mais bem preparado para identificá-las antes que causem dano irreversível.

Tabela 36. Comparação dos Três Ciclos Eleitorais: 2018, 2022 e 2024

Dimensão	2018	2022	2024
Principal vetor de desinformação	WhatsApp: disparos em massa de conteúdo textual e imagens.	Redes sociais públicas: vídeos manipulados e narrativas de fraude eleitoral.	Múltiplos canais: deepfakes, chatbots não declarados, conteúdo gerado por IA.
Tecnologia de IA utilizada	Automação básica de disparos; segmentação por dados demográficos.	Primeiros deepfakes de vídeo; edição seletiva de conteúdo.	IA generativa avançada (LLMs, síntese de voz, face swap); chatbots com LLMs.
Marco regulatório disponível	Lei das Eleições sem disposições específicas para IA ou disparos em massa digital.	Sem regulação específica sobre deepfakes; aplicação de normas gerais por analogia.	Resolução TSE nº 23.732/2024: primeiro marco normativo específico.

Capacidade de resposta do TSE	Limitada: sem instrumentos específicos; acordos voluntários com plataformas.	Intermediária: jurisprudência em construção; CIEDDE criado em 2022.	Melhorada: instrumentos específicos; CIEDDE operacional; parcerias consolidadas.
Principal lacuna identificada	Rastreamento de financiadores; fiscalização de comunicação privada.	Definição jurídica precisa de deepfake; mecanismos de detecção técnica.	Detecção sistemática de conteúdo IA sem rotulagem; protocolo ANPD-TSE.

Fonte: Elaborado pelo autor com base em TSE (2019, 2022, 2024), FGV DAPP)

Lições para 2026: o que os Três Ciclos Ensinam

A análise comparada dos três ciclos eleitorais brasileiros recentes permite identificar padrões e tendências com grau de regularidade suficiente para informar a preparação dos diferentes atores para as eleições gerais de 2026, o primeiro ciclo presidencial brasileiro sob plena vigência da Resolução TSE nº 23.732/2024 e em contexto de disponibilidade generalizada de ferramentas de inteligência artificial generativa de alta qualidade a custo marginal próximo de zero. O que os três ciclos ensinam não é um conjunto de recomendações técnicas pontuais: é uma lógica de adaptação que os atores que buscam preservar a integridade eleitoral precisam internalizar como princípio operacional permanente.

A primeira lição é sobre a velocidade de adaptação dos atores maliciosos em relação à velocidade de resposta regulatória. Em cada ciclo eleitoral, as práticas de desinformação evoluíram para explorar precisamente os canais e as tecnologias que não estavam adequadamente cobertas pela regulação do ciclo anterior. Em 2018, o WhatsApp foi o canal não regulado: a legislação eleitoral foi projetada para meios de comunicação de massa identificáveis, não para comunicação privada criptografada em escala. Em 2022, os deepfakes foram a tecnologia não regulada: os instrumentos disponíveis para enquadrá-los dependiam de analogias com tipos eleitorais preexistentes que não capturavam adequadamente suas especificidades.

Em 2024, a combinação de chatbots com modelos de linguagem e de IA generativa acessível foram as fronteiras que a Resolução TSE nº 23.732/2024 buscou cobrir, com os resultados mistos documentados na seção anterior. Em 2026, por extensão do padrão, a atenção regulatória deve se concentrar nas tecnologias emergentes que ainda não têm cobertura normativa adequada: agentes de IA autônomos capazes de conduzir campanhas

de influência sem intervenção humana contínua, sistemas de vídeo sintético em tempo real que eliminam a necessidade de pós-produção, e ferramentas de personalização hipersegmentada que operam abaixo do limiar de detecção dos sistemas de monitoramento atuais. Como observa Howard (2020), a assimetria entre a velocidade de inovação dos atores maliciosos e a velocidade de resposta dos sistemas regulatórios é estrutural e não será eliminada: o objetivo realista é reduzi-la a um nível em que o dano eleitoral seja limitável, não eliminá-la (HOWARD, 2020, p. 148-151).

A segunda lição é sobre a irreductibilidade da literacia midiática como defesa estrutural que nenhuma estratégia regulatória pode substituir. Regulação e fiscalização são necessárias, mas insuficientes por uma razão que os três ciclos eleitorais confirmaram de forma consistente: a camada de amplificação orgânica do ecossistema de desinformação é composta majoritariamente por cidadãos que redistribuem conteúdo de boa-fé, acreditando genuinamente no que compartilham. Essa camada não pode ser alcançada por sanção jurídica, porque seus participantes não cometem ato ilícito. Ela só pode ser reduzida por investimento em capacidade crítica que aumente o ceticismo do eleitor em relação ao conteúdo que recebe antes de redistribuí-lo.

As pesquisas de Pennycook e Rand (2019) sobre intervenções de literacia midiática sugerem que mesmo exposições breves a conteúdo educativo sobre técnicas de desinformação produzem reduções mensuráveis na intenção de compartilhar conteúdo não verificado, particularmente em populações com menor experiência prévia no ambiente digital (PENNYCOOK; RAND, 2019, p. 2522-2524). Os investimentos em educação midiática que o Brasil precisa realizar para 2026 não são despesas eleitorais de curto prazo: são investimentos em capital cívico que produzirão dividendos democráticos por décadas, independentemente de quem vencer as eleições em que forem aplicados.

A terceira lição é sobre a necessidade de integração entre as instituições com competências complementares no combate à desinformação eleitoral. Os três ciclos revelaram repetidamente que a fragmentação de competências entre TSE, Anatel, Polícia Federal, Receita Federal, ANPD e Ministério Público Eleitoral cria janelas de irresponsabilidade que operações sofisticadas exploram deliberadamente: cada instituição tem jurisdição sobre uma parte da cadeia, e nenhuma tem instrumentos para alcançar a cadeia inteira. O CIEDDE, criado em 2022, foi um passo relevante na direção da integração, mas sua efetividade depende de protocolos de cooperação e de compartilhamento de inteligência que ainda estão em construção. Para 2026, a pergunta relevante não é se as instituições têm competência individual suficiente, mas se têm protocolos de ação conjunta suficientemente ágeis para operar na velocidade que o ambiente exige.

A quarta lição, talvez a mais incômoda politicamente, é que a efetividade da regulação de desinformação eleitoral depende, em última instância, da vontade dos atores políticos de submetê-la a ela. Marcos regulatórios são instrumentos: sua eficácia depende de quem os opera e de quais incentivos governam esse operador. Um sistema eleitoral em que os principais beneficiários potenciais da desinformação são também os que controlam os instrumentos de fiscalização enfrenta um problema de governança que nenhuma norma bem escrita resolve por si mesma. A independência institucional da Justiça Eleitoral brasileira é uma proteção relevante contra esse risco, mas não é uma garantia absoluta.

Como argumentam Levitsky e Ziblatt (2018), democracias se enfraquecem não apenas por golpes abruptos, mas pela erosão incremental das normas e instituições que garantem a competição eleitoral justa, e essa erosão frequentemente começa com a instrumentalização seletiva dos mecanismos de fiscalização (LEVITSKY; ZIBLATT, 2018, p. 97-100). A robustez do sistema eleitoral brasileiro para 2026 depende tanto da qualidade de suas normas quanto da qualidade das instituições e dos atores que as operarão.

■ RESUMO DO CAPÍTULO:

Este capítulo analisou os três ciclos eleitorais brasileiros mais recentes como estudos de caso integrados, combinando as dimensões técnica, jurídica e política que a tendência de especialização da literatura tende a tratar separadamente. O argumento central que atravessa toda a análise é o de que cada ciclo eleitoral cria uma janela de vulnerabilidade específica determinada pela combinação de tecnologia disponível, capacidade regulatória existente e incentivos dos atores com recursos para explorar a assimetria entre elas. A revisão da literatura estabeleceu três linhas de pesquisa que estruturam o campo: a análise quantitativa do ecossistema informacional digital, representada pelo FGV DAPP, NetLab UFRJ e UnB; a análise jurídica das respostas institucionais do TSE, com Gomes, Cândido e Salgado; e a análise comparada, que situa a experiência brasileira no contexto internacional. A contribuição específica deste capítulo é precisamente a articulação entre essas três linhas, que individualmente produzem conhecimento robusto mas que raramente dialogam de forma sistemática. A análise das eleições de 2018 revelou três fenômenos convergentes: a massificação do WhatsApp como canal político para populações que chegavam ao ambiente digital pela primeira vez, com o perfil de confiança estrutural que isso criou; a organização de esquemas hierárquicos de disparo em massa que exploraram deliberadamente esse ambiente, com financiamento externo às campanhas e dissociação deliberada entre financiadores e beneficiários; e a ausência de instrumentos regulatórios adequados para operar em tempo real num ambiente de comunicação privada criptografada. A lição mais durável de 2018 não é descritiva: é estrutural, a combinação de novo canal, ausência regulatória e financiamento organizado produz janela de vulnerabilidade que atores motivados reconhecerão e explorarão. O primeiro teste da Resolução TSE nº 23.732/2024 nas eleições municipais de 2024 revelou um quadro misto que a análise honesta não pode simplificar. Do lado positivo, a norma específica conferiu instrumentos mais precisos, acelerou as liminares de remoção e criou incentivos que melhoraram a cooperação das plataformas. Do lado negativo, a detecção de violações continua dependendo majoritariamente de denúncias formais, e o volume de conteúdo gerado por IA tornou impraticável a verificação sistemática de conformidade com a obrigação de rotulagem. A resposta estrutural para esse problema não está na expansão da capacidade humana de fiscalização, mas no desenvolvimento de sistemas automatizados de detecção que operem na mesma escala das violações. A comparação dos três ciclos evidenciou um padrão persistente: a lacuna entre tecnologia de desinformação e capacidade regulatória não diminuiu proporcionalmente ao avanço normativo. Sua

natureza mudou, de lacuna de marco normativo em 2018, para lacuna de definição jurídica precisa em 2022, para lacuna de detecção sistemática em escala em 2024, mas sua existência se manteve constante. Isso não é pessimismo: é o diagnóstico que orienta a pergunta mais relevante para 2026. As quatro lições para 2026 sintetizaram o argumento do capítulo. A primeira é que atores maliciosos adaptam-se mais rapidamente do que sistemas regulatórios, e que para 2026 a atenção deve se concentrar em agentes autônomos de IA, vídeo sintético em tempo real e personalização hipersegmentada. A segunda é que literacia midiática é a única defesa que alcança a camada de amplificação orgânica, que não pode ser atingida por sanção jurídica. A terceira é que a fragmentação de competências entre TSE, Anatel, PF, Receita Federal e ANPD cria janelas de irresponsabilidade que operações sofisticadas exploram deliberadamente. A quarta, a mais incômoda, é que a efetividade da regulação depende, em última instância, da vontade dos atores políticos de submetê-la, e que democracias se enfraquecem não apenas por golpes abruptos, mas pela erosão incremental das normas que garantem a competição eleitoral justa.

QUESTÕES PARA REFLEXÃO E DEBATE

1. O capítulo argumenta que a combinação de novo canal de comunicação, ausência regulatória e financiamento organizado produz janela de vulnerabilidade eleitoral que atores motivados explorarão. Identificando as tecnologias emergentes mencionadas para 2026, como agentes autônomos de IA e vídeo sintético em tempo real, quais medidas preventivas específicas o TSE poderia adotar antes das eleições gerais para reduzir essa janela, levando em conta as limitações de competência, orçamento e tempo disponível?

2. A segunda lição para 2026 sustenta que literacia midiática é a defesa mais robusta contra a camada de amplificação orgânica, que não pode ser alcançada por sanção jurídica. Como seria possível implementar programas de literacia midiática com escala e velocidade suficientes para produzir impacto mensurável nas eleições de 2026, considerando que o Brasil tem 120 milhões de usuários do WhatsApp e que intervenções de educação midiática tradicionalmente levam anos para produzir efeitos populacionais?

3. A comparação dos três ciclos eleitorais revelou que a lacuna entre tecnologia de desinformação e capacidade regulatória mudou de natureza mas não diminuiu. Esse padrão sugere que a regulação está estruturalmente condenada a chegar sempre depois, ou existem modelos regulatórios prospectivos, baseados em obrigações de avaliação de risco prévia ao lançamento de novas ferramentas de IA, que poderiam inverter essa lógica?

4. A quarta lição do capítulo é que a efetividade da regulação depende, em última instância, da vontade dos atores políticos de submetê-la. Essa afirmação tem implicações para o design institucional da Justiça Eleitoral: que características estruturais, como mandatos fixos, composição plural, autonomia orçamentária e mecanismos de prestação de contas, são mais relevantes para preservar a independência operacional do TSE num ciclo eleitoral em que a pressão política sobre as instituições eleitorais tende a ser máxima?

5. O capítulo identifica que a fragmentação de competências entre TSE, Anatel, Polícia Federal, Receita Federal e ANPD cria janelas de irresponsabilidade exploradas por operações sofisticadas. Que modelo de governança integrada, inspirado em experiências comparadas como o *Electoral Commission britânico* ou o modelo escandinavo de agência única de integridade eleitoral, seria mais compatível com o sistema federativo e o modelo de Justiça Eleitoral brasileiros?

6. A análise das eleições de 2024 concluiu que a resposta estrutural para o problema de detecção em escala não está na expansão da capacidade humana, mas no desenvolvimento de sistemas automatizados de detecção de violações. Que riscos específicos, como viés algorítmico, erros sistemáticos e opacidade decisória, esse modelo de fiscalização automatizada introduz, e como o direito eleitoral deveria regulamentá-los para evitar que o remédio seja pior do que a doença?

7. O caso das eleições de 2018 mostrou que a dissociação deliberada entre financiadores e candidatos, construída para criar negação plausível, foi eficaz em atrasar a responsabilização por quatro anos. Que

reformas específicas na legislação eleitoral, nos procedimentos do TSE e nos mecanismos de cooperação financeira com Receita Federal e Banco Central seriam necessárias para reduzir esse prazo para um horizonte compatível com a utilidade eleitoral da sanção?

8. Levitsky e Ziblatt argumentam que democracias se enfraquecem pela erosão incremental de normas e instituições antes de sucumbirem a rupturas abruptas. Aplicando essa análise ao contexto eleitoral digital brasileiro, quais indicadores específicos, como volume de desinformação não sancionada, tempo médio de resposta a denúncias, taxa de cumprimento de ordens de remoção, seriam os mais úteis para monitorar continuamente a saúde democrática do sistema eleitoral entre ciclos?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Mapeie a janela de vulnerabilidade do seu ciclo. Ela sempre existirá; a questão é de qual natureza.

Os três ciclos eleitorais analisados neste capítulo mostram que a janela de vulnerabilidade nunca fecha: ela muda de forma. Em 2018 era o WhatsApp não regulado. Em 2022 eram os deepfakes sem tipificação precisa. Em 2024 eram os chatbots sem identificação obrigatória. Em 2026 serão os agentes autônomos de IA e o vídeo sintético em tempo real. Para o profissional de campanha e para o advogado eleitoral, a pergunta estratégica não é 'estamos protegidos pela regulação vigente?', mas 'qual é a fronteira não regulada que nossos adversários têm incentivo para explorar?' Identificar essa fronteira antes da campanha é mais valioso do que reagir a violações depois que o dano já ocorreu.

2. A responsabilização por disparos chega quatro anos depois. O dano eleitoral é imediato.

O RQ-0060125/2022 foi concluído quatro anos após as eleições de 2018. Isso não é falha específica do TSE: é a consequência estrutural de investigar financiamento fragmentado em microtransações, acessar dados de comunicações privadas criptografadas e construir provas em cadeias de custódia que resistam ao contraditório. Para o assessor jurídico de campanha, isso significa que a ameaça de responsabilização retrospectiva não é desincentivo suficiente para quem planeja uma operação de desinformação que vai decidir uma eleição. A estratégia defensiva eficaz é a prevenção e a resposta comunicacional em tempo real, não a expectativa de que o adversário seja sancionado antes do pleito.

3. Em 2026, a Resolução TSE 23.732/2024 estará em plena vigência. Prepare sua campanha agora.

As eleições municipais de 2024 foram o primeiro teste da Resolução. As eleições gerais de 2026 serão o segundo, com escrutínio incomparavelmente maior, adversários que já estudaram as lacunas do primeiro ciclo e uma Justiça Eleitoral que aprendeu com os casos de 2024. Campanhas que chegarem a 2026 sem protocolos internos de conformidade com a Resolução, sem contratos adequados com fornecedores de tecnologia eleitoral e sem advogados eleitorais que conheçam a norma em detalhes estarão competindo com uma desvantagem estrutural. O tempo de preparação é agora, não na véspera do período de campanha.

4. Chatbot de campanha sem identificação é violação da Resolução. E é facilmente detectável.

O artigo 15 da Resolução TSE nº 23.732/2024 exige identificação obrigatória de sistemas automatizados que interajam com eleitores. Um chatbot que simula um atendente humano da campanha sem se identificar como sistema de IA viola a norma e está sujeito a multa e remoção. Diferente dos deepfakes que circulam no WhatsApp, chatbots em plataformas abertas são facilmente identificáveis por adversários, jornalistas e pelo próprio TSE. A tentação de usar chatbots não declarados para escalar o

atendimento eleitoral é compreensível do ponto de vista operacional, mas o risco jurídico e reputacional é desproporcional ao ganho. Use chatbots com identificação clara e rotulagem conforme a Resolução.

5. A literacia midiática dos seus eleitores é ativo de campanha. Invista nela.

Eleitores com maior capacidade crítica para identificar desinformação são mais difíceis de ser mobilizados contra o seu candidato por operações de fake news. Isso não é argumento altruísta: é argumento estratégico. Campanhas que distribuem conteúdo educativo sobre como identificar deepfakes, como verificar fontes de áudios políticos e como questionar mensagens de WhatsApp antes de repassá-las estão construindo imunidade eleitoral no seu eleitorado. Esse investimento produz retorno não apenas para a eleição em curso, mas para os ciclos seguintes, porque eleitores mais críticos são eleitores mais fiéis a quem os tratou com respeito.

6. O TikTok em 2026 merece o mesmo planejamento que o WhatsApp deveria ter recebido em 2018.

Em 2018, ninguém tinha estratégia específica para o WhatsApp porque ninguém havia mapeado seu potencial como vetor de desinformação eleitoral em escala. Em 2026, o TikTok está nessa mesma posição: crescimento acelerado entre eleitores jovens de primeira viagem, algoritmo de recomendação opaco, conteúdo em vídeo curto de difícil moderação em escala e ausência de regulação específica consolidada. Campanhas que chegarem a 2026 sem estratégia específica para o TikTok, tanto ofensiva quanto defensiva, estarão repetindo o erro que as campanhas de 2018 cometeram com o WhatsApp. A diferença é que agora o padrão é documentado e não há desculpa para a surpresa.

7. Documente tudo. A prova que você não produziu hoje não existirá quando precisar amanhã.

A maior fragilidade das investigações de 2018 e 2022 foi a dificuldade de produzir provas sobre operações conduzidas em canais privados e com pagamentos informais. Para a campanha que quer se defender de acusações de desinformação, e para a que quer responsabilizar adversários por operações ilegais, a documentação prévia é o único instrumento que funciona no tempo útil do processo eleitoral. Isso significa: registrar e autenticar notarialmente screenshots de conteúdo irregular assim que identificado; documentar todos os contratos com fornecedores de tecnologia eleitoral; manter registros de todas as decisões sobre uso de IA na campanha; e ter um protocolo de custódia de evidências digitais que resista ao contraditório num processo eleitoral.

CAPÍTULO 11
A Resolução TSE nº 23.732/2024

Análise do Primeiro Marco Normativo Mundial sobre IA em Eleições

"Regular a IA eleitoral é regular a própria condição de possibilidade da democracia na era digital. Não se trata de tecnologia. Trata-se de soberania popular."

Adaptado do Relatório de Justificação da Resolução TSE nº 23.732/2024

Regulação de IA Eleitoral no Direito Comparado

A Resolução TSE nº 23.732/2024 inseriu o Brasil num debate regulatório global que, até poucos anos antes de sua edição, era inteiramente dominado por democracias do hemisfério norte. Para compreender adequadamente sua relevância e suas limitações, é necessário situá-la no contexto mais amplo desse debate, que se intensificou dramaticamente após as eleições de 2016 nos Estados Unidos e no Reino Unido e acelerou de forma exponencial com a proliferação de ferramentas de IA generativa a partir de 2022.

O que estava em discussão teórica antes de 2016 tornou-se urgência regulatória após Cambridge Analytica. O que era urgência regulatória após Cambridge Analytica tornou-se crise regulatória após o ChatGPT. A Resolução TSE foi editada precisamente nesse segundo momento de aceleração, quando a janela para uma resposta normativa preventiva ainda estava aberta.

O debate regulatório internacional sobre IA e eleições ainda não produziu consenso sobre qual abordagem é mais eficaz, e a divergência não é apenas técnica: ela reflete visões distintas sobre onde deve residir a responsabilidade pela integridade eleitoral num ambiente tecnológico que as autoridades eleitorais não controlam. A literatura regulatória organiza esse debate em torno de três abordagens fundamentais que correspondem a filosofias regulatórias distintas.

A abordagem setorial, adotada pelo Brasil e, de forma mais fragmentada, pelos Estados Unidos, trata a IA eleitoral como subdomínio da regulação eleitoral geral, conferindo às autoridades eleitorais competência normativa específica para o período e o contexto do processo eleitoral. Sua vantagem é a especificidade: normas produzidas por autoridades eleitorais tendem a capturar com maior precisão os riscos específicos ao processo democrático. Sua limitação é o escopo temporal e espacial: a regulação eleitoral

opera sobre plataformas globais que existem independentemente do calendário eleitoral e que têm incentivos estruturais, como demonstrado no Capítulo 8, que não são neutralizados pelo fato de haver uma eleição em curso.

A abordagem horizontal, cuja expressão mais desenvolvida é o AI Act europeu aprovado em 2024, classifica os sistemas de IA por nível de risco e aplica obrigações proporcionais independentemente do setor de aplicação. Sistemas de IA utilizados para influenciar eleitores ou para produzir conteúdo político são classificados como de alto risco, sujeitando-se a obrigações de transparência, auditoria e avaliação de conformidade antes de sua disponibilização ao público. A vantagem dessa abordagem é a abrangência: ela alcança os sistemas de IA na fase de desenvolvimento e implantação, antes que produzam danos eleitorais. Sua limitação é a velocidade de resposta: processos de classificação e conformidade que levam meses ou anos são estruturalmente inadequados para um ambiente tecnológico em que novos modelos são lançados em semanas (UNIÃO EUROPEIA, 2024, considerandos 84-86).

A abordagem mista, emergente em diversas democracias que avaliaram as limitações dos dois modelos anteriores, combina regulação setorial eleitoral com obrigações gerais de transparência algorítmica aplicáveis às plataformas digitais independentemente do período eleitoral. O raciocínio subjacente é que os riscos eleitorais das plataformas não se iniciam na data do pleito: eles são produto de padrões de amplificação que operam continuamente e que apenas se tornam mais salientes no período eleitoral. Regular apenas o período eleitoral sem alterar os incentivos estruturais das plataformas é tratar o sintoma sem tocar na causa, argumento desenvolvido em detalhe no Capítulo 8.

Do conjunto dessas abordagens, a experiência brasileira destaca-se por dois elementos que merecem reconhecimento antes da análise crítica de suas limitações. O primeiro é a especificidade: a Resolução TSE nº 23.732/2024 é, ao tempo de sua edição, o instrumento normativo mais específico sobre IA eleitoral disponível em qualquer democracia, incluindo definições operacionais precisas, vedações absolutas e regime sancionatório próprio. Nenhum equivalente internacional combina os três elementos com o mesmo grau de detalhe normativo. O segundo elemento é a velocidade: o TSE produziu esse marco normativo antes do ciclo eleitoral de 2024, num processo que incluiu consulta pública e análise comparada das regulações internacionais disponíveis, demonstrando capacidade institucional de resposta adaptativa que contrasta favoravelmente com a inércia legislativa observada em democracias com sistemas mais lentos de produção normativa.

Contexto e Justificativas da Resolução

A Resolução TSE nº 23.732/2024 foi editada em 28 de março de 2024, com vigência imediata para as eleições municipais de outubro daquele ano. Sua edição não foi um exercício de precaução abstrata diante de riscos hipotéticos: foi uma resposta a um conjunto de fatores convergentes que tornavam a ausência de regulação específica juridicamente insustentável e politicamente indefensável para uma instituição com a tradição de inovação regulatória do TSE.

O primeiro fator foi a transformação radical do mercado de ferramentas de IA generativa entre novembro de 2022 e março de 2024. O lançamento do ChatGPT em novembro de 2022 não foi apenas um marco tecnológico: foi o momento em que a capacidade de produzir conteúdo sintético convincente migrou de ambientes de pesquisa especializada para o consumo de massa.

Em menos de dezoito meses, qualquer ator com acesso à internet e um cartão de crédito havia adquirido a capacidade de gerar texto, imagem, áudio e vídeo sintético com qualidade suficiente para enganar eleitores sem treinamento específico para identificar conteúdo artificial.

Ferramentas como Midjourney, DALL-E 3, ElevenLabs e Runway ML reduziram o custo marginal de produção de desinformação sofisticada a um nível que tornava o modelo regulatório anterior, baseado na premissa de que deepfakes exigiam recursos técnicos e financeiros significativos, empiricamente obsoleto.

O segundo fator foi a evidência internacional acumulada durante 2023 e 2024, analisada em detalhe no Capítulo 7. O áudio deepfake imitando a voz do presidente Biden nas primárias democratas do New Hampshire, o deepfake do candidato liberal eslovaco Michal Simecka divulgado no período de silêncio eleitoral e os vídeos sintéticos disseminados antes das eleições de Bangladesh demonstraram, em contextos eleitorais concretos e documentados, que a tecnologia estava sendo efetivamente utilizada para interferir em processos democráticos em múltiplos países. Esses casos removeram o argumento de que o risco era especulativo: era real, estava ocorrendo e seguia padrões operacionais identificáveis que uma regulação bem desenhada poderia endereçar.

O terceiro fator foi a lacuna normativa identificada pelo próprio TSE no mapeamento de sua regulação existente. A Resolução nº 23.610/2019, que havia estabelecido o marco para a propaganda eleitoral na internet, foi editada num contexto tecnológico anterior ao advento da IA generativa de qualidade comercial. Ela não contemplava a obrigação de identificação de conteúdo sintético, não definia deepfake eleitoral como categoria jurídica

autônoma e não previa regime sancionatório específico para as novas modalidades de violação. Aplicar essa norma às eleições de 2024 exigiria analogias jurídicas que, além de juridicamente questionáveis, seriam insuficientes para gerar o grau de previsibilidade e deterrência que uma regulação efetiva requer.

A convergência desses três fatores, democratização das ferramentas, evidência empírica de uso eleitoral e lacuna normativa identificada, criou a janela de oportunidade regulatória que o TSE utilizou para editar a Resolução nº 23.732/2024. A questão relevante para a análise que se segue não é por que a Resolução foi editada, o que é claro, mas o que ela conseguiu regular, o que ficou fora de seu alcance e o que isso implica para a preparação do sistema eleitoral brasileiro para os ciclos seguintes.

Estrutura da Resolução: Definições, Obrigações e Vedações

A Resolução TSE nº 23.732/2024 organiza-se em torno de três eixos normativos interdependentes que, em conjunto, constituem o que se pode denominar uma arquitetura regulatória integrada: um eixo definitório, que estabelece o vocabulário técnico-jurídico do campo e sem o qual os dois eixos seguintes não poderiam operar com precisão; um eixo obrigacional, que define as condutas exigidas dos diferentes atores da cadeia de produção e distribuição de conteúdo eleitoral; e um eixo proibitório, que estabelece as vedações absolutas e as sanções correspondentes, conferindo ao sistema sua capacidade dissuasória.

A interdependência entre os três eixos não é acidental: ela reflete uma escolha de técnica normativa que merece ser reconhecida como tal. Uma regulação que proibisse deepfakes eleitorais sem defini-los operacionalmente estaria criando uma norma aberta demais para gerar previsibilidade jurídica e restrita demais para ser aplicada com consistência pela jurisprudência. Uma regulação que definisse deepfakes sem estabelecer obrigações de rotulagem estaria alcançando apenas a ponta mais visível do problema, deixando sem cobertura as formas mais sutis de manipulação por IA que não chegam ao patamar da fabricação completa. E uma regulação que combinasse definições e obrigações sem regime sancionatório específico dependeria da aplicação analógica das sanções gerais da legislação eleitoral, com a imprecisão e a litigiosidade que isso inevitavelmente gera.

O Eixo Definitório: os Conceitos Operacionais

A precisão conceitual é o alicerce de qualquer regulação efetiva, e sua ausência é frequentemente o principal vetor de litígio e de evasão regulatória. Regulações que não definem com precisão o que proíbem oferecem aos operadores do sistema, tanto aos que buscam cumpri-la quanto aos que buscam evitá-la, um espaço de indeterminação que inevitavelmente se traduz em insegurança jurídica para uns e em oportunidade de evasão para outros.

A Resolução TSE nº 23.732/2024 respondeu a esse desafio com um conjunto de definições operacionais que, como analisado em detalhe no Capítulo 7, capturam com precisão técnica adequada à finalidade normativa os principais conceitos que estruturam o campo. A definição de inteligência artificial adota abordagem funcional e não exaustiva, descrevendo o que os sistemas de IA fazem em vez de como são construídos internamente, o que garante longevidade normativa sem sacrificar a precisão operacional. Uma definição baseada em arquitetura técnica específica, como redes neurais ou modelos de transformadores, estaria obsoleta antes do próximo ciclo eleitoral; uma definição baseada em capacidade funcional, como a da Resolução, acompanha a evolução tecnológica sem exigir emendas normativas contínuas.

A definição de conteúdo sintético, analisada pormenorizadamente no Capítulo 7, é o conceito central da Resolução e o que mais diretamente estrutura as vedações e obrigações subsequentes. Sua formulação em torno dos três verbos, gerar, manipular e modificar, cobre o espectro completo de intervenção algorítmica sobre conteúdo audiovisual, desde a criação integral por IA até as modificações pontuais que, sem fabricar o conteúdo do zero, distorcem a percepção do receptor sobre o que aconteceu. Essa abrangência é a que permite à Resolução alcançar tanto os deepfakes tecnicamente sofisticados quanto os cheap fakes descritos por Paris e Donovan (2019), que operam por manipulações simples mas igualmente capazes de enganar o eleitor.

📌 Definições Normativas Fundamentais da Resolução TSE nº 23.732/2024

Inteligência Artificial (art. 2º, I): conjunto de técnicas computacionais capazes de executar tarefas que normalmente requerem inteligência humana, incluindo, mas não se limitando a: processamento de linguagem natural, geração de texto, síntese de imagem, áudio e vídeo, reconhecimento de padrões, tomada de decisão automatizada e personalização de conteúdo, quando aplicadas em contexto de propaganda eleitoral, comunicação de campanha ou interação com eleitores. Conteúdo Sintético (art. 2º, II): qualquer conteúdo audiovisual (imagem, vídeo, áudio) gerado ou significativamente modificado por IA, de forma a alterar a aparência, a voz ou

as declarações de uma pessoa, com qualidade suficiente para induzir o observador comum ao erro sobre sua autenticidade. Agente Virtual (art. 2º, III): sistema de software que interage com eleitores de forma automatizada, total ou parcialmente, simulando ou assistindo comunicação que de outra forma seria conduzida por humanos.

Brasil. TSE. Resolução nº 23.732/2024, art. 2º.

O Eixo Obrigacional: Transparência e Rotulagem

Se o eixo definitório estabelece o vocabulário que torna a regulação inteligível, o eixo obrigacional é o que a torna operacional: ele define o que os atores da cadeia de produção e distribuição de conteúdo eleitoral são positivamente obrigados a fazer, independentemente de qualquer denúncia ou processo sancionatório. A distinção entre obrigações positivas e vedações absolutas é mais do que formal: ela reflete filosofias regulatórias distintas sobre onde deve recair o ônus da conformidade.

Vedações absolutas são instrumentos reativos, acionados quando uma conduta proibida já ocorreu. Obrigações positivas são instrumentos preventivos, que impõem custos de conformidade antes que qualquer violação se materialize e que, idealmente, tornam as violações menos prováveis por criarem um ambiente informacional em que o eleitor tem condições de avaliar criticamente o que recebe.

O mecanismo central de transparência estabelecido pela Resolução é a obrigação de rotulagem, prevista no artigo 15: todo conteúdo gerado ou significativamente modificado por inteligência artificial utilizado em propaganda eleitoral deve trazer identificação clara e ostensiva, em linguagem compreensível ao eleitor comum, de que se trata de conteúdo criado ou alterado por IA. Essa obrigação repousa sobre um fundamento normativo que é simultaneamente epistêmico e democrático. O fundamento epistêmico é o direito do eleitor de conhecer a origem e a natureza do conteúdo que recebe e que influenciará sua decisão política, o que a literatura denomina, na articulação com Fricker (2007) desenvolvida no Capítulo 7, direito epistêmico do eleitor. O fundamento democrático é que a legitimidade do resultado eleitoral depende, em alguma medida, de que as preferências dos eleitores sejam formadas em condições informacionais que preservem sua capacidade de avaliação autônoma.

A obrigação de rotulagem articula-se com a análise do modelo de negócios das plataformas desenvolvida no Capítulo 8 de forma que merece ser explicitada. Como demonstrado naquele capítulo, as plataformas têm incentivo estrutural para amplificar

conteúdo emocionalmente ativador independentemente de sua autenticidade. A rotulagem obrigatória de conteúdo gerado por IA introduz uma fricção informacional nesse processo: ao sinalizar ao eleitor que o conteúdo que está recebendo foi produzido artificialmente, ela ativa o ceticismo crítico que a confiança no mensageiro naturalmente suprime. A pesquisa empírica de Pennycook e Rand (2019) sugere que essa fricção produz efeitos mensuráveis sobre a intenção de compartilhar conteúdo não verificado, mesmo quando a intervenção é mínima (PENNYCOOK; RAND, 2019, p. 2523).

A expressão "significativamente modificado", contudo, introduz a principal zona de indeterminação normativa do artigo 15, cuja delimitação a jurisprudência do TSE ainda está construindo. A Resolução não especifica o limiar quantitativo ou qualitativo a partir do qual uma modificação por IA torna-se "significativa" para fins de ativação da obrigação de rotulagem, o que cria um espectro de incerteza com duas bordas razoavelmente claras e um vasto território intermediário juridicamente nebuloso.

Na borda inferior do espectro, ajustes elementares como correção automática de brilho e contraste por software de edição fotográfica que utilize algoritmos de aprendizado de máquina são intuitivamente insuficientes para acionar a obrigação. Na borda superior, a substituição facial completa num vídeo ou a clonagem de voz para produzir áudio sintético são igualmente intuitivas como acionadoras da obrigação. O problema regulatório está no território intermediário: uma fotografia de candidato em que o algoritmo de IA aprimorou a nitidez e removeu imperfeições de pele foi "significativamente modificada"? Um vídeo em que o modelo de IA removeu ruído de fundo e melhorou a qualidade do áudio está sujeito à rotulagem? Um texto de campanha redigido com assistência de modelo de linguagem mas substancialmente revisado e adaptado por redator humano é "gerado por IA" para fins da Resolução?

Essas perguntas não têm resposta unívoca no texto da norma, e a ausência de critérios mais precisos é a principal vulnerabilidade do eixo obrigacional da Resolução. Ela cria dois riscos simétricos: o risco de sub-rotulagem, em que atores que deveriam identificar conteúdo como gerado por IA não o fazem porque a zona cinzenta lhes oferece argumento de boa-fé; e o risco de sobre-rotulagem performática, em que atores identificam todo conteúdo como gerado por IA mesmo quando a assistência algorítmica foi mínima, diluindo o sinal informacional que a obrigação pretendia criar e treinando o eleitor a ignorar as identificações como ruído regulatório.

A resposta regulatória mais eficaz para essa indeterminação não é a emenda da Resolução para incluir critérios técnicos rígidos, que estariam desatualizados antes do

próximo ciclo eleitoral. É a construção jurisprudencial progressiva de parâmetros interpretativos que, partindo dos casos concretos levados ao TSE e aos TREs, vão delimitando o escopo da obrigação com base no critério material que a própria Resolução sugere: o potencial de engano sobre o receptor. Conteúdo que, por sua natureza ou pelo grau de intervenção algorítmica, tem capacidade de induzir o eleitor a acreditar que está diante de um registro autêntico quando não está é o que a obrigação de rotulagem foi projetada para identificar, independentemente do método técnico específico pelo qual a modificação foi produzida.

A obrigação de rotulagem não se aplica apenas aos produtores do conteúdo: ela vincula também os responsáveis pela veiculação da propaganda eleitoral, o que na prática significa os candidatos e os partidos que autorizam e publicam o material. Essa extensão da obrigação tem uma implicação que o assessor jurídico de campanha precisa compreender com precisão: o candidato que veicula conteúdo gerado por IA sem rotulagem adequada responde pela omissão mesmo que não tenha sido ele o produtor do conteúdo, e mesmo que a agência de marketing contratada não o tenha informado sobre os métodos utilizados. A responsabilidade pela conformidade é do responsável pela propaganda, não da cadeia de produção. Isso cria um incentivo estrutural para que campanhas incluam cláusulas contratuais específicas sobre uso de IA e rotulagem nos contratos com fornecedores de serviços de comunicação eleitoral.

O quadro a seguir traduz as obrigações de rotulagem em termos operacionais por tipo de conteúdo, especificando o formato exigido e a sanção correspondente pelo descumprimento. A leitura revela dois aspectos que complementam a análise da zona de indeterminação desenvolvida acima.

O primeiro é que a Resolução, ao especificar formatos distintos para cada tipo de mídia, reconhece implicitamente que a eficácia da rotulagem depende da compatibilidade entre o formato da identificação e o formato do conteúdo: uma legenda textual num vídeo que é assistido sem som não cumpre a função informativa que a obrigação pretende criar.

O segundo é a assimetria sancionatória entre os tipos de conteúdo: o chatbot eleitoral sujeita-se a multa substancialmente maior do que os demais formatos, o que reflete a avaliação regulatória de que a simulação de interação humana sem identificação constitui a forma mais grave de engano sobre a natureza do conteúdo, porque o eleitor acredita estar se comunicando com uma pessoa real, não apenas consumindo material de campanha.

Tabela 37. Obrigações de Rotulagem por Tipo de Conteúdo

Tipo de Conteúdo	Obrigação de Rotulagem	Local e Formato	Sanção pelo Descumprimento
Peça impressa de campanha gerada por IA	Obrigatória	Texto visível: 'Conteúdo gerado com auxílio de Inteligência Artificial', em fonte legível, em local de destaque.	Multa de R$ 5.000 a R$ 30.000 por peça.
Vídeo de campanha gerado ou modificado por IA	Obrigatória	Identificação ao início e ao final do vídeo, com duração mínima de 3 segundos em cada exibição.	Multa de R$ 5.000 a R$ 30.000 por peça; direito de resposta.
Áudio de campanha com síntese de voz	Obrigatória	Locução clara: 'Esta mensagem foi gerada por inteligência artificial', ao início e ao final.	Multa de R$ 5.000 a R$ 30.000 por peça; remoção imediata.
Publicação em rede social com conteúdo gerado por IA	Obrigatória	Identificação na legenda ou texto de apresentação da publicação, em linguagem clara.	Multa de R$ 5.000 a R$ 30.000; notificação à plataforma para remoção.
Chatbot eleitoral	Obrigatória em toda interação	Identificação imediata como sistema automatizado no início de cada conversa.	Multa de R$ 10.000 a R$ 50.000; cessação imediata do funcionamento.

Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024, arts. 15 a 19.

A obrigação de rotulagem tem duas funções normativas distintas que precisam ser compreendidas separadamente para que sua aplicação seja adequada e sua eficácia possa ser avaliada com precisão. A função direta é informar o eleitor sobre a natureza sintética do conteúdo que recebe, restituindo-lhe a condição de avaliar criticamente o que recebe com plena consciência de sua origem. Essa função opera sobre o receptor da mensagem: ela pressupõe que um eleitor informado de que está diante de conteúdo gerado por IA processará essa informação de forma diferente do que processaria se acreditasse estar diante de um registro autêntico.

A validade dessa premissa é empiricamente suportada, com as qualificações necessárias: pesquisas sobre o efeito de rótulos de advertência em conteúdo desinformativo sugerem que a identificação reduz a intenção de compartilhar e aumenta o ceticismo sobre a veracidade do conteúdo, mas com magnitudes que variam

significativamente conforme o perfil do receptor, o formato da identificação e o grau de congruência entre o conteúdo e as crenças preexistentes do eleitor (CLAYTON et al., 2020, p. 2039-2040). Em outras palavras: a rotulagem funciona melhor para eleitores que já estão dispostos a questionar o conteúdo e pior precisamente para aqueles que mais precisam de proteção, os que têm menor literacia midiática e maior propensão a aceitar conteúdo que confirma suas crenças preexistentes.

A função indireta é criar um registro de conformidade que facilita a fiscalização pela Justiça Eleitoral e que, pela via da dissuasão, incentiva a conformidade espontânea antes de qualquer processo sancionatório. Essa função opera sobre o produtor do conteúdo: ela pressupõe que o custo esperado da sanção pelo descumprimento, ponderado pela probabilidade de detecção, é suficiente para induzir comportamento conforme. A eficácia dessa função depende crucialmente de dois fatores que o texto da Resolução não resolve sozinho: a probabilidade real de detecção de violações, que como analisado nas seções anteriores é estruturalmente baixa num ambiente de produção em escala industrial; e a magnitude da sanção, que precisa ser suficientemente elevada para superar o benefício esperado da violação. O quadro sancionatório atual, com multas que variam de R$ 5.000 a R$ 50.000 por peça, pode ser insuficiente para campanhas bem financiadas que produzem conteúdo em volumes que tornam a detecção estatisticamente improvável.

Ambas as funções são necessárias e nenhuma, isoladamente, é suficiente. A rotulagem que informa mas não dissuade é vulnerável à não-conformidade calculada. A rotulagem que dissuade mas não informa falha em sua missão democrática central quando, apesar da ameaça de sanção, o conteúdo irregular já circulou e o dano epistêmico já foi causado antes de qualquer intervenção. A eficácia real da obrigação de rotulagem depende, portanto, da combinação das duas funções com os mecanismos de detecção e sanção que as tornam operacionais na prática.

O Eixo Proibitório: Vedações Absolutas

A Resolução distingue, com precisão técnica relevante, condutas reguladas de condutas vedadas absolutamente. Condutas reguladas são aquelas que a norma permite, desde que observados requisitos específicos: o uso de IA generativa na produção de conteúdo eleitoral é, em si, conduta regulada, sujeita à obrigação de rotulagem mas não proibida. Condutas vedadas absolutamente são aquelas para as quais não existe possibilidade de regularização mediante cumprimento de qualquer requisito: a proibição é

total, independentemente da intenção declarada, da qualidade técnica do conteúdo ou do contexto em que seria utilizado.

Essa distinção reflete uma escolha de política regulatória que merece ser explicitada, porque ela é frequentemente mal compreendida no debate público. A existência de vedações absolutas não significa que o regulador considerou toda a conduta no campo como igualmente danosa: significa que identificou um subconjunto de condutas em relação às quais o risco de dano à integridade eleitoral é suficientemente elevado e a possibilidade de uso legítimo suficientemente baixa para que a ponderação caso a caso seja indesejável. Normas que exigem ponderação caso a caso são inevitavelmente mais lentas na aplicação, mais custosas na fiscalização e mais vulneráveis à manipulação por atores que constroem narrativas de justificação sofisticadas. Vedações absolutas sacrificam a flexibilidade em favor da previsibilidade e da deterrência, e esse é um trade-off regulatório consciente que o TSE fez ao estruturar o eixo proibitório da Resolução.

As vedações absolutas estabelecidas pela Resolução TSE nº 23.732/2024 concentram-se, fundamentalmente, na representação sintética não autorizada de candidatos: a criação ou disseminação de conteúdo que faça um candidato aparentar dizer, fazer ou expressar posições que nunca manifestou, sem seu consentimento e com potencial de enganar o eleitor sobre a autenticidade do conteúdo. O artigo 9º, que estabelece essa vedação central, não exige a demonstração de dano efetivo para que a conduta seja tipificada: o potencial de engano é suficiente. Essa escolha é normativamente relevante porque reconhece que, em ambiente de rápida circulação de conteúdo, exigir a demonstração de dano efetivo como condição de responsabilização é conferir aos produtores de deepfakes eleitorais uma janela de impunidade estrutural, porque o dano, quando demonstrável, já é irreversível.

A vedação não alcança, contudo, todo uso de representação sintética de candidatos. O uso autorizado e transparentemente identificado de avatares digitais, como o caso argentino do candidato Milei analisado no Capítulo 7, permanece fora do escopo da proibição quando a identificação obrigatória é observada. Essa delimitação do perímetro da vedação é analiticamente importante: ela confirma que o objeto da proibição não é a tecnologia em si, mas o engano. Tecnologia usada com transparência é regulada; tecnologia usada para enganar é vedada. O critério estruturante do eixo proibitório é o mesmo que estrutura os cinco princípios éticos desenvolvidos no Capítulo 7: a relação que cada uso estabelece com a autonomia do eleitor.

O quadro a seguir sistematiza as cinco vedações absolutas estabelecidas pelo artigo 9º da Resolução, com seus fundamentos normativos específicos e as sanções correspondentes. A leitura por linhas revela que, embora todas as vedações compartilhem o mesmo fundamento geral, a proteção da autonomia do eleitor e da autenticidade do processo eleitoral, cada uma endereça um vetor de dano distinto: a primeira protege a honra do candidato adversário; a segunda protege o eleitor do engano sobre a natureza da interação; a terceira protege o direito de imagem e voz do candidato; a quarta protege a confiança nas instituições eleitorais como tal; e a quinta, a mais abrangente e a mais difícil de provar, protege a capacidade deliberativa do eleitor contra manipulação psicológica que opera abaixo do limiar da consciência. A leitura por colunas é igualmente reveladora: a progressão das sanções reflete a avaliação do legislador normativo sobre a gravidade relativa de cada conduta, com a representação não autorizada de autoridades e instituições eleitorais sujeita à multa mais elevada, porque seu dano não é apenas individual mas sistêmico, comprometendo a confiança no processo eleitoral como um todo e não apenas numa candidatura específica.

Tabela 38. Vedações Absolutas ao Uso de IA em Campanhas Eleitorais

Conduta Vedada	Base Legal	Fundamento da Vedação	Sanção Aplicável
Criação de deepfake de candidato adversário fazendo declaração falsa	Art. 9º, I	Comprometimento da autenticidade do debate eleitoral; dano irreparável à honra do candidato.	Multa de R$ 20.000 a R$ 100.000 por peça; crime eleitoral de calúnia (art. 324 CE).
Simulação de interação direta com candidato sem identificação como IA	Art. 9º, II	Engano sobre a natureza da comunicação; violação da autonomia informacional do eleitor.	Multa de R$ 10.000 a R$ 50.000; cassação do registro do candidato se reiterada.
Clonagem de voz de candidato para produzir áudio com declaração falsa	Art. 9º, III	Violação do direito de imagem e voz; dano à honra; risco de supressão de voto.	Multa de R$ 20.000 a R$ 100.000; responsabilização penal eleitoral.
Uso de IA para representação não autorizada de autoridades ou instituições eleitorais	Art. 9º, IV	Risco de erosão da confiança nas instituições; desinformação sobre o processo eleitoral.	Multa de R$ 50.000 a R$ 200.000; comunicação ao Ministério Público.
Uso de técnicas subliminares ou	Art. 9º, V	Violação da autonomia do eleitor; manipulação	Multa de R$ 20.000 a R$ 100.000; representação por propaganda irregular.

deliberadamente manipulativas por IA		psicológica incompatível com a democracia deliberativa.	

Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024, art. 9º.

Extensão às Normas de Pré-Campanha

Uma das inovações mais relevantes da Resolução TSE nº 23.732/2024 é a extensão explícita de suas obrigações ao período de pré-campanha, fechando uma lacuna que havia sido sistematicamente explorada em ciclos eleitorais anteriores com uma racionalidade jurídica aparentemente defensável: se as normas eleitorais ainda não estavam formalmente vigentes, a conduta não poderia ser por elas alcançada. Esse argumento tinha uma lógica formal que a jurisprudência do TSE progressivamente recusou, mas que continuava gerando insegurança jurídica e incentivando a concentração de operações de desinformação no período pré-eleitoral precisamente para explorar a janela de menor regulação.

O fenômeno não era novo nem circunstancial. Em cada ciclo eleitoral, atores com recursos e assessoria jurídica sofisticada identificavam o período de pré-campanha como a janela de menor exposição regulatória e concentravam nele as operações mais agressivas de comunicação política, incluindo a construção de narrativas negativas sobre adversários e a disseminação de conteúdo desinformativo que, uma vez estabelecido no ecossistema informacional, continuaria circulando e produzindo efeitos durante o período oficial de campanha sem poder ser rastreado até sua origem pré-eleitoral. A introdução da IA generativa amplificou esse problema porque reduziu o tempo necessário para produzir conteúdo em escala: uma operação que antes exigia meses de preparação passou a ser executável em dias, tornando a janela pré-eleitoral ainda mais atraente para atores que queriam maximizar impacto com mínima exposição regulatória.

O critério adotado pela Resolução para a extensão ao período de pré-campanha é funcional e não temporal, e essa escolha de técnica normativa é precisamente o que confere à inovação sua eficácia. O que importa para fins de incidência da norma não é a data em que o conteúdo foi produzido ou publicado, mas sua natureza eleitoral. Conteúdo que tenha por finalidade, ainda que implícita, a promoção de candidatura ou a captação de apoio eleitoral está sujeito às normas da Resolução independentemente de ter sido produzido antes ou durante o período oficial de campanha. Esse critério funcional é coerente com a jurisprudência consolidada do TSE sobre propaganda eleitoral antecipada,

que há anos rejeita a argumentação temporal como critério determinante e adota a finalidade eleitoral como elemento central de tipificação da conduta irregular.

A extensão funcional tem, contudo, uma zona de indeterminação própria que o intérprete precisa reconhecer. A fronteira entre atividade política ordinária de um pré-candidato, que inclui legitimamente a construção de sua imagem pública, a comunicação de suas posições sobre temas de interesse coletivo e a manutenção de presença nas redes sociais, e a propaganda eleitoral antecipada não é sempre nítida. Um vídeo gerado com auxílio de IA em que o pré-candidato comenta uma questão de política pública sem fazer referência explícita à sua candidatura está sujeito à obrigação de rotulagem? A resposta depende da avaliação da finalidade predominante do conteúdo, que é uma questão de prova e de interpretação que a jurisprudência construirá caso a caso.

Para o assessor jurídico de campanha, essa indeterminação tem uma implicação prática simples: a estratégia mais segura é tratar todas as obrigações da Resolução como vigentes desde o momento em que o cliente passa a ter interesse eleitoral identificável, independentemente de qualquer análise sobre se determinado conteúdo específico poderia ser enquadrado como pré-eleitoral ou não. O custo da conformidade preventiva é sistematicamente menor do que o custo da defesa em processos de propaganda eleitoral antecipada, especialmente quando o conteúdo em questão foi produzido com IA e está, portanto, sujeito ao regime sancionatório específico da Resolução em cumulação com os instrumentos gerais de combate à propaganda irregular.

A extensão ao período de pré-campanha tem ainda uma implicação que vai além da conformidade individual de cada campanha: ela cria um incentivo para que plataformas digitais mantenham seus sistemas de monitoramento de conteúdo eleitoral ativos durante todo o ano eleitoral, e não apenas durante o período oficial de campanha. Plataformas que desativam ou reduzem seus sistemas de fiscalização eleitoral fora do período oficial, argumentando que a regulação eleitoral não se aplica, passam a ter exposição regulatória no modelo da Resolução: se o conteúdo tem finalidade eleitoral, a obrigação de cooperação com o TSE se aplica independentemente do calendário.

Essa expansão temporal das obrigações das plataformas é um avanço normativo que o DSA europeu também buscou operacionalizar, ao exigir que Very Large Online Platforms mantenham capacidades de avaliação de risco eleitoral ativas ao longo de todo o período em que eleições relevantes estejam previstas nos países em que operam (UNIÃO EUROPEIA, 2022, art. 34, §1º).

ATENÇÃO PRÁTICA — Para o Assessor de Campanha

A extensão das normas ao período de pré-campanha tem implicação prática imediata: todo conteúdo gerado com auxílio de IA para fins políticos, mesmo que publicado antes do início oficial da campanha, está sujeito à obrigação de rotulagem e às vedações absolutas da Resolução TSE nº 23.732/2024. A produção e publicação de conteúdo sintético de candidatos adversários em período de pré-campanha, com o argumento de que ainda não há campanha formal, configura violação da norma e expõe o responsável a todo o regime sancionatório previsto.

Lacunas Identificadas e Perspectivas de Aperfeiçoamento

A análise crítica honesta da Resolução TSE nº 23.732/2024 exige identificar, além de seus avanços genuínos, as lacunas que persistem e que podem ser exploradas por atores maliciosos ou que podem gerar insegurança jurídica para atores de boa-fé. Esse exercício não é destrutivo: é a condição para que o aperfeiçoamento regulatório seja orientado por diagnóstico preciso em vez de por reação a crises já ocorridas. Uma regulação que não é submetida a análise crítica sistemática tende a ser aperfeiçoada por pressão de escândalos, o que produz reformas apressadas e frequentemente mal calibradas. Uma regulação que é continuamente avaliada por seus aplicadores, pelos atores que a ela se submetem e pela comunidade acadêmica que a estuda tem melhores condições de evoluir de forma coerente com seus objetivos originais.

A primeira lacuna diz respeito à definição técnica precisa do limiar de intervenção algorítmica que aciona as obrigações da Resolução. Como analisado na seção sobre o eixo obrigacional, a expressão "significativamente modificado" por inteligência artificial cria uma zona de indeterminação que afeta tanto a conformidade de boa-fé quanto a fiscalização. A Resolução define conteúdo sintético de forma funcional, com base no efeito produzido sobre o receptor, mas não especifica os critérios técnicos que determinam quando uma intervenção algorítmica é suficientemente substancial para acionar a obrigação de rotulagem.

Essa imprecisão produz consequências assimétricas conforme o perfil do ator. Campanhas com assessoria jurídica sofisticada e acesso a consultores técnicos especializados conseguem navegar a zona cinzenta com relativa segurança, construindo argumentações defensáveis sobre porque determinado conteúdo não está sujeito à rotulagem. Campanhas menores, sem esses recursos, enfrentam a escolha entre rotular tudo indiscriminadamente, diluindo o sinal informacional da obrigação, ou não rotular nada

além do óbvio, correndo o risco de violação não intencional. Essa assimetria de capacidade de conformidade é, ela própria, um problema regulatório: uma norma que protege melhor os atores mais sofisticados está produzindo resultado oposto ao pretendido.

A solução mais promissora para essa lacuna não é a emenda da Resolução para incluir critérios técnicos rígidos, que estariam desatualizados antes do próximo ciclo eleitoral dado o ritmo de evolução das ferramentas. É a edição de instruções normativas complementares que, sem alterar a norma principal, estabeleçam exemplos ilustrativos de condutas conformes e não conformes organizados por tipo de intervenção, funcionando como safe harbors regulatórios que reduzam a incerteza para atores de boa-fé sem fechar a norma para condutas ainda não antecipadas.

A segunda lacuna é a ausência de mecanismo de certificação prévia de ferramentas de IA utilizadas em campanhas eleitorais. A Resolução estabelece obrigações para os responsáveis pelo conteúdo, candidatos e campanhas, mas não cria nenhum processo pelo qual as ferramentas utilizadas na produção desse conteúdo possam ser previamente avaliadas quanto à sua conformidade com os requisitos normativos. Isso transfere integralmente a responsabilidade da conformidade técnica para o candidato e sua equipe, que frequentemente não têm capacidade técnica para avaliar se uma ferramenta específica gera conteúdo que exige rotulagem, se seus outputs incluem os marcadores de autenticidade recomendados ou se seus processos internos de geração deixam rastros auditáveis que poderiam ser utilizados em investigações eleitorais.

O modelo de certificação de ferramentas existe em outros campos regulatórios que enfrentam desafios técnicos similares. A regulação de dispositivos médicos, por exemplo, certifica as ferramentas antes de responsabilizar os usuários por condutas que dependem de características técnicas que esses usuários não podem verificar individualmente. O campo eleitoral poderia desenvolver modelo análogo: um sistema voluntário de certificação pelo qual desenvolvedores de ferramentas de IA utilizadas em comunicação política submetem seus produtos a avaliação técnica do TSE ou de auditores credenciados, obtendo certificação que tanto serve como safe harbor para as campanhas que as utilizam quanto como instrumento de monitoramento técnico para a Justiça Eleitoral. Esse sistema não eliminaria a responsabilidade das campanhas, mas redistribuiria de forma mais equitativa o ônus de conformidade técnica entre quem tem capacidade de avaliá-la, os desenvolvedores, e quem não tem, os candidatos e seus assessores.

A terceira lacuna é a ausência de protocolo técnico integrado de detecção e remoção de conteúdo de IA sem rotulagem em escala. A obrigação de rotulagem pressupõe, para sua efetividade, que as violações possam ser identificadas: um sistema que exige

identificação mas não tem meios de verificar se ela está presente é um sistema que depende inteiramente da conformidade voluntária ou da denúncia por terceiros. Ambos os mecanismos são insuficientes num ambiente de produção de conteúdo em escala industrial.

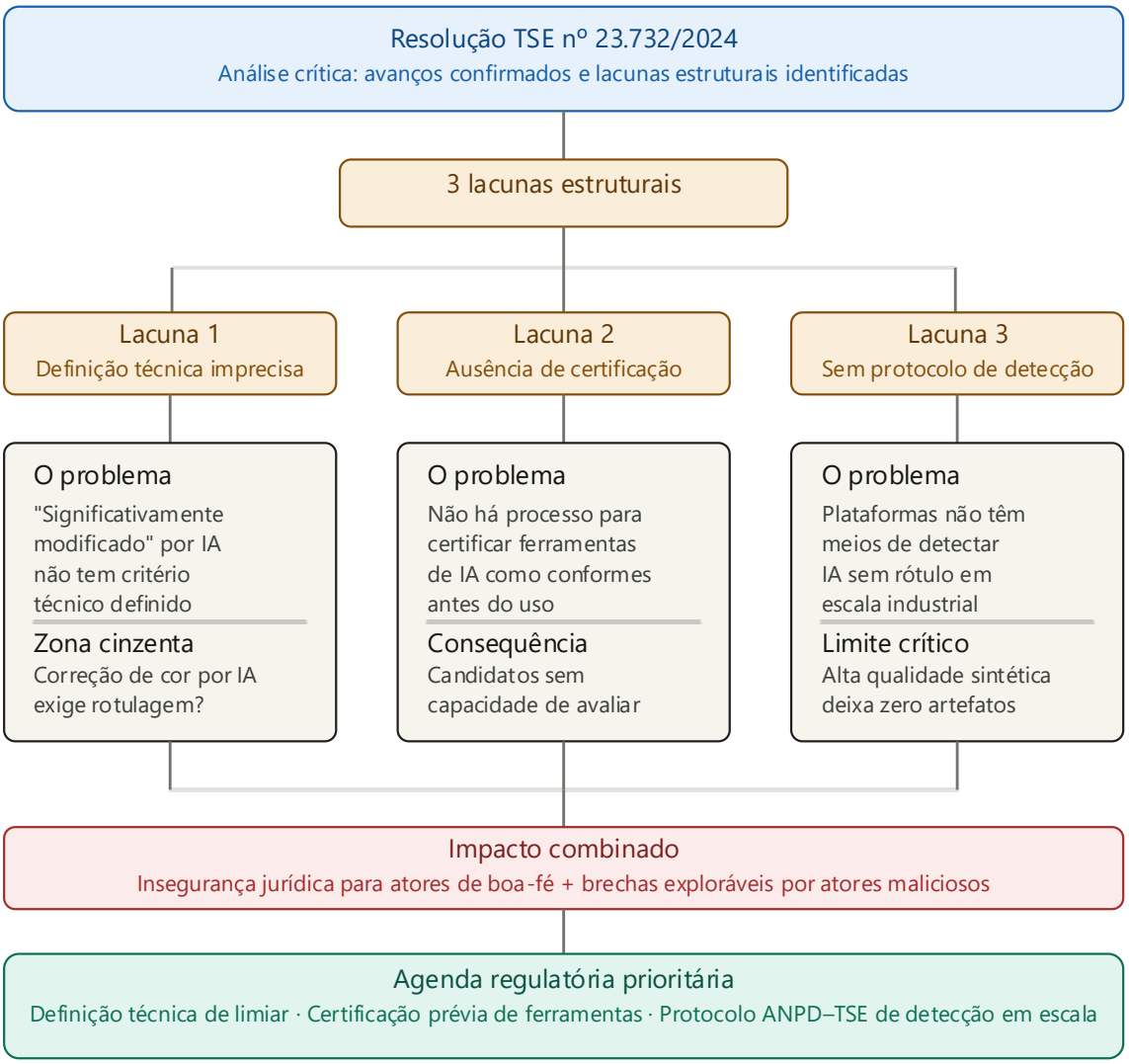
A capacidade técnica de detecção de conteúdo sintético por sistemas automatizados ainda é limitada, especialmente para conteúdo de alta qualidade produzido por modelos de difusão avançados que não deixam os artefatos visuais e auditivos característicos das gerações anteriores de deepfakes. Como documentaram Nightingale e Farid (2022) em pesquisa sobre a capacidade humana e automatizada de detecção de faces sintéticas, mesmo observadores treinados e sistemas de detecção especializados têm taxas de erro significativas diante de conteúdo produzido pelos modelos mais avançados (NIGHTINGALE; FARID, 2022, p. 12974). O problema é estruturalmente assimétrico: os modelos de geração evoluem mais rapidamente do que os sistemas de detecção, porque detecção é um problema mais difícil do que geração, o que significa que a lacuna técnica entre produzir e identificar conteúdo sintético tende a se ampliar com o tempo, não a se reduzir.

A resposta regulatória mais promissora para essa lacuna não está na competição entre sistemas de geração e detecção, corrida que os sistemas de geração tendem a vencer. Está na adoção de soluções de proveniência de conteúdo, como os padrões desenvolvidos pela Coalition for Content Provenance and Authenticity (C2PA), que estabelecem mecanismos de autenticação criptográfica para registrar a origem e o histórico de modificações de conteúdo audiovisual desde sua criação. Num sistema baseado em proveniência, a questão regulatória deixa de ser "este conteúdo foi gerado por IA?", que é tecnicamente difícil de responder após o fato, e passa a ser "este conteúdo tem cadeia de proveniência verificável?", que é tecnicamente muito mais acessível. A integração desse padrão aos requisitos técnicos de conformidade da Resolução é uma das recomendações mais concretas que a literatura técnico-regulatória tem produzido para o aperfeiçoamento de marcos normativos similares ao brasileiro (C2PA, 2024).

A quarta lacuna, que o texto original não identificou mas que emerge com clareza da análise dos três ciclos eleitorais brasileiros, é a ausência de mecanismo de revisão periódica obrigatória da Resolução. Normas tecnológicas que não têm prazo ou gatilho de revisão tendem a envelhecer silenciosamente: continuam formalmente em vigor enquanto sua efetividade se deteriora progressivamente à medida que a tecnologia que regulam evolui. A Resolução TSE nº 23.732/2024 foi bem calibrada para o estado da arte tecnológico de março de 2024. Em 2026, quando as eleições gerais ocorrerão, esse estado da arte será substancialmente diferente. Sem um mecanismo institucionalizado de revisão que inclua

consulta pública, análise técnica especializada e comparação com as regulações internacionais mais avançadas disponíveis, o TSE dependerá da iniciativa ad hoc de seus ministros para manter a norma atualizada, o que é um mecanismo frágil para uma responsabilidade regulatória dessa magnitude.

O modelo mais robusto disponível como referência é o ciclo de revisão obrigatória estabelecido pelo AI Act europeu, que prevê avaliação periódica das classificações de risco e dos requisitos de conformidade à medida que a tecnologia evolui, com participação de stakeholders e publicação de relatórios de avaliação (UNIÃO EUROPEIA, 2024, art. 112). Adaptar esse modelo ao contexto do direito eleitoral brasileiro, com seus prazos e restrições institucionais específicos, é uma tarefa que a doutrina e a prática eleitoral precisarão endereçar antes das eleições de 2026.



Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024; Nightingale e Farid (2022); e C2PA Technical Specification 2.0 (2024).

O TSE como Pioneiro Mundial: Avaliação Comparada

A afirmação de que o Brasil ocupa posição de vanguarda na regulação de IA eleitoral não é retórica institucional: é uma conclusão que a análise comparada dos marcos normativos disponíveis confirma, com as qualificações necessárias para que o reconhecimento não se converta em complacência.

O contexto em que essa liderança emergiu merece ser compreendido em sua especificidade. Democracias com sistemas legislativos tradicionais, nos quais a produção normativa depende de processos parlamentares longos e sujeitos a múltiplos vetos, enfrentaram dificuldade estrutural para responder com agilidade à aceleração tecnológica posterior ao ChatGPT. Os Estados Unidos, com seu sistema federativo e sua tradição de resistência à regulação de conteúdo amparada na Primeira Emenda, chegaram às eleições de 2024 sem lei federal abrangente sobre IA eleitoral. O Reino Unido editou o Online Safety Act em 2023, mas com foco em moderação de conteúdo prejudicial em geral, sem tratamento específico dos riscos da IA generativa no contexto eleitoral. A União Europeia, com o AI Act e o DSA, produziu os marcos horizontais mais ambiciosos disponíveis, mas com implementação gradual que se estende até 2027 e territorialidade limitada ao espaço europeu.

O TSE, operando num modelo de autoridade eleitoral com poder normativo próprio que não depende de aprovação parlamentar para editar resoluções no âmbito de sua competência, tinha uma vantagem institucional estrutural nesse contexto: podia agir com velocidade que legislaturas não conseguem replicar. A Resolução nº 23.732/2024 foi o produto dessa vantagem, combinada com a experiência acumulada nos ciclos eleitorais de 2018, 2020 e 2022 e com a análise sistemática das regulações internacionais disponíveis.

O reconhecimento da liderança brasileira, porém, exige duas qualificações que a análise honesta não pode omitir. A primeira é que pioneirismo normativo não equivale a efetividade regulatória: uma norma pode ser a mais abrangente e específica do mundo e ainda assim ter limitações operacionais que reduzem sua eficácia real, como as lacunas identificadas na seção anterior demonstram. A segunda é que a liderança é temporária por natureza: outros sistemas regulatórios estão em construção acelerada, e a vantagem comparativa do Brasil será progressivamente reduzida à medida que o AI Act europeu entre em plena vigência e que democracias como Índia e Canadá aperfeiçoem seus instrumentos. O valor do pioneirismo está em estabelecer padrões que outros adotem e em construir capacidade institucional que se acumula ao longo dos ciclos eleitorais, não em

manter indefinidamente uma posição de liderança que o avanço global inevitavelmente relativiza.

O quadro a seguir sistematiza a comparação entre os seis marcos regulatórios mais relevantes disponíveis, permitindo identificar com precisão em que dimensões o modelo brasileiro se destaca, em que dimensões outros sistemas oferecem soluções complementares e quais lacunas permanecem sem resposta adequada em qualquer dos marcos analisados.

Tabela 39. Comparação Internacional de Marcos Regulatórios sobre IA Eleitoral

País/Bloco	Marco Principal	Data	Abrangência	Ponto Forte	Limitação Principal
Brasil	Resolução TSE nº 23.732/2024	Mar. 2024	Específica para IA eleitoral	Mais abrangente e específica do mundo; vedações absolutas; sanções proporcionais.	Dependência de cooperação voluntária das plataformas; lacuna em definições técnicas.
União Europeia	AI Act + DSA (2024)	Jun. 2024	Horizontal (todos os setores) + plataformas	Obrigações vinculantes de auditoria algorítmica para Very Large Online Platforms.	Implementação gradual até 2027; territorialidade limitada ao espaço europeu.
Estados Unidos	Diretrizes FEC (2024)	2024	Parcial (publicidade política)	Foco em anúncios políticos pagos com identificação obrigatória.	Fragmentado: sem lei federal abrangente; varia por estado.
Reino Unido	Online Safety Act (2023)	Out. 2023	Plataformas digitais em geral	Obrigações de moderação de conteúdo prejudicial.	Não trata especificamente de IA eleitoral.
Índia	Código de Conduta Eleitoral (2024)	2024	Específica para IA eleitoral (eleições 2024)	Implementação rápida em contexto de 970 milhões de eleitores.	Autorregulação sem mecanismos robustos de enforcement.
Canadá	Elections Canada	2024	Parcial (transparência)	Diretrizes sobre identificação de	Caráter orientativo, sem força vinculante.

	Guidelines (2024)			conteúdo gerado por IA.	

Fonte: Elaborado pelo autor com base em Resolução TSE 23.732/2024, AI Act (2024), DSA (2022), Online Safety Act (2023) e relatórios eleitorais do IFES (2024).

O Calendário Eleitoral de 2026 e as Novas Obrigações Regulatórias

A Resolução TSE nº 23.760, de 2 de março de 2026, que estabelece o Calendário Eleitoral para as Eleições Gerais de 2026, incorpora de forma expressa e inéditas dispositivos relacionados ao uso de inteligência artificial em propaganda eleitoral, consolidando a integração entre o marco normativo de IA eleitoral e o calendário processual do pleito. Essa integração é mais do que simbólica: ela transforma obrigações que na Resolução TSE nº 23.732/2024 eram tratadas como matéria regulatória autônoma em prazos e vedações com marco temporal preciso, dotados da mesma força cogente das demais obrigações eleitorais.

O dispositivo mais relevante para fins de regulação de IA eleitoral é o que incide sobre os dias imediatamente anteriores e posteriores ao pleito. Tanto para o primeiro turno, fixado em 4 de outubro de 2026, quanto para o segundo, em 25 de outubro, o Calendário estabelece que a partir de três dias antes de cada turno e até vinte e quatro horas após o encerramento da votação ficam vedadas a publicação e a republicação, ainda que gratuitas, bem como o impulsionamento pago de novos conteúdos sintéticos produzidos ou alterados por inteligência artificial ou por tecnologias equivalentes que utilizem imagem, voz ou manifestação de candidato ou de pessoa pública, mesmo que rotulados e em conformidade com as demais exigências previstas no artigo 9º-B da Resolução TSE nº 23.610/2019.

Essa vedação ampliada no período crítico pré e pós-eleitoral representa uma resposta direta ao padrão de timing estratégico documentado nos casos internacionais e brasileiros analisados nos capítulos anteriores: ao proibir não apenas o conteúdo irregular, mas qualquer novo conteúdo sintético, mesmo o rotulado, nos dias de maior vulnerabilidade eleitoral, o TSE fechou a janela que operações sofisticadas exploravam deliberadamente.

A extensão da vedação ao conteúdo rotulado nos três dias anteriores ao pleito merece análise específica porque representa uma inflexão na lógica regulatória da Resolução nº 23.732/2024. Se a norma de 2024 distinguia conteúdo sintético irregular, sujeito a vedação absoluta, de conteúdo sintético transparente, sujeito apenas à obrigação de rotulagem, a nova vedação temporal colapsa essa distinção: nesse período, nem a rotulagem é suficiente para legitimar a publicação de novo conteúdo sintético. O fundamento é o mesmo que orienta todo o eixo proibitório da regulação de IA eleitoral: o

dano epistêmico causado pela exposição a conteúdo sintético nos dias imediatamente anteriores ao voto é de natureza distinta e mais grave do que o dano causado em outros momentos do ciclo eleitoral, porque o tempo para processamento crítico, verificação e formação de julgamento é mínimo.

O Calendário revela também o estágio de maturidade que a regulação de IA eleitoral brasileira atingiu ao ser incorporada à arquitetura processual do pleito. Em 2018, a desinformação digital operava num vazio regulatório. Em 2022, a regulação existia mas precisava ser aplicada por analogia. Em 2024, a Resolução TSE nº 23.732 forneceu o marco normativo específico. Em 2026, esse marco foi integrado ao calendário eleitoral oficial, com prazos, vedações e instrumentos de enforcement que tratam o conteúdo sintético com o mesmo rigor processual aplicado à propaganda eleitoral tradicional. A distância percorrida em oito anos é substancial; a distância ainda a percorrer, como as lacunas identificadas na seção anterior demonstram, permanece significativa.

Para o profissional de campanha e para o assessor jurídico eleitoral, os marcos do Calendário de 2026 com impacto direto sobre o uso de IA na comunicação política exigem atenção específica. A data de 16 de agosto de 2026, a partir da qual é permitida a propaganda eleitoral inclusive na internet, é também a data a partir da qual as obrigações de rotulagem da Resolução nº 23.732/2024 passam a incidir sobre todo conteúdo publicitado. A data de 1º de outubro de 2026, três dias antes do primeiro turno, é o marco a partir do qual qualquer novo conteúdo sintético, mesmo rotulado, está vedado. E a data de 22 de outubro de 2026, três dias antes do segundo turno, replica a mesma vedação para o segundo ciclo de votação. Essas três datas constituem os marcos regulatórios de IA mais críticos do ciclo eleitoral de 2026, e campanhas que não os incorporarem em seus protocolos internos de conformidade estarão operando com risco jurídico relevante e desnecessário.

■ RESUMO DO CAPÍTULO:

Este capítulo analisou a Resolução TSE nº 23.732/2024 como o primeiro marco normativo mundial especificamente voltado para a regulação de inteligência artificial em processos eleitorais, situando-a no debate regulatório internacional e examinando sua arquitetura interna, seus avanços, suas lacunas e sua integração ao calendário eleitoral de 2026.A análise comparada das três abordagens regulatórias disponíveis, a setorial adotada pelo Brasil, a horizontal do AI Act europeu e a mista emergente em outras democracias, revelou que cada uma reflete uma filosofia distinta sobre onde deve residir a responsabilidade pela integridade eleitoral num ambiente tecnológico que as autoridades eleitorais não controlam. O modelo brasileiro destaca-se pela especificidade e pela velocidade de resposta, duas vantagens estruturais derivadas do poder normativo próprio do TSE que legislaturas parlamentares não conseguem replicar. O reconhecimento dessa liderança foi qualificado por duas ressalvas essenciais: pioneirismo normativo não equivale a efetividade regulatória, e a liderança é temporária num campo em

aceleração global.A análise dos fatores que motivaram a edição da Resolução identificou três convergências: a democratização radical das ferramentas de IA generativa após o ChatGPT, que tornou empiricamente obsoleta a premissa de que deepfakes exigiam recursos significativos; a evidência internacional acumulada em 2023 e 2024 de uso eleitoral concreto dessas tecnologias em múltiplos países; e a lacuna normativa identificada pelo próprio TSE no mapeamento de sua regulação existente.A arquitetura da Resolução foi examinada em torno de seus três eixos interdependentes. O eixo definitório, com a abordagem funcional das definições de inteligência artificial, conteúdo sintético e agente virtual, garante longevidade normativa sem sacrificar precisão operacional. O eixo obrigacional, centrado na obrigação de rotulagem do artigo 15, foi analisado em suas duas funções normativas, a função direta de informar o eleitor e a função indireta de criar registro de conformidade e dissuasão, com as qualificações empíricas sobre a eficácia de cada uma e os dois riscos simétricos que a zona de indeterminação do "significativamente modificado" cria. O eixo proibitório, com as cinco vedações absolutas do artigo 9º, foi analisado como trade-off consciente entre flexibilidade e previsibilidade, com a lógica específica de cada vedação e a progressão sancionatória que reflete a gravidade relativa dos danos endereçados.A extensão das obrigações ao período de pré-campanha, por critério funcional e não temporal, foi identificada como inovação que fecha uma lacuna sistematicamente explorada, com as zonas de indeterminação residuais que a jurisprudência precisará delimitar. Quatro lacunas estruturais foram identificadas: a imprecisão do limiar de intervenção algorítmica, a ausência de certificação prévia de ferramentas, a falta de protocolo técnico de detecção em escala e a ausência de mecanismo de revisão periódica obrigatória, com propostas específicas para endereçar cada uma. O capítulo encerrou com a análise da Resolução TSE nº 23.760/2026, que integra a regulação de IA ao calendário eleitoral oficial e estabelece a vedação ampliada dos três dias pré-eleitorais, que colapsa a distinção entre conteúdo irregular e conteúdo rotulado no período de maior vulnerabilidade democrática. Os três marcos práticos mais críticos do calendário de 2026 foram identificados: 16 de agosto, início das obrigações de rotulagem; 1º de outubro, vedação total de novo conteúdo sintético para o primeiro turno; e 22 de outubro, a mesma vedação para o segundo turno.

QUESTÕES PARA REFLEXÃO E DEBATE

1. A Resolução TSE nº 23.732/2024 adota abordagem funcional para definir inteligência artificial, ancorando a definição no que os sistemas fazem em vez de como são construídos. Essa escolha garante longevidade normativa, mas pode gerar imprecisão na fronteira entre ferramentas de IA e softwares de edição convencional com componentes algorítmicos. Como o TSE deveria construir os parâmetros interpretativos para delimitar essa fronteira sem engessar a norma com critérios técnicos que estariam desatualizados antes do próximo ciclo eleitoral?

2. A obrigação de rotulagem tem eficácia empiricamente limitada precisamente para o segmento do eleitorado que mais precisa de proteção: eleitores com menor literacia midiática tendem a ignorar ou a não compreender os rótulos de identificação de IA. Se a rotulagem funciona melhor para quem já questiona o conteúdo, qual seria o instrumento regulatório complementar mais eficaz para alcançar os eleitores estruturalmente mais vulneráveis à desinformação sintética?

3. A vedação ampliada dos três dias pré-eleitorais, estabelecida pela Resolução TSE nº 23.760/2026, colapsa a distinção entre conteúdo irregular e conteúdo rotulado no período crítico. Esse colapso é justificável do ponto de vista normativo, mas pode ser arguido como restrição desproporcional à liberdade de expressão de candidatos que utilizam IA de forma transparente e legítima. Como o direito eleitoral deveria equilibrar a proteção da integridade eleitoral no período pré-pleito com a garantia do espaço comunicacional dos candidatos?

4. A quarta lacuna identificada no capítulo é a ausência de mecanismo de revisão periódica obrigatória da Resolução. Considerando que o TSE opera com calendário institucional próprio e que os ministros têm mandatos fixos, qual seria o formato institucional mais adequado para um ciclo de revisão normativa que inclua consulta técnica especializada, participação de stakeholders e análise comparada internacional, sem criar rigidez processual incompatível com a velocidade de evolução tecnológica?

5. A comparação internacional revelou que o modelo brasileiro é o mais específico do mundo em termos de regulação de IA eleitoral, mas que o modelo europeu do DSA é mais robusto em termos de responsabilização sistêmica das plataformas. Um modelo híbrido que combinasse a especificidade setorial brasileira com a responsabilização sistêmica europeia seria desejável? Quais seriam os obstáculos jurídicos, políticos e técnicos à sua implementação no contexto constitucional brasileiro?

6. A Resolução estabelece que a responsabilidade pela conformidade com a obrigação de rotulagem recai sobre o responsável pela propaganda, não sobre a cadeia de produção. Esse regime de responsabilidade cria incentivo estrutural para a due diligence contratual, mas pode ser insuficiente para campanhas de menor porte sem acesso a assessoria jurídica especializada. Como o direito eleitoral deveria tratar a assimetria de capacidade de conformidade entre campanhas bem financiadas e candidaturas com recursos limitados, sem criar regimes diferenciados que violem o princípio da igualdade eleitoral?

7. O pioneirismo normativo do TSE foi possível porque o modelo brasileiro de autoridade eleitoral com poder normativo próprio permite resposta sem aprovação parlamentar. Esse mesmo modelo concentra, porém, o poder de definir os limites do discurso político legítimo numa instituição não eleita diretamente. Quais mecanismos de controle democrático sobre o poder normativo do TSE em matéria de IA eleitoral seriam adequados para preservar a legitimidade das normas sem comprometer a agilidade que tornou o modelo brasileiro um referencial internacional?

8. O capítulo identifica que a lacuna técnica entre produzir e identificar conteúdo sintético tende a se ampliar com o tempo, porque detecção é um problema mais difícil do que geração. Se essa assimetria é estrutural e irredutível, a estratégia regulatória baseada em detecção de violações está condenada a ser progressivamente menos eficaz. Quais abordagens alternativas, como a proveniência criptográfica de conteúdo pelo padrão C2PA, a responsabilização dos desenvolvedores de ferramentas ou a tributação sobre engajamento, seriam mais promissoras como fundamentos de uma regulação de segunda geração para as eleições de 2030?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Três datas do calendário eleitoral de 2026 que toda campanha precisa ter na agenda agora.

16 de agosto de 2026: início da propaganda eleitoral e, simultaneamente, início da vigência plena das obrigações de rotulagem da Resolução TSE 23.732/2024 sobre todo conteúdo publicitado. 1º de outubro de 2026: vedação total de qualquer novo conteúdo sintético, mesmo rotulado, até 24 horas após o encerramento do primeiro turno. 22 de outubro de 2026: a mesma vedação para o segundo turno. Campanhas que não incorporarem essas três datas em seus protocolos internos de conformidade estarão operando com risco jurídico desnecessário nas janelas temporais mais críticas do ciclo eleitoral.

2. Você é responsável pelo que sua agência faz com IA, mesmo que ela não te conte.

A Resolução estabelece que a responsabilidade pela conformidade com a obrigação de rotulagem recai sobre o responsável pela propaganda, não sobre a cadeia de produção. Se sua agência de marketing usou IA generativa para produzir conteúdo eleitoral sem a rotulagem obrigatória e sem te informar, você responde pela omissão. A solução prática é simples: inclua nos contratos com fornecedores de tecnologia eleitoral cláusula expressa exigindo conformidade com a Resolução TSE 23.732/2024, com obrigação de notificação sobre o uso de IA em qualquer peça produzida e com responsabilização contratual do fornecedor em caso de violação.

3. A zona cinzenta do 'significativamente modificado' pode ser sua proteção ou seu risco.

A Resolução não define com precisão quando uma intervenção de IA é 'significativa' o suficiente para acionar a obrigação de rotulagem. Isso cria dois riscos simétricos: não rotular o que deveria ser rotulado, por acreditar que a intervenção foi mínima; ou rotular tudo indiscriminadamente, diluindo o sinal informacional. A estratégia mais segura para 2026 é adotar o critério material que a própria Resolução sugere: se o conteúdo tem capacidade de induzir o eleitor a acreditar que está diante de um registro autêntico quando não está, rotule. Se tem dúvida, rotule. O custo da rotulagem desnecessária é zero. O custo da omissão de rotulagem pode chegar a R$ 30.000 por peça.

4. Chatbot de campanha: oportunidade estratégica com risco jurídico alto se mal implementado.

Chatbots eleitorais são permitidos pela Resolução, mas com obrigação de identificação imediata como sistema automatizado no início de cada conversa, sob pena de multa de R$ 10.000 a R$ 50.000 e cessação imediata do funcionamento. Essa multa é substancialmente maior do que as aplicáveis a outras formas de conteúdo sem rotulagem, o que reflete a avaliação do TSE de que simular interação humana é a forma mais grave de engano. Um chatbot bem implementado, com identificação clara e linguagem adequada, é um instrumento poderoso de escala de atendimento eleitoral. Um chatbot não identificado é um passivo jurídico e reputacional que pode comprometer toda a campanha.

5. Conteúdo de pré-campanha gerado por IA já está sujeito à Resolução.

A extensão funcional das obrigações ao período de pré-campanha significa que todo conteúdo produzido com auxílio de IA para fins políticos, mesmo antes do início formal da campanha, está sujeito à obrigação de rotulagem e às vedações absolutas da Resolução. O critério é a finalidade eleitoral do conteúdo, não a data de sua produção. Um vídeo gerado com IA em janeiro de 2026, antes do período oficial, promovendo implicitamente a candidatura, está sujeito à norma. A estratégia mais segura é tratar todas as obrigações da Resolução como vigentes desde o momento em que seu cliente passa a ter interesse eleitoral identificável.

6. Deepfake de adversário é crime eleitoral, não apenas irregularidade.

A criação de deepfake de candidato adversário fazendo declaração falsa não configura apenas violação da Resolução TSE 23.732/2024: configura, em cumulação, crime eleitoral de calúnia previsto no artigo 324 do Código Eleitoral. As sanções são cumulativas: multa de R$ 20.000 a R$ 100.000 por peça mais responsabilização penal eleitoral. A vedação é absoluta: não existe contextualização, intenção declarada ou argumento de boa-fé que afaste a tipificação. E vale para qualquer ciclo eleitoral, incluindo o período de pré-campanha. Campanhas que produzem ou encomendam deepfakes de adversários estão expondo seus responsáveis a processo criminal, não apenas a multa administrativa.

7. A Resolução protege você tanto quanto restringe você. Use isso estrategicamente.

As vedações absolutas da Resolução funcionam nos dois sentidos: sua campanha não pode produzir deepfakes de adversários, mas adversários também não podem produzir deepfakes de você. O regime de responsabilidade solidária inclui as plataformas, que devem remover conteúdo irregular em 24 horas sob pena de multa diária de R$ 150.000. Isso significa que, se sua campanha identificar deepfakes ou conteúdo sintético irregular circulando contra você, tem instrumentos jurídicos efetivos para exigir remoção rápida. A condição é ter a infraestrutura de monitoramento pronta antes do ataque, não depois. Campanha que chega a 2026 sem protocolo de monitoramento de deepfakes e de acionamento emergencial da Justiça Eleitoral não está usando os instrumentos que a Resolução criou.

CAPÍTULO 12

Crimes Eleitorais Digitais e Sanções

Tipificação, Responsabilização e a Jurisprudência Emergente do TSE

Responsabilização Eleitoral na Era Digital

A questão da responsabilização por condutas eleitorais ilícitas praticadas com o auxílio de tecnologias digitais e de inteligência artificial constitui um dos campos de maior fermentação doutrinária e jurisprudencial no direito eleitoral contemporâneo, e sua relevância prática aumenta a cada ciclo eleitoral na proporção em que essas tecnologias se tornam acessíveis a um número crescente de atores com motivação e recursos para usá-las irregularmente.

O problema da responsabilização na era digital não é inteiramente novo. O direito eleitoral já havia enfrentado, em versões anteriores, o desafio de aplicar tipos normativos construídos para um ambiente analógico a condutas mediadas por tecnologia: a jurisprudência sobre propaganda eleitoral na internet, o regime de impulsionamento pago estabelecido pela Resolução TSE nº 23.610/2019 e as normas sobre disparo em massa do artigo 222-A da Lei nº 9.504/1997 são produtos de ciclos anteriores de adaptação. O que a inteligência artificial generativa introduz é uma descontinuidade qualitativa, não apenas quantitativa: ela não apenas acelera e escala condutas que já existiam, mas cria condutas inteiramente novas, como a fabricação de declarações falsas com aparência de autenticidade, para as quais os tipos normativos anteriores foram construídos sem qualquer antecipação.

A literatura especializada tem enfrentado esse problema a partir de dois ângulos complementares que estruturam o debate doutrinário atual. O primeiro é o ângulo da tipificação, que examina em que medida as condutas habilitadas pela IA se enquadram nos tipos penais e infracionais eleitorais existentes. O segundo é o ângulo da atribuição, que examina a quem deve ser imputada a responsabilidade quando o conteúdo irregular foi produzido por um sistema algorítmico, frequentemente sem intervenção humana direta no ato específico de produção.

Sobre o ângulo da tipificação, Gomes (2023), em *Direito Eleitoral*, obra de referência que sintetiza décadas de doutrina e jurisprudência no campo, argumenta que o princípio da legalidade estrita em matéria penal eleitoral exige que novas condutas tecnológicas sejam

enquadradas nos tipos já existentes com cautela hermenêutica redobrada. O argumento é de base constitucional: a vedação à analogia *in malam partem* em direito penal, derivada do artigo 5º, inciso XXXIX da Constituição Federal, impede que o intérprete amplie o alcance de um tipo penal para abranger condutas que, ao tempo de sua edição, o legislador não antecipava e não pretendia criminalizar. Aplicar o tipo penal de calúnia eleitoral do artigo 324 do Código Eleitoral à criação de deepfakes eleitorais é hermeneuticamente defensável, porque o núcleo da ilicitude, a imputação falsa de fato definido como crime, está presente em ambas as situações. Mas enquadrar a desinformação estrutural de 2022, a narrativa de fraude nas urnas, como crime eleitoral pela via da analogia é problemático, porque essa conduta não se ajusta com precisão a nenhum tipo penal existente (GOMES, 2023, p. 387-392).

Cândido (2022), em posição doutrinária que ganhou relevância nos debates sobre a Resolução TSE nº 23.732/2024, defende perspectiva distinta sobre o eixo da tipificação. Para o autor, a interpretação evolutiva dos tipos eleitorais não constitui analogia vedada quando o elemento normativo central do tipo, o dolo de enganar o eleitor ou de comprometer a lisura do processo, está presente na conduta nova. Nessa leitura, a proibição constitucional alcança a criação de novos tipos por analogia, não a interpretação extensiva de tipos existentes à luz de novos meios de execução. Um deepfake que coloca em candidato adversário declarações que nunca proferiu preenche o elemento objetivo do crime de calúnia eleitoral, a imputação de fato falso definido como crime, independentemente do método tecnológico pelo qual a falsidade foi produzida. O método é circunstância do crime, não elemento constitutivo do tipo (CÂNDIDO, 2022, p. 445-449).

A tensão entre essas duas posições doutrinárias não é meramente acadêmica: ela tem consequências diretas sobre a previsibilidade das decisões judiciais eleitorais e sobre a estratégia de defesa em casos concretos. Quando a Justiça Eleitoral aplica o tipo de calúnia eleitoral a um deepfake, está implicitamente adotando a posição de Cândido. Quando recorre à analogia com propaganda eleitoral irregular para enquadrar condutas que nenhum tipo específico captura, está operando num terreno hermeneuticamente mais frágil que potencialmente invalida a decisão em grau recursal. A Resolução TSE nº 23.732/2024 procurou resolver essa tensão estrutural ao criar tipos infracionais administrativos específicos para as condutas mais problemáticas, complementando os tipos penais do Código Eleitoral com um regime sancionatório próprio que não está sujeito às restrições da legalidade penal estrita. Esse movimento é analiticamente relevante: ao criar infrações administrativas eleitorais específicas para o uso irregular de IA, o TSE deslocou parte da responsabilização do campo penal, onde a legalidade estrita impõe restrições

hermenêuticas significativas, para o campo administrativo-eleitoral, onde a interpretação evolutiva encontra menor resistência formal (LEONARDI, 2012, p. 203-207).

Sobre o ângulo da atribuição, o problema é ainda mais complexo e menos resolvido pela doutrina disponível. Quando um candidato contrata uma agência de marketing que utiliza IA generativa para produzir conteúdo irregular, sem que o candidato tenha autorizado especificamente essa modalidade de produção, a quem deve ser imputada a responsabilidade? A resposta do direito eleitoral tradicional, que atribui ao candidato responsabilidade por todo conteúdo veiculado em seu nome, tem fundamento na lógica de que a campanha é uma unidade de responsabilidade: quem aufere o benefício eleitoral do conteúdo responde por ele, independentemente de ter sido o autor material. Essa lógica é compatível com a estrutura da Resolução TSE nº 23.732/2024, que estabelece responsabilidade solidária entre candidatos, prestadores de serviço e plataformas, mas levanta questões sobre equidade quando a campanha não tinha capacidade técnica de verificar os métodos utilizados pela agência contratada.

Howard (2020), ao analisar os padrões de responsabilização em operações de desinformação documentadas internacionalmente, observa que a dissociação deliberada entre financiadores, produtores e distribuidores é frequentemente construída precisamente para explorar essa dificuldade de atribuição: quando ninguém pode ser inequivocamente responsabilizado por todo o nexo causal entre a conduta e o dano, o sistema sancionatório perde eficácia deterrente (HOWARD, 2020, p. 118-122). A resposta regulatória mais eficaz para esse problema, como demonstrado no capítulo anterior, é o modelo de responsabilidade compartilhada em cadeia, que distribui a responsabilidade proporcionalmente à contribuição causal de cada ator, sem exigir que um único ator responda por toda a operação.

O Sistema Sancionatório Eleitoral: Três Camadas

O sistema sancionatório aplicável ao uso irregular de IA em eleições não é monolítico: é construído pela articulação de três camadas distintas e cumulativamente aplicáveis, cada uma com pressuposto, procedimento e consequência distintos, que podem ser acionadas simultaneamente a partir de um mesmo fato ilícito. Compreender a arquitetura dessas três camadas é condição necessária tanto para a defesa jurídica de candidatos acusados de violações quanto para a estratégia de responsabilização de adversários que as praticaram.

A primeira camada é a penal eleitoral, que incide quando a conduta praticada por meio de IA preenche os elementos constitutivos de um tipo penal previsto no Código Eleitoral ou na legislação penal especial aplicável ao contexto eleitoral. Os tipos mais relevantes para o universo das condutas habilitadas por IA são a calúnia eleitoral do artigo 324, que pune a imputação falsa de fato definido como crime a candidato ou partido com pena de detenção de seis meses a dois anos mais multa; a difamação eleitoral do artigo 325, que pune a imputação de fato ofensivo à reputação sem necessidade de que o fato seja definido como crime; e o uso indevido de meios de comunicação do artigo 57-H da Lei nº 9.504/1997, que pune condutas de propaganda eleitoral irregular realizadas por meios eletrônicos. A responsabilização penal eleitoral exige dolo específico, prova do nexo causal entre a conduta e o resultado e observância do devido processo legal penal, com as garantias constitucionais que isso implica. Sua vantagem é a força simbólica e deterrente da sanção criminal; sua limitação é a dificuldade probatória num ambiente em que a atribuição de autoria a conteúdo produzido por IA é tecnicamente complexa e frequentemente inconclusiva no tempo útil do processo eleitoral.

A segunda camada é a infracionaleleitoral administrativa, que é a que mais diretamente se relaciona com o marco normativo da Resolução TSE nº 23.732/2024. Essa camada é acionada quando a conduta viola as normas eleitorais sem necessariamente configurar crime, sujeitando o responsável às sanções administrativas previstas na Resolução e na legislação eleitoral geral: multas, cassação de registro ou diploma, direito de resposta e responsabilidade solidária das plataformas. A distinção fundamental em relação à camada penal é o padrão probatório: a infração administrativa eleitoral pode ser reconhecida com base em preponderância de evidências, sem a necessidade de prova além de dúvida razoável que o processo penal exige. Isso torna a responsabilização administrativa estruturalmente mais acessível no tempo útil do processo eleitoral, especialmente considerando que a Resolução nº 23.732/2024 não exige a demonstração de dano efetivo para a maioria de suas vedações absolutas, bastando o potencial de engano. Como argumenta Salgado (2015), a construção de um sistema sancionatório eleitoral robusto depende precisamente da articulação entre a camada penal, com sua força simbólica, e a camada administrativa, com sua agilidade procedimental (SALGADO, 2015, p. 298-302).

A terceira camada é a cível eleitoral, que incide quando a conduta irregular causa dano mensurável a candidatos, partidos ou ao próprio processo eleitoral, gerando obrigação de reparação. Essa camada é acionada por meio de ações de responsabilidade civil eleitorais que podem ser cumuladas com as representações administrativas e com as

denúncias criminais. Sua relevância cresceu com o regime de responsabilidade solidária estabelecido pela Resolução TSE nº 23.732/2024: ao tornar as plataformas digitais solidariamente responsáveis pelos danos decorrentes de conteúdo mantido após notificação, a norma abriu o caminho para que candidatos e partidos busquem reparação econômica não apenas dos produtores do conteúdo irregular, mas também das plataformas que o mantiveram em circulação após serem notificadas. Para operações de desinformação conduzidas por atores que não têm patrimônio identificável no Brasil, a responsabilização das plataformas pode ser o único mecanismo de reparação economicamente efetivo disponível.

A articulação das três camadas no mesmo caso concreto não é apenas possível: é frequentemente necessária para uma resposta proporcional e eficaz. Um deepfake de candidato adversário disseminado nos dias antes do pleito pode dar origem simultaneamente a representação administrativa perante o TSE pelo descumprimento das vedações absolutas da Resolução, com pedido de liminar de remoção e aplicação de multa; a queixa-crime por calúnia eleitoral perante a Justiça Eleitoral, com a força simbólica e deterrente que a responsabilização criminal carrega; e a ação de responsabilidade civil contra os produtores identificados e contra as plataformas que mantiveram o conteúdo após notificação, com pedido de reparação pelos danos causados à imagem e ao resultado eleitoral do candidato afetado. A estratégia de acionamento simultâneo dessas três camadas é mais eficaz do que a escolha entre elas, precisamente porque cada uma alcança dimensões do problema que as outras não cobrem integralmente.

O quadro a seguir sistematiza as três camadas sancionatórias em perspectiva comparada, permitindo visualizar com precisão as diferenças estruturais entre elas que determinam a estratégia de acionamento em cada caso concreto. A leitura por colunas é mais reveladora do que a leitura por linhas: o que mais distingue as três camadas não é a gravidade das sanções em abstrato, mas o prazo para acionamento, o padrão probatório exigido e o procedimento disponível.

A camada administrativa eleitoral é a mais ágil e a mais acessível probatoriamente, o que a torna o instrumento preferencial para a resposta em tempo real durante o período eleitoral. A camada cível eleitoral, com seus instrumentos de cassação de registro e diploma, é a que produz consequências mais definitivas sobre o resultado eleitoral, mas exige prazos e procedimentos que raramente se completam antes do pleito. A camada penal eleitoral é a de maior força simbólica e deterrente, mas a que mais frequentemente chega depois do dano, pela combinação de padrão probatório elevado, prazo prescricional de quatro anos e estrutura processual que não foi desenhada para a urgência do calendário

eleitoral. A estratégia jurídica mais eficaz diante de uso irregular de IA eleitoral não é escolher entre essas três camadas: é acionar simultaneamente as que forem pertinentes ao caso, extraindo de cada uma o que ela oferece de melhor, que na prática significa a liminar administrativa para remoção imediata, a representação cível para pressão sobre o resultado eleitoral e o registro criminal para deterrência de longo prazo.

Tabela 40. O Sistema Sancionatório Eleitoral Aplicável à IA

Camada	Fundamento	Sanções Típicas	Procedimento	Prazo Prescricional
Administrativa Eleitoral	Resolução TSE 23.732/2024; Lei 9.504/1997	Multa (R$ 5.000 a R$ 200.000); remoção de conteúdo; suspensão de impulsionamento; direito de resposta.	Representação ao TSE/TRE; julgamento liminar possível; recurso ao TSE.	1 ano após a eleição (regra geral).
Eleitoral Cível	Lei 9.504/1997; LC 64/1990	Cassação de registro/diploma; inelegibilidade por abuso de poder.	AIJE (Ação de Investigação Judicial Eleitoral); AIME; RCED.	Até a diplomação (AIJE); 15 dias após diplomação (AIME).
Penal Eleitoral	Código Eleitoral (arts. 289-354); Lei 14.197/2021	Detenção (6 meses a 5 anos, conforme o tipo); multa cumulativa.	Ação penal proposta pelo MP Eleitoral; TRE em primeira instância.	4 anos (crimes eleitorais em geral).

Fonte: Elaborado pelo autor com base em Gomes (2023), Cândido (2022), CE e Resolução TSE 23.732/2024.

Tipificação Penal: Crimes Eleitorais Digitais Aplicáveis

O Código Eleitoral, editado em 1965 num contexto em que a comunicação política era inteiramente analógica e em que a manipulação de imagem e voz exigia recursos técnicos e humanos que poucos atores poderiam mobilizar, não contempla tipos penais específicos para crimes cometidos com o auxílio de inteligência artificial ou de tecnologias digitais de desinformação. Essa ausência não é uma falha que passou despercebida: ela reflete a inevitável limitação temporal de qualquer codificação, que captura o estado das práticas ao tempo de sua edição sem capacidade de antecipar as transformações tecnológicas subsequentes.

A ausência de tipos penais específicos para crimes eleitorais digitais cria o problema hermenêutico central que a doutrina e a jurisprudência têm enfrentado nos últimos ciclos eleitorais: em que medida os tipos penais eleitorais clássicos alcançam condutas mediadas por IA sem violação do princípio constitucional da legalidade estrita? A resposta não é uniforme e depende da análise específica de cada tipo em relação a cada conduta concreta. Como ponto de partida, é necessário distinguir dois planos que o debate público frequentemente confunde: o plano dos elementos do tipo, que pergunta se a conduta preenche os requisitos objetivos e subjetivos do tipo penal independentemente do meio utilizado; e o plano do meio de execução, que pergunta se o uso de IA ou de tecnologia digital constitui circunstância agravante, elemento qualificador ou dado irrelevante para a tipificação.

A hermenêutica judicial tem progressivamente construído a ponte entre os tipos clássicos e as condutas tecnológicas contemporâneas a partir de um critério que Cândido (2022) denomina de invariância do núcleo da ilicitude: o que o tipo penal protege, a honra do candidato, a lisura do processo, a autenticidade do debate eleitoral, permanece protegido independentemente do método pelo qual a violação é praticada, e a tipificação deve seguir essa invariância (CÂNDIDO, 2022, p. 447). A Lei nº 14.197/2021, que introduziu o crime de violência política no ordenamento jurídico brasileiro e atualizou parcialmente os crimes contra o Estado Democrático de Direito, representou um passo relevante nessa direção ao reconhecer que condutas que afetam a integridade do processo eleitoral podem ser praticadas por meios digitais sem que isso exclua a tipicidade.

Os tipos penais eleitorais com maior relevância para o universo de condutas habilitadas por IA generativa são os seguintes, analisados tanto em seus elementos constitutivos quanto em sua aplicabilidade específica ao contexto digital.

A calúnia eleitoral, tipificada no artigo 324 do Código Eleitoral com pena de detenção de seis meses a dois anos mais multa, pune quem calunia alguém na propaganda eleitoral, imputando-lhe falsamente fato definido como crime. É o tipo mais diretamente aplicável à criação de deepfakes eleitorais que fazem candidatos aparentarem confessar irregularidades, cometer crimes ou expressar posições criminosas que nunca manifestaram. O elemento central do tipo, a imputação falsa de fato definido como crime, está presente com precisão nos deepfakes que colocam na boca de candidatos confissões de corrupção, coordenação com organizações criminosas ou prática de outros ilícitos. A questão hermenêutica relevante é se o deepfake preenche o elemento de "imputação": como argumenta Gomes (2023), imputar pressupõe atribuir a alguém uma autoria que não lhe pertence, e o deepfake é precisamente um instrumento de atribuição falsa de autoria,

tornando a aplicação do tipo não apenas defensável mas hermeneuticamente natural (GOMES, 2023, p. 389). A ação penal é pública condicionada à representação, o que significa que o candidato afetado precisa manifestar interesse na persecução penal, criando um filtro estratégico que pode ser relevante para campanhas que preferem não amplificar o alcance do conteúdo irregular por meio do processo criminal.

A difamação eleitoral, tipificada no artigo 325 do Código Eleitoral com pena de detenção de três meses a um ano mais multa, pune quem difama candidato na propaganda eleitoral, imputando-lhe fato ofensivo à reputação, sem necessidade de que esse fato seja definido como crime. Tem espectro de aplicação mais amplo do que a calúnia porque não exige que o fato imputado seja criminoso: basta que seja ofensivo à reputação e falso. Deepfakes que fazem candidatos aparentarem comportamento moral ou pessoal reprovável, como imagens sintéticas em situações comprometedoras que nunca ocorreram, enquadram-se com precisão nesse tipo. A distinção prática entre calúnia e difamação eleitoral é relevante para a estratégia processual: a difamação tem pena menor, mas é mais fácil de provar porque dispensa a demonstração de que o fato imputado é definido como crime.

A injúria eleitoral, tipificada no artigo 326 do Código Eleitoral com pena de detenção de um a seis meses mais multa, pune quem injuria candidato na propaganda eleitoral, ofendendo-lhe a dignidade ou o decoro. Distingue-se da difamação por não exigir a imputação de fato específico: basta a ofensa à dignidade por meio de xingamento, humilhação ou depreciação. Conteúdos gerados por IA que ridicularizam candidatos por características pessoais, associam sua imagem a símbolos pejorativos ou constroem representações deliberadamente degradantes podem preencher esse tipo, embora a linha entre crítica política legítima, que goza de proteção constitucional ampla no contexto eleitoral, e injúria eleitoral seja frequentemente objeto de litígio. A jurisprudência do TSE tem adotado critério de que o conteúdo satírico claramente identificado como tal, mesmo que ácido e exagerado, não configura injúria eleitoral, o que alinha a aplicação do tipo com a exclusão expressa da sátira identificada prevista na Resolução TSE nº 23.732/2024.

O crime de falsa identidade eleitoral e uso indevido de meios de comunicação, cujas bases legais estão no artigo 57-H da Lei nº 9.504/1997 e em dispositivos complementares, pune condutas de propaganda eleitoral irregular realizadas por meios eletrônicos, incluindo a criação de perfis falsos, a simulação de identidades e a distribuição de conteúdo eleitoral por meios não autorizados. É o tipo mais diretamente voltado para operações coordenadas de influência que utilizam contas automatizadas, personas sintéticas e redes de amplificação artificial. Sua aplicação a chatbots eleitorais não identificados como sistemas

automatizados, que simulam interação humana em violação ao artigo 15 da Resolução TSE nº 23.732/2024, é hermeneuticamente defensável: o chatbot não identificado está operando sob falsa identidade, atribuindo ao sistema automatizado a aparência de interação humana que o eleitor não autorizou e não foi informado de estar recebendo.

O crime de disparo em massa irregular, tipificado no artigo 222-A da Lei nº 9.504/1997 com a redação dada pela Lei nº 14.197/2021, pune o envio em lote de mensagens eletrônicas para fins eleitorais sem consentimento dos destinatários. Com a introdução da IA generativa no ciclo de produção de desinformação, esse tipo ganhou relevância adicional: operações que antes exigiam conteúdo uniforme em escala passaram a poder personalizar cada mensagem individualmente, adaptando o conteúdo ao perfil do receptor sem alterar a estrutura de disparo em massa que o tipo visa coibir. A personalização por IA não descaracteriza o disparo em massa: ela o torna mais eficaz sem alterar sua natureza jurídica.

A correta identificação do tipo aplicável em cada caso não é apenas exercício de precisão jurídica: tem consequências práticas sobre a competência, o procedimento, o prazo prescricional e as sanções disponíveis. Um deepfake que atribui a candidato declaração falsa de ter praticado crime de corrupção é simultaneamente calúnia eleitoral, uso indevido de meios de comunicação e, se disseminado por disparo em massa, violação do artigo 222-A. A cumulação de tipos é juridicamente possível quando as condutas têm elementos distintos e tutelam bens jurídicos diferentes, e a estratégia de denúncia que articula múltiplos tipos aumenta a probabilidade de que pelo menos um deles seja reconhecido pelo julgador, mesmo que outros sejam afastados por insuficiência probatória ou por divergência hermenêutica.

O quadro a seguir sistematiza os cinco tipos penais eleitorais com maior relevância para o universo de condutas habilitadas por inteligência artificial, organizados por base legal, conduta correspondente, pena e elemento subjetivo exigido. A coluna do elemento subjetivo é a mais relevante para a estratégia processual: em todos os tipos, o dolo direto é exigido, o que significa que a responsabilização penal depende da demonstração de que o agente conhecia a falsidade do conteúdo e agiu com intenção específica de causar o dano correspondente. Essa exigência cria o principal obstáculo probatório da camada penal em contexto de IA: quando o conteúdo irregular foi produzido por um sistema algorítmico mediante instruções de um operador humano, a prova do dolo exige demonstrar não apenas que o conteúdo falso foi produzido e disseminado, mas que o responsável conhecia sua falsidade e atuou deliberadamente para que ela causasse o dano tipificado. A leitura da coluna de penas revela também a progressão da gravidade: os crimes contra o Estado

Democrático de Direito, introduzidos pela Lei nº 14.197/2021 especificamente para endereçar ataques sistêmicos à legitimidade do processo eleitoral, têm pena de reclusão de dois a oito anos, substancialmente superior à dos crimes contra a honra eleitoral individual, o que reflete a avaliação legislativa de que a desinformação estrutural que ataca o processo como um todo é mais grave do que a desinformação direcionada a candidatos específicos.

Tabela 41. Tipos Penais Eleitorais Aplicáveis a Condutas com IA

Tipo Penal	Base Legal	Conduta com IA Correspondente	Pena	Elemento Subjetivo Exigido
Calúnia eleitoral	Art. 324 do CE	Criação de deepfake ou geração de texto por IA atribuindo falsamente crime a candidato.	Detenção de 6 meses a 2 anos e multa.	Dolo direto: intenção de ofender a honra imputando fato criminoso falso.
Difamação eleitoral	Art. 325 do CE	Uso de IA para gerar conteúdo que atribui fato desonroso falso a candidato.	Detenção de 3 a 12 meses e multa.	Dolo direto: intenção de ofender a reputação.
Injúria eleitoral	Art. 326 do CE	Uso de IA para gerar conteúdo que atinge a dignidade ou decoro de candidato.	Detenção de 1 a 6 meses e multa.	Dolo direto: animus injuriandi.
Propaganda enganosa	Art. 323 do CE	Geração de IA de declaração falsa atribuída a candidato (deepfake declarativo).	Detenção de 6 meses a 1 ano.	Dolo: conhecimento da falsidade.
Crimes contra o Estado Democrático	Arts. 359-L a 359-S da Lei 14.197/2021	Campanha sistemática de desinformação gerada por IA visando deslegitimar o processo eleitoral.	Reclusão de 2 a 8 anos.	Dolo direto: finalidade específica de atacar o Estado Democrático.

Fonte: Elaborado pelo autor com base no Código Eleitoral (1965), Lei 14.197/2021 e jurisprudência do TSE.

Abuso de Poder Político e Econômico com IA

O abuso de poder constitui, ao lado dos crimes eleitorais, um dos instrumentos centrais de responsabilização no direito eleitoral brasileiro, e sua relevância para o contexto da inteligência artificial eleitoral é particularmente significativa por uma razão estrutural:

enquanto a responsabilização penal depende da demonstração de dolo específico, exigência que como analisado na seção anterior cria obstáculos probatórios relevantes em contexto de IA, o abuso de poder tem natureza predominantemente objetiva em sua dimensão eleitoral. O que se exige provar não é a intenção do agente, mas o efeito potencial de sua conduta sobre a igualdade da disputa eleitoral.

Essa distinção não é apenas técnica: ela tem implicações estratégicas diretas para a escolha dos instrumentos de responsabilização em cada caso. Quando a prova do dolo é difícil, como frequentemente ocorre em operações de desinformação que utilizam intermediários e infraestrutura tecnológica para criar distância entre o mandante e o conteúdo produzido, o abuso de poder pode ser o caminho mais acessível para responsabilização efetiva. Como argumenta Salgado (2015), o conceito de abuso de poder no direito eleitoral brasileiro foi progressivamente construído pela jurisprudência do TSE com uma lógica que privilegia a proteção da igualdade de chances na competição eleitoral sobre a punição de intenções específicas: o que importa é se a conduta distorceu as condições da disputa, não se o agente pretendia distorcê-las (SALGADO, 2015, p. 312-316).

O fundamento normativo do abuso de poder está na Lei Complementar nº 64/1990, artigos 19 a 25, que disciplina a Ação de Investigação Judicial Eleitoral, e no artigo 22 da mesma lei, que estabelece o rito processual para apuração de abuso do poder econômico e uso indevido dos meios de comunicação. A jurisprudência do TSE consolidou, ao longo de décadas, os pressupostos para a configuração do abuso: a conduta deve ser grave, o que significa que deve superar o patamar da mera irregularidade para atingir o de desequilíbrio relevante na competição; deve ser generalizada, o que significa que deve ter alcance suficiente para influenciar o resultado eleitoral ou sua percepção pelo eleitorado; e deve ter nexo de causalidade com a potencial alteração do resultado, embora a jurisprudência mais recente tenha flexibilizado esse último requisito para situações em que a prova da influência efetiva é estruturalmente impossível (TSE, REspe 12.677/SP, 2016).

No contexto da inteligência artificial eleitoral, o abuso de poder político manifesta-se primariamente em três modalidades que a prática dos ciclos eleitorais recentes permite identificar com precisão crescente. A primeira é o uso de infraestrutura governamental para produção ou distribuição de conteúdo de IA em benefício de candidato: quando servidores públicos, equipamentos ou sistemas governamentais são utilizados para gerar, personalizar ou disseminar conteúdo eleitoral produzido por modelos de linguagem ou outras ferramentas de IA, há desvio de recursos públicos em benefício de candidatura específica que configura abuso de poder político independentemente da demonstração de intenção

específica do agente. A prova dessa modalidade é relativamente acessível quando o uso da infraestrutura governamental deixa rastros em sistemas de logs, registros de acesso e dados de produção que investigações técnicas podem recuperar.

A segunda modalidade é o uso de dados coletados por órgão público para orientação de microtargeting eleitoral. Quando dados pessoais de cidadãos coletados por órgãos governamentais no exercício de suas funções, como registros de beneficiários de programas sociais, dados de saúde, histórico tributário ou informações cadastrais, são utilizados para direcionar propaganda eleitoral personalizada por IA, há violação simultânea da Lei Geral de Proteção de Dados, da legislação eleitoral sobre uso de dados em campanha e da vedação constitucional ao desvio de finalidade no uso de dados públicos. Essa modalidade é analiticamente relevante porque o microtargeting eleitoral orientado por dados públicos produz uma assimetria de informação estrutural entre o incumbente, que tem acesso privilegiado a dados sobre o eleitorado através de sua posição governamental, e os demais candidatos, que dependem de dados comercialmente adquiridos ou produzidos pela própria campanha. Essa assimetria é precisamente o tipo de distorção da igualdade de chances que a doutrina do abuso de poder político foi desenvolvida para combater.

A terceira modalidade é a utilização de IA para disseminação coordenada de ataques falsos a adversários com recursos do erário. Quando campanhas de desinformação contra adversários políticos são financiadas, ainda que indiretamente, por recursos públicos, e quando essa conexão pode ser demonstrada por meio de análise de fluxo financeiro, contratação de serviços com verbas públicas ou uso de infraestrutura governamental de comunicação, há abuso de poder político que pode ser apurado pela via da Ação de Investigação Judicial Eleitoral. A dificuldade dessa modalidade é probatória: a dissociação deliberada entre a origem dos recursos e o destino das operações de desinformação, que Howard (2020) identificou como característica estrutural das operações de desinformação sofisticadas, é especialmente eficaz em dificultar a demonstração do nexo entre recursos públicos e conteúdo produzido (HOWARD, 2020, p. 119-121).

O abuso de poder econômico, por sua vez, é configurado quando o uso de IA em campanha é financiado por recursos que não aparecem na prestação de contas eleitoral obrigatória ou que excedem os limites legais de gastos estabelecidos pelo TSE para cada cargo. O caso paradigmático dos disparos em massa de 2018, analisado no Capítulo 9, é o modelo mais documentado dessa modalidade: grupos de empresários financiaram operações de desinformação por valores estimados em dezenas de milhões de reais sem que esses recursos fossem contabilizados como despesas de campanha, configurando abuso de poder econômico por dupla violação, o excesso de gasto e a omissão contábil.

Com a IA generativa, o abuso de poder econômico adquire uma dimensão adicional que a doutrina ainda está elaborando. O custo de produção de conteúdo político em escala caiu dramaticamente com a democratização das ferramentas: operações que antes exigiam dezenas de funcionários e orçamentos de centenas de milhares de reais podem hoje ser executadas por equipes pequenas com assinaturas de ferramentas que custam centenas de reais por mês. Isso cria uma assimetria entre o valor econômico do impacto eleitoral gerado e o custo registrado na prestação de contas: uma campanha que utiliza IA generativa para produzir e distribuir conteúdo equivalente a uma operação de publicidade tradicional de grande escala, mas declara apenas os custos das assinaturas de software, pode estar gerando impacto eleitoral desproporcional ao que aparece nos registros oficiais de gastos. A questão regulatória que ainda não tem resposta consolidada é se o valor a ser declarado na prestação de contas deve corresponder ao custo efetivo das ferramentas utilizadas ou ao valor de mercado equivalente do impacto produzido. A segunda interpretação seria mais adequada para preservar a função protetora dos limites de gastos, mas encontra dificuldades práticas de mensuração que a regulação vigente ainda não endereçou.

A sanção pelo abuso de poder, tanto político quanto econômico, pode incluir a cassação do registro da candidatura antes da eleição ou a cassação do diploma após o pleito, além da declaração de inelegibilidade pelo prazo de oito anos previsto na Lei Complementar nº 64/1990. Essas são as sanções eleitorais de maior impacto sobre a trajetória política do responsável, e sua possibilidade de aplicação em decorrência do uso irregular de IA constitui o argumento deterrente mais poderoso disponível no sistema jurídico eleitoral brasileiro para influenciar o comportamento de candidatos e campanhas.

Responsabilização de Candidatos, Partidos e Plataformas

O regime de responsabilização no direito eleitoral brasileiro é construído sobre um princípio que a jurisprudência do TSE consolidou ao longo de décadas e que a introdução da IA generativa torna simultaneamente mais relevante e mais complexo de aplicar: o princípio da responsabilidade do candidato pelo conteúdo de campanha veiculado em seu nome, com progressiva extensão desse princípio à cadeia de atores que participam da produção e distribuição desse conteúdo. Esse princípio tem fundamento lógico na estrutura da competição eleitoral: quem aufere o benefício político do conteúdo deve responder por ele, independentemente de ter sido o autor material, porque a alternativa, permitir que

candidatos se beneficiem de irregularidades praticadas por terceiros com a proteção do argumento do desconhecimento, tornaria o sistema sancionatório inoperante.

A responsabilidade do candidato pelo conteúdo gerado com o auxílio de IA apresenta uma dualidade normativa que o operador do direito eleitoral precisa compreender com precisão para definir a estratégia processual adequada a cada caso. Em sua dimensão penal, essa responsabilidade é subjetiva: a aplicação de tipos como calúnia eleitoral, difamação ou uso indevido de meios de comunicação ao candidato exige a demonstração de que ele conhecia o conteúdo irregular, autorizou sua produção ou participou de sua concepção, não sendo suficiente a mera veiculação em seu nome por terceiro sem seu conhecimento. Em sua dimensão administrativa eleitoral, a responsabilidade é objetiva: basta demonstrar que o conteúdo irregular foi produzido ou distribuído por agente que integrava ou prestava serviço à campanha, independentemente de o candidato ter conhecimento ou autorizado especificamente a conduta. Essa dualidade não é uma contradição do sistema: ela reflete escolhas distintas sobre a intensidade da responsabilização em campos com consequências distintas. A sanção penal, por sua gravidade, exige o elemento subjetivo como condição de legitimidade constitucional. A sanção administrativa eleitoral, voltada primariamente para a preservação da igualdade da disputa e não para a punição do agente, pode ser aplicada com base em nexo objetivo entre a conduta e a campanha beneficiada.

A Resolução TSE nº 23.732/2024 consolidou esse regime ao estabelecer expressamente que candidatos são responsáveis pelo conteúdo gerado por agentes automatizados utilizados em suas campanhas, mesmo quando operados por terceiros contratados. Essa disposição tem três implicações práticas que merecem desenvolvimento. A primeira é que o candidato não pode invocar o desconhecimento sobre os métodos utilizados pela agência contratada como excludente de responsabilidade administrativa: a due diligence sobre os fornecedores de tecnologia eleitoral é uma obrigação de cuidado que o regime de responsabilidade pressupõe, não uma faculdade. A segunda é que os contratos com agências de marketing eleitoral e fornecedores de tecnologia precisam incluir cláusulas específicas sobre uso de IA e rotulagem obrigatória, não apenas como proteção contratual do candidato, mas como instrumento de governança que demonstra a adoção de medidas razoáveis de conformidade. A terceira é que, em caso de violação, o candidato pode ter ação regressiva contra o fornecedor que utilizou IA de forma irregular sem informá-lo, mas essa ação regressiva não afasta a responsabilidade eleitoral perante a Justiça Eleitoral: ela apenas redistribui os custos entre os atores da cadeia após a aplicação da sanção.

A responsabilidade dos partidos políticos pelo conteúdo de IA produzido em campanha tem fundamento no artigo 6º da Lei nº 9.504/1997, que estabelece a responsabilidade das agremiações pelos atos de seus órgãos de direção e de candidatos que integram suas listas, e foi progressivamente expandida pela jurisprudência do TSE para alcançar atos praticados por agentes que atuam em nome do partido mesmo sem vínculo formal de mandato. No contexto da IA eleitoral, a responsabilidade partidária é especialmente relevante em duas situações. A primeira é a das campanhas de desinformação que operam com infraestrutura compartilhada entre múltiplos candidatos do mesmo partido: quando o conteúdo irregular é produzido de forma centralizada pelo partido e distribuído para as campanhas individuais, a responsabilidade da agremiação pela operação como um todo é mais fácil de demonstrar do que a responsabilidade de cada candidato individualmente. A segunda é a das operações financiadas pelo fundo partidário: quando recursos do Fundo Especial de Financiamento de Campanha ou do Fundo Partidário são utilizados para contratar ferramentas de IA que produzem conteúdo irregular, a responsabilidade financeira do partido pelo conteúdo produzido é mais direta e mais facilmente rastreável do que quando o financiamento ocorre por vias informais.

A responsabilidade das plataformas digitais é o componente mais novo e mais dinamicamente desenvolvido do regime de responsabilização eleitoral, e a Resolução TSE nº 23.732/2024 representa o avanço mais significativo nessa dimensão ao estabelecer responsabilidade solidária das plataformas pelos danos decorrentes de conteúdo mantido em circulação após notificação formal do TSE ou dos TREs. Como analisado no Capítulo 8, essa responsabilidade solidária rompe com o modelo de imunidade dos intermediários estabelecido pelo artigo 19 do Marco Civil da Internet para o contexto eleitoral, criando um regime especial que reconhece que as plataformas não são intermediários neutros, mas atores com papel ativo na amplificação de conteúdo e com capacidade técnica de intervir quando notificadas.

A operacionalização da responsabilidade solidária das plataformas envolve três momentos processuais distintos que o advogado eleitoral precisa dominar. O primeiro é a notificação formal, que deve ser feita por meio dos canais estabelecidos nos acordos de cooperação entre o TSE e cada plataforma, com identificação precisa do conteúdo irregular, da norma violada e do prazo de vinte e quatro horas para remoção. Notificações informais, por e-mail ou por canais de usuário comum, não acionam o regime de responsabilidade solidária da Resolução: apenas a notificação pelo canal formal estabelecido gera a obrigação de remoção cujo descumprimento sujeita a plataforma à multa diária de R$ 150.000. O segundo momento é o monitoramento do cumprimento, que deve verificar,

dentro do prazo de vinte e quatro horas, se o conteúdo foi efetivamente removido de todas as superfícies da plataforma, incluindo não apenas a publicação original mas os compartilhamentos e as reedições que circulam com o mesmo conteúdo irregular. O terceiro momento é a representação pelo descumprimento, que deve ser protocolada imediatamente após o vencimento do prazo sem remoção, com documentação do conteúdo ainda em circulação e dos dados que demonstram a ausência de providência pela plataforma.

A questão doutrinariamente mais disputada sobre a responsabilidade das plataformas no contexto da IA eleitoral é se ela alcança não apenas o conteúdo irregular que circula nas plataformas, mas os próprios algoritmos de recomendação que o amplificam. Benkler, Faris e Roberts (2018) argumentam que a distinção entre hospedar conteúdo e amplificá-lo é artificial no modelo de negócios das plataformas: quando um algoritmo decide que um deepfake deve ser exibido para dez vezes mais usuários do que o conteúdo de desmentido, a plataforma não está apenas intermediando a circulação, está ativamente contribuindo para o dano (BENKLER; FARIS; ROBERTS, 2018, p. 348-351). Essa posição ainda não encontrou acolhida expressa na jurisprudência do TSE, mas é o horizonte para o qual a doutrina aponta como próximo passo necessário na construção do regime de responsabilização das plataformas no contexto eleitoral.

A responsabilidade dos desenvolvedores de ferramentas de IA utilizadas na produção de conteúdo eleitoral irregular é o elo menos desenvolvido do sistema de responsabilização e aquele para o qual a legislação brasileira oferece os instrumentos mais fragmentados. Não existe, no ordenamento eleitoral vigente, dispositivo que responsabilize diretamente os desenvolvedores de ferramentas de IA pelo uso eleitoral irregular de seus produtos. A responsabilidade por esse elo da cadeia precisaria ser construída pela via da responsabilidade civil geral do Código Civil, particularmente pelo artigo 927, parágrafo único, que prevê responsabilidade objetiva por danos causados por atividade normalmente desenvolvida que implique risco para terceiros, combinado com a teoria do risco criado aplicada ao desenvolvimento de tecnologias com potencial de uso eleitoral irregular. Essa construção doutrinária está em estágio inicial no Brasil, mas é precisamente para ela que o PL 2.338/2023, o projeto de lei de regulação geral da IA em tramitação no Congresso, aponta ao estabelecer obrigações específicas para desenvolvedores de sistemas de alto risco, categoria que inclui sistemas de IA utilizados para influenciar processos eleitorais.

O quadro a seguir organiza o regime de responsabilidade em perspectiva comparada entre os quatro atores centrais da cadeia de produção e distribuição de conteúdo de IA eleitoral, permitindo visualizar com precisão as diferenças de fundamento, intensidade e

excludentes disponíveis para cada um. A coluna dos excludentes é a mais relevante para a estratégia de defesa: ela revela que a arquitetura do sistema foi desenhada de forma que cada elo da cadeia tem um caminho possível de exclusão de responsabilidade, mas que esse caminho é progressivamente mais estreito conforme o ator se aproxima do benefício eleitoral gerado pelo conteúdo irregular.

O candidato pode invocar desconhecimento combinado com adoção de medidas preventivas adequadas na dimensão penal, mas não na administrativa. A empresa de marketing pode invocar cumprimento das instruções formais do contratante, mas apenas se demonstrar ausência de conhecimento da irregularidade, o que em operações que envolvem conteúdo claramente falso é hermeneuticamente difícil de sustentar.

A plataforma pode invocar impossibilidade técnica demonstrada de identificar e remover o conteúdo, mas a demonstração precisa ser robusta: alegações genéricas sobre volume de conteúdo não constituem impossibilidade técnica para fins da Resolução. O candidato, por sua vez, é o único ator para o qual não existe excludente de responsabilidade administrativa: ele responde pelo conteúdo veiculado em seu nome independentemente de qualquer outra circunstância, o que cria o incentivo mais forte disponível no sistema para que a *due diligence* sobre fornecedores de tecnologia eleitoral seja tratada como prioridade de conformidade e não como formalidade contratual.

Tabela 42. Regime de Responsabilidade por Conteúdo de IA em Campanhas

Ator	Fundamento da Responsabilidade	Tipo de Responsabilidade	Excludente Possível
Candidato	Responsabilidade pelo material de campanha em seu nome (Lei 9.504/1997, art. 36-A).	Objetiva (administrativa); subjetiva (penal).	Prova de desconhecimento e adoção de medidas preventivas adequadas (penal).
Partido/Coligação	Responsabilidade solidária por propaganda veiculada em canais oficiais (Lei 9.096/1995).	Solidária com o candidato.	Inexistência de relação com o conteúdo irregular e ausência de benefício.
Empresa de marketing digital	Responsabilidade própria pela produção do conteúdo irregular; solidária com a campanha.	Solidária e própria.	Cumprimento das instruções formais do contratante sem conhecimento da irregularidade.

Plataforma digital	Responsabilidade após notificação formal de irregularidade e não remoção no prazo (Res. TSE 23.732/2024, art. 27).	Objetiva após notificação.	Impossibilidade técnica demonstrada de identificar e remover o conteúdo.

Fonte: Elaborado pelo autor com base em Gomes (2023), Resolução TSE 23.732/2024 e jurisprudência do TSE.

Jurisprudência do TSE: Precedentes Paradigmáticos

O TSE construiu, ao longo dos últimos ciclos eleitorais, um conjunto de precedentes que progressivamente define os contornos da responsabilização por condutas eleitorais digitais e que constitui, para o advogado eleitoral que precisa orientar clientes sobre riscos concretos e estratégias defensivas, a fonte normativa mais diretamente aplicável ao cotidiano da prática. A relevância desses precedentes supera a de qualquer análise doutrinária abstrata precisamente porque eles revelam como o TSE opera na interface entre a norma geral e o caso concreto: que tipo de prova considera suficiente, que nível de distância entre o candidato e a conduta irregular ainda configura responsabilidade, que elementos determinam a graduação da sanção e quais argumentos defensivos foram aceitos ou rejeitados em situações análogas.

A jurisprudência do TSE sobre condutas eleitorais digitais desenvolveu-se em três fases que refletem a evolução tecnológica dos próprios fenômenos regulados. A primeira fase, que abrange aproximadamente o período de 2010 a 2017, foi dominada por casos de propaganda eleitoral irregular na internet, perfis falsos em redes sociais e uso não declarado de impulsionamento pago. Nessa fase, o TSE construiu os princípios gerais que ainda hoje estruturam a responsabilização: a teoria da responsabilidade objetiva do candidato pelo conteúdo de campanha, a doutrina do benefício como critério de imputação e o entendimento de que a dificuldade de identificação do autor material não exclui a responsabilização do candidato beneficiado.

A segunda fase, que abrange o período de 2018 a 2022, foi marcada pelos casos decorrentes dos disparos em massa pelo WhatsApp e pelas primeiras ocorrências de desinformação estrutural documentada. O precedente mais significativo dessa fase é o RQ-0060125/2022, que estabeleceu que o financiamento coletivo por grupos de empresários de operações de disparo em massa configura abuso de poder econômico independentemente da demonstração de conhecimento formal do candidato beneficiado, desde que o nexo entre o financiamento e a campanha possa ser estabelecido por prova

indiciária consistente. Esse precedente foi analiticamente relevante por dois motivos: primeiro, por flexibilizar o padrão probatório do nexo causal em operações propositalmente opacas; segundo, por estabelecer que a dissociação deliberada entre financiadores e candidatos como estratégia de criação de negação plausível não é, por si mesma, excludente de responsabilidade quando os demais elementos do abuso estão presentes (TSE, RQ-0060125/2022).

A terceira fase, inaugurada pelas eleições municipais de 2024 com a vigência da Resolução TSE nº 23.732/2024, está em construção e ainda não produziu precedentes suficientemente consolidados para análise sistemática. Os primeiros casos de aplicação da Resolução, envolvendo deepfakes de candidatos a prefeito em municípios de médio porte e chatbots eleitorais sem identificação, confirmaram a disposição do TSE de aplicar o novo marco normativo com velocidade superior à dos ciclos anteriores, particularmente nas liminares de remoção, mas ainda não geraram julgamentos definitivos que permitam extrair parâmetros claros sobre os critérios de graduação das multas e sobre os limites da responsabilidade solidária das plataformas em casos específicos.

Quatro precedentes paradigmáticos merecem análise detalhada pela sua influência direta sobre o entendimento atual do TSE e pela sua aplicabilidade a casos que envolvem o uso de IA eleitoral.

O primeiro é o conjunto de decisões sobre responsabilidade objetiva do candidato pelo conteúdo de campanha, cuja formulação mais clara está no REspe nº 39.681/MG, em que o TSE estabeleceu que o candidato responde pelo material veiculado em seu nome ainda que não tenha autorizado expressamente a veiculação, desde que o conteúdo tenha sido produzido no âmbito de sua campanha e que ele não tenha adotado medidas razoáveis para evitar sua veiculação ou para providenciar sua remoção após tomar conhecimento. Esse precedente é diretamente aplicável ao contexto de IA: um candidato cuja agência produziu conteúdo irregular com IA sem seu conhecimento responde administrativamente pelo conteúdo se não demonstrar que adotou medidas razoáveis de conformidade, como cláusulas contratuais específicas, auditorias de conteúdo e protocolos de revisão (TSE, REspe nº 39.681/MG, 2014).

O segundo precedente paradigmático é o RCED nº 833/DF, em que o TSE analisou caso de uso de dados pessoais de cidadãos para orientação de propaganda eleitoral sem consentimento dos titulares. Embora o caso seja anterior à LGPD e à regulação específica de IA, ele estabeleceu o princípio de que o uso de informações pessoais dos eleitores para fins eleitorais sem autorização constitui conduta que compromete a liberdade do voto e a

igualdade da disputa, podendo configurar abuso de poder político quando o acesso às informações derivou de posição governamental. A aplicação desse precedente ao microtargeting eleitoral por IA orientado por dados obtidos de fontes não autorizadas é direta e foi reconhecida por doutrinadores como Gomes (2023), que identifica o precedente como fundamento jurisprudencial para a responsabilização por uso irregular de dados em campanhas que utilizem sistemas de personalização automatizada (GOMES, 2023, p. 412-415).

O terceiro precedente é a série de decisões sobre uso indevido dos meios de comunicação por WhatsApp, cujo acúmulo nos ciclos de 2018 e 2020 construiu o entendimento de que o aplicativo, apesar de sua natureza de comunicação privada, pode ser objeto de responsabilização eleitoral quando o padrão de envio revela uso coordenado e em escala para fins eleitorais. O parâmetro desenvolvido pela jurisprudência, que distingue comunicação individual legítima de disparo em massa irregular, combina critérios quantitativos, volume de mensagens enviadas, com critérios qualitativos, identidade de conteúdo e coordenação entre remetentes, numa análise que se aplica integralmente ao uso de IA para personalização em escala de disparos que, individualmente, parecem comunicações pessoais mas coletivamente revelam a estrutura de uma operação coordenada.

O quarto precedente é o conjunto de decisões sobre o momento da responsabilização das plataformas digitais, em que o TSE consolidou o entendimento de que a notificação formal ao canal de suporte eleitoral estabelecido no acordo de cooperação é condição necessária e suficiente para ativar o dever de remoção e, consequentemente, a responsabilidade solidária pelo descumprimento. A precisão sobre o que constitui notificação formal adequada, distinguindo-a de denúncias por canais comuns de usuário ou de comunicações informais, tem importância prática significativa para a estratégia de responsabilização de plataformas: uma notificação mal direcionada não ativa o prazo de vinte e quatro horas da Resolução e não produz o efeito de solidariedade que o sistema pressupõe.

A análise comparada desses quatro precedentes revela um padrão consistente na jurisprudência do TSE que orienta tanto a estratégia de responsabilização quanto a estratégia de defesa: o Tribunal valoriza a prova de adoção de medidas razoáveis de conformidade como atenuante ou excludente na dimensão administrativa, tende a flexibilizar o padrão probatório do nexo causal quando a dissociação entre condutas e beneficiários é estruturalmente artificial, e aplica o critério do benefício eleitoral como

substituto da prova de autoria quando esta é impossível de obter no tempo útil do processo eleitoral. Para o advogado que orienta candidatos, esses três padrões traduzem-se em três recomendações operacionais: documente as medidas de conformidade antes que o processo se inicie, não depois; não invista em estruturas de dissociação que o TSE progressivamente aprende a atravessar; e calcule o risco de responsabilização a partir do benefício eleitoral potencial do conteúdo irregular, não apenas da distância formal entre o candidato e a conduta.

⚖ JURISPRUDÊNCIA TSE — RO nº 0600814-51.2022.6.00.0000

EMENTA:

Recurso Ordinário. Propaganda eleitoral irregular. Vídeo manipulado de candidato com edição seletiva que distorcia o sentido de suas declarações originais. Candidato adversário alegou que o vídeo era deepfake; análise forense do TSE concluiu que se tratava de edição simples (shallowfake) sem uso de IA generativa.

RESULTADO:

O TSE determinou a remoção imediata e aplicou multa ao candidato responsável pela divulgação, independentemente da classificação técnica do conteúdo como deepfake ou shallowfake. O acórdão estabeleceu que a irregularidade não depende da sofisticação tecnológica da manipulação, mas de seu efeito sobre a percepção do eleitor. Precedente relevante para a interpretação do conceito de conteúdo sintético da Resolução nº 23.732/2024.

⚖ JURISPRUDÊNCIA TSE — Representação nº 0604891-33.2024.6.00.0000

EMENTA:

Representação por propaganda eleitoral irregular. Candidato a vereador utilizou chatbot não identificado como sistema automatizado para interagir com eleitores via WhatsApp, coletando dados pessoais sem informar o objetivo e simulando interação pessoal com o candidato.

RESULTADO:

O TSE aplicou multa de R$ 45.000 ao candidato, determinou a cessação imediata do funcionamento do chatbot, a eliminação dos dados coletados sem consentimento e comunicou os fatos à ANPD para análise da conformidade com a LGPD. O caso estabeleceu o entendimento de que a responsabilidade do candidato pelo chatbot é objetiva e que a ignorância sobre o funcionamento técnico da ferramenta não é excludente de responsabilidade.

Direito de Resposta: Instrumento Essencial na Era dos Deepfakes

O direito de resposta eleitoral, previsto no artigo 58 da Lei nº 9.504/1997 e regulamentado pela Resolução TSE nº 23.608/2019, é o mecanismo processual que permite ao candidato atingido por conteúdo eleitoral irregular obter, de forma célere e

proporcional, espaço equivalente para veicular sua resposta no mesmo veículo e com o mesmo alcance em que ocorreu a ofensa. Na era dos deepfakes e do conteúdo sintético gerado por IA, esse instrumento adquiriu relevância renovada e complexidade adicional que a jurisprudência está progressivamente endereçando, porque os problemas que ele enfrenta, identificar o veículo onde ocorreu a ofensa, dimensionar o alcance da resposta proporcional e garantir que ela chegue ao mesmo público que foi exposto ao conteúdo irregular, tornaram-se exponencialmente mais difíceis num ecossistema informacional fragmentado e de circulação acelerada.

O fundamento do direito de resposta eleitoral é distinto do fundamento das demais sanções do sistema: ele não é primariamente punitivo, mas restaurativo. Seu objetivo não é castigar quem produziu o conteúdo irregular, para isso existem as multas e as sanções penais, mas restabelecer as condições informacionais que a irregularidade distorceu, oferecendo ao eleitorado exposto ao conteúdo falso a oportunidade de receber a versão correta com alcance equivalente. Essa natureza restaurativa tem uma implicação processual relevante que frequentemente é subestimada: o direito de resposta pode ser concedido mesmo quando a autoria do conteúdo irregular não está identificada, porque a restauração das condições informacionais não depende de que o autor seja conhecido, apenas de que a distorção seja demonstrada e de que exista veículo com capacidade de veicular a resposta.

A jurisprudência do TSE tem progressivamente adaptado os critérios clássicos de concessão de direito de resposta ao contexto digital com resultados que ainda estão em evolução, mas que já permitem identificar as linhas interpretativas dominantes.

O critério da equivalência, que exige que a resposta seja veiculada no mesmo canal e com alcance equivalente ao do conteúdo irregular, foi interpretado de forma funcionalmente ampla para o contexto digital. Numa aplicação literal, esse critério exigiria que a resposta fosse veiculada no mesmo perfil ou canal que publicou o conteúdo irregular, o que frequentemente é impossível quando o autor é anônimo ou o canal foi encerrado após a publicação. A interpretação funcional adotada pelo TSE substitui o critério do mesmo canal pelo critério do mesmo público: o que se exige é que a resposta alcance, com intensidade equivalente, a audiência que foi exposta ao conteúdo irregular. Isso permite determinar que a plataforma amplifique a resposta para o mesmo público-alvo que recebeu o deepfake, utilizando os mesmos mecanismos de segmentação que foram eventualmente utilizados na distribuição do conteúdo irregular. Como argumenta Gomes (2023), essa interpretação funcional do critério de equivalência é a única que preserva a finalidade

restaurativa do direito de resposta num ambiente em que a identidade do canal e a identidade do público frequentemente divergem (GOMES, 2023, p. 421-424).

O critério da proporcionalidade, que determina que a extensão e o formato da resposta sejam proporcionais à ofensa, foi adaptado para considerar uma dimensão que os casos analógicos não apresentavam com a mesma intensidade: o timing da violação em relação ao pleito. Um deepfake divulgado sessenta dias antes da eleição deixa ao candidato afetado tempo suficiente para que a resposta circule e seja processada pelo eleitorado com alguma profundidade.

Um deepfake divulgado quarenta e oito horas antes do primeiro turno, seguindo o padrão de timing estratégico documentado nos casos internacionais e brasileiros analisados nos capítulos anteriores, opera numa janela temporal em que a resposta, mesmo que veiculada imediatamente, não consegue reverter o dano epistêmico já causado ao eleitor que recebeu o conteúdo irregular e formou sua percepção antes da correção.

Nessa segunda situação, a jurisprudência do TSE tem reconhecido que o direito de resposta precisa ser combinado com a medida liminar de remoção imediata: sem a remoção, o direito de resposta chega ao eleitor depois que o deepfake já construiu o enquadramento de referência pelo qual o eleitor avaliará qualquer informação subsequente, incluindo a resposta. A combinação das duas medidas, remoção e resposta, é mais eficaz do que qualquer uma delas isoladamente precisamente porque endereça dimensões distintas do problema: a remoção interrompe a exposição ao conteúdo falso, a resposta oferece o conteúdo correto ao público que já foi exposto (PENNYCOOK; RAND, 2019, p. 2523-2524).

A aplicação do direito de resposta a conteúdo sintético disseminado por WhatsApp apresenta a limitação estrutural mais difícil de superar. Como analisado no Capítulo 9, o WhatsApp é um ambiente de mensageria privada criptografado em que a Justiça Eleitoral não tem acesso ao conteúdo circulante e em que não existe mecanismo técnico de veiculação compulsória de resposta: não há como determinar ao WhatsApp que envie a resposta do candidato para os mesmos grupos que receberam o deepfake, porque o WhatsApp não tem registro desses grupos nem mecanismo técnico de envio compulsório para ambientes privados. Nessa situação, o direito de resposta produz efeitos apenas nas superfícies públicas onde o conteúdo também eventualmente circula, como Facebook, Instagram, X e YouTube, e a resposta que chegará ao público exposto pelo WhatsApp depende inteiramente da disseminação orgânica da resposta por esses canais secundários.

A consciência dessa limitação estrutural é relevante para a estratégia de comunicação de crise da campanha: quando o deepfake circula primariamente pelo

WhatsApp, a resposta jurídica pelo direito de resposta precisa ser complementada por uma estratégia comunicacional ativa que utilize os mesmos canais de distribuição informal para alcançar o mesmo público, independentemente do que a Justiça Eleitoral pode determinar.

A legitimidade ativa para o pedido de direito de resposta é ampla: o artigo 58 da Lei nº 9.504/1997 confere o direito ao candidato, ao partido e à coligação atingidos, ainda que de forma indireta, por conceito, imagem ou afirmação caluniosa, difamatória, injuriosa ou notoriamente inverídica difundida por qualquer veículo de comunicação social, inclusive por provedores de aplicativos de internet e redes sociais. A expressão "notoriamente inverídica" é especialmente relevante para o contexto dos deepfakes: ela dispensa a demonstração de que o conteúdo é intencionalmente falso, bastando que seja objetivamente inverídico e que essa inveracidade seja verificável pelos meios de prova disponíveis. A análise forense que conclui com probabilidade superior a 95% que um vídeo é sintético constitui prova de inveracidade notória suficiente para fundamentar o pedido de direito de resposta sem necessidade de identificação do autor ou de demonstração do dolo.

O prazo para o pedido de direito de resposta, estabelecido pela Resolução TSE nº 23.608/2019 em até vinte e quatro horas após o conhecimento do fato pelo atingido, é o elemento processual que mais frequentemente compromete o instrumento na prática. Em ambiente de circulação acelerada de conteúdo sintético, vinte e quatro horas é um prazo que pode ser suficiente para que o deepfake já tenha atingido centenas de milhares de pessoas antes que o candidato afetado perceba a disseminação, obtenha análise forense, formalize o pedido e obtenha a decisão judicial.

A estratégia mais eficaz para enfrentar essa compressão temporal é a constituição prévia de uma equipe de monitoramento e resposta que opere em tempo real durante o período eleitoral e que tenha protocolos pré-definidos de acionamento do direito de resposta sem necessidade de decisão ad hoc sobre cada passo do processo. Campanhas que dependem de improvisação para responder a deepfakes estão sistematicamente em desvantagem em relação ao prazo de vinte e quatro horas; campanhas com protocolo de resposta pré-definido têm condições de cumpri-lo.

■ RESUMO DO CAPÍTULO:

Este capítulo analisou o sistema de responsabilização por condutas eleitorais ilícitas praticadas com o auxílio de inteligência artificial, partindo do diagnóstico de que a IA generativa não apenas acelera e escala condutas preexistentes, mas cria condutas inteiramente novas para as quais os tipos normativos foram construídos sem qualquer antecipação. O debate doutrinário entre Gomes e Cândido sobre tipificação foi apresentado como tensão com consequências práticas diretas: enquanto Gomes defende cautela hermenêutica rigorosa em respeito à legalidade estrita penal, Cândido sustenta que a interpretação evolutiva de tipos existentes à luz de novos meios de execução não constitui analogia vedada quando o

núcleo da ilicitude permanece o mesmo. A Resolução TSE nº 23.732/2024 procurou resolver essa tensão criando tipos infracionais administrativos específicos, deslocando parte da responsabilização do campo penal para o administrativo-eleitoral, onde a interpretação evolutiva encontra menor resistência formal. O sistema sancionatório foi analisado em suas três camadas cumulativamente aplicáveis: a administrativa eleitoral, mais ágil e acessível probatoriamente, é o instrumento preferencial para resposta em tempo real; a cível eleitoral, com seus instrumentos de cassação de registro e diploma, produz as consequências mais definitivas sobre o resultado eleitoral; e a penal eleitoral, de maior força simbólica e deterrente, estruturalmente chega depois do dano pela combinação de padrão probatório elevado e prazo prescricional incompatível com a urgência eleitoral. A estratégia mais eficaz é o acionamento simultâneo das três camadas, extraindo de cada uma o que oferece de melhor. Os cinco tipos penais eleitorais com maior relevância para condutas habilitadas por IA foram analisados individualmente: calúnia eleitoral, difamação eleitoral, injúria eleitoral, propaganda enganosa e crimes contra o Estado Democrático de Direito da Lei nº 14.197/2021. O elemento subjetivo exigido em todos os tipos, o dolo direto, foi identificado como o principal obstáculo probatório da camada penal em contexto de IA, porque demonstrar que o responsável conhecia a falsidade do conteúdo produzido algoritmicamente é tecnicamente complexo e frequentemente inconclusivo no tempo útil do processo eleitoral. O abuso de poder político e econômico foi analisado como alternativa estratégica quando a prova do dolo é difícil, dada sua natureza predominantemente objetiva: o que importa é o efeito potencial sobre a igualdade da disputa, não a intenção do agente. Três modalidades de abuso de poder político no contexto de IA foram identificadas: uso de infraestrutura governamental para produção de conteúdo de IA, uso de dados públicos para microtargeting eleitoral e disseminação coordenada de ataques com recursos do erário. O abuso de poder econômico ganhou dimensão adicional com a IA generativa: o colapso do custo de produção cria assimetria entre o impacto eleitoral gerado e o custo declarado na prestação de contas, questão que a regulação vigente ainda não resolveu. O regime de responsabilização foi analisado em perspectiva comparada entre quatro atores: o candidato, com responsabilidade objetiva na dimensão administrativa e subjetiva na penal; o partido, com responsabilidade solidária especialmente relevante em operações de infraestrutura compartilhada; as plataformas, com responsabilidade solidária após notificação formal, cujo protocolo preciso é condição necessária para ativar o prazo de vinte e quatro horas; e os desenvolvedores de ferramentas, o elo menos desenvolvido do sistema, para o qual a responsabilização depende de construção doutrinária ainda em estágio inicial. A jurisprudência do TSE foi periodizada em três fases, com análise de quatro precedentes paradigmáticos que revelam um padrão consistente: valorização das medidas razoáveis de conformidade como atenuante, flexibilização do nexo causal em operações propositalmente opacas e aplicação do critério do benefício eleitoral como substituto da prova de autoria. O capítulo encerrou com a análise do direito de resposta como instrumento restaurativo, não punitivo, com suas adaptações jurisprudenciais ao contexto digital, suas limitações estruturais no WhatsApp e a conclusão de que a combinação de remoção liminar e direito de resposta é mais eficaz do que qualquer um dos instrumentos isoladamente.

QUESTÕES PARA REFLEXÃO E DEBATE

1. A tensão doutrinária entre Gomes e Cândido sobre tipificação de condutas eleitorais digitais reflete uma tensão constitucional mais profunda entre o princípio da legalidade estrita penal, que protege o acusado de extensões analógicas não autorizadas, e o princípio da proteção da integridade eleitoral, que exige respostas eficazes a condutas que o legislador não antecipou. Como o sistema jurídico brasileiro deveria resolver essa tensão para condutas que são funcionalmente idênticas às tipificadas mas tecnologicamente novas?

2. O deslocamento da responsabilização do campo penal para o administrativo-eleitoral pela Resolução TSE nº 23.732/2024 resolve o problema da legalidade estrita, mas cria uma assimetria: o mesmo fato ilícito pode ser sancionado administrativamente com multa imediata e negado na esfera penal por falta de tipicidade. Essa assimetria é sistemicamente desejável, ou ela compromete a coerência do ordenamento e a percepção de justiça pelos atores do processo eleitoral?

3. O abuso de poder econômico com IA generativa cria uma lacuna específica: o custo efetivo das ferramentas é desproporcional ao impacto eleitoral gerado, permitindo que campanhas declarem gastos mínimos e produzam efeito equivalente a operações de publicidade de grande escala. Qual critério de mensuração, custo efetivo das ferramentas ou valor de mercado equivalente do impacto produzido, seria mais adequado para a prestação de contas eleitoral, e como ele seria operacionalizável na prática?

4. A responsabilidade objetiva do candidato pelo conteúdo gerado por agentes automatizados contratados, mesmo sem conhecimento específico da conduta, foi estabelecida pela Resolução TSE nº 23.732/2024. Esse regime cria incentivo para *due diligence*, mas também pode penalizar candidatos de menor porte que não têm recursos para contratar assessoria jurídica especializada em conformidade. Como o direito eleitoral deveria tratar essa assimetria de capacidade de conformidade sem violar o princípio da igualdade eleitoral?

5. A questão doutrinariamente mais disputada sobre responsabilidade das plataformas é se ela alcança os algoritmos de recomendação que amplificam o conteúdo irregular, não apenas o conteúdo em si. Benkler, Faris e Roberts argumentam que a distinção entre hospedar e amplificar é artificial no modelo de negócios das plataformas. Que instrumentos jurídicos do ordenamento brasileiro, existentes ou a serem criados, poderiam fundamentar a responsabilização pelo ato de amplificação algorítmica, distinto do ato de hospedagem?

6. O direito de resposta tem natureza restaurativa, não punitiva, e pode ser concedido mesmo sem identificação do autor do conteúdo irregular. Essa característica o torna o instrumento mais adequado para o contexto de deepfakes disseminados por atores anônimos. Contudo, sua limitação estrutural no WhatsApp, onde não existe mecanismo técnico de veiculação compulsória de resposta, significa que o instrumento falha precisamente no canal onde a desinformação mais circula. Que reforma regulatória ou técnica poderia endereçar essa lacuna sem comprometer a criptografia de ponta a ponta que protege a privacidade das comunicações?

7. O padrão jurisprudencial identificado no TSE valoriza a prova de adoção de medidas razoáveis de conformidade como atenuante ou excludente. Isso cria incentivo para que candidatos documentem suas práticas de governança de IA eleitoral antes que qualquer processo se inicie. Que conjunto mínimo de documentos e protocolos deveria uma campanha produzir para demonstrar que adotou medidas razoáveis de conformidade com a Resolução TSE nº 23.732/2024?

8. A Lei nº 14.197/2021 introduziu o crime de ataques ao Estado Democrático de Direito com pena de reclusão de dois a oito anos, aplicável a campanhas sistemáticas de desinformação que visem deslegitimar o processo eleitoral. Esse tipo penal foi construído para endereçar fenômenos como a narrativa de fraude nas urnas de 2022. Em que medida ele é aplicável a campanhas de desinformação gerada por IA que não têm por objetivo declarado atacar o processo eleitoral, mas cujo efeito agregado produz erosão da confiança institucional? O elemento subjetivo da "finalidade específica de atacar o Estado Democrático" pode ser demonstrado por prova indireta em operações de IA coordenadas?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Acione as três camadas simultaneamente. Cada uma faz o que as outras não conseguem.

Diante de um deepfake ou conteúdo de IA irregular contra seu candidato, a estratégia mais eficaz não é escolher entre representação administrativa, ação cível ou denúncia criminal: é acionar as três ao mesmo tempo. A representação administrativa ao TSE ou TRE busca a liminar de remoção imediata, que é o instrumento mais urgente. A ação de investigação judicial eleitoral (AIJE) pressiona sobre o resultado eleitoral com a ameaça de cassação. A queixa-crime por calúnia eleitoral carrega força simbólica e deterrente que as sanções administrativas não têm. Cada camada alcança uma dimensão do problema que as outras não cobrem. A escolha entre elas não é estratégia: é desperdício de instrumentos.

2. Acionar o direito de resposta: você tem 24 horas e precisa estar pronto antes do ataque.

O prazo de vinte e quatro horas para o pedido de direito de resposta, contados do conhecimento do fato, é o elemento processual que mais frequentemente compromete o instrumento na prática. Quando o deepfake é identificado, a campanha precisa simultaneamente obter análise forense, preparar o pedido, protocolar na Justiça Eleitoral e acompanhar a decisão, tudo em menos de um dia. Isso é impossível por improvisação. A única forma de cumprir esse prazo é ter protocolos pré-definidos: equipe de monitoramento em tempo real, perito forense de confiança já contratado, modelo de petição já revisado, canal direto com o advogado eleitoral ativo 24 horas. Campanhas que chegam a outubro de 2026 sem esse protocolo estão sistematicamente em desvantagem.

3. Notificação de plataforma: canal errado não ativa o prazo de 24 horas.

Para acionar a responsabilidade solidária da plataforma e o prazo de vinte e quatro horas para remoção, a notificação precisa ser feita pelo canal formal estabelecido nos acordos de cooperação entre o TSE e cada plataforma. E-mail para suporte geral, denúncia pelo botão de usuário comum ou comunicação informal com funcionários da plataforma não ativam o regime da Resolução TSE 23.732/2024. Se o prazo de vinte e quatro horas não for acionado corretamente, a multa diária de R$ 150.000 e a responsabilidade solidária não se aplicam. O advogado eleitoral precisa conhecer o canal formal de cada plataforma com antecedência, não no momento da crise.

4. Ignorar o que sua agência faz com IA não é excludente de responsabilidade administrativa.

O precedente do REspe nº 39.681/MG estabeleceu que o candidato responde pelo material veiculado em seu nome ainda que não tenha autorizado expressamente a veiculação, desde que não demonstre que adotou medidas razoáveis de conformidade. No contexto da IA eleitoral, isso significa que o candidato que não sabe o que sua agência está fazendo com ferramentas de IA, não tem cláusula contratual específica sobre uso de IA e rotulagem, e não auditou o conteúdo produzido, responde administrativamente por qualquer violação que sua agência cometa. O desconhecimento não é excludente: é a ausência das medidas razoáveis que o precedente exige para excluir a responsabilidade.

5. Use o abuso de poder quando a prova do dolo do adversário for difícil de obter.

Quando seu adversário produziu ou financiou conteúdo irregular por IA mas a prova de seu conhecimento específico da conduta é difícil de obter, o abuso de poder pode ser o caminho mais acessível de responsabilização. O abuso de poder tem natureza predominantemente objetiva: o que precisa ser demonstrado não é a intenção do agente, mas o efeito potencial de sua conduta sobre a igualdade da disputa eleitoral. Se o conteúdo irregular foi financiado com recursos não contabilizados, distribuído em escala suficiente para influenciar o resultado e pode ser vinculado à campanha adversária por prova indiciária consistente, os pressupostos do abuso estão presentes mesmo sem prova direta de dolo.

6. Direito de resposta no WhatsApp: instrumento jurídico não resolve. Comunicação ativa resolve.

Quando o deepfake ou conteúdo irregular circula primariamente por grupos de WhatsApp, o direito de resposta produz efeitos apenas nas superfícies públicas onde o conteúdo também circula, como Facebook, Instagram e YouTube. Não existe mecanismo técnico para que a Justiça Eleitoral determine ao WhatsApp que envie a resposta aos mesmos grupos que receberam o deepfake. Isso significa que a resposta que alcançará o público exposto pelo WhatsApp depende de estratégia comunicacional ativa, não de instrumentos jurídicos. A campanha precisa usar os mesmos canais de distribuição informal, grupos de confiança, líderes de opinião locais, redes religiosas e empresariais, para disseminar a resposta com a mesma capilaridade que o conteúdo irregular usou para chegar ao eleitor.

7. Documente as medidas de conformidade hoje. Você vai precisar delas no processo de amanhã.

O padrão jurisprudencial do TSE valoriza a prova de adoção de medidas razoáveis de conformidade como atenuante ou excludente na dimensão administrativa eleitoral. Isso significa que o conjunto de documentos produzidos antes de qualquer processo, contratos com cláusulas específicas sobre IA, auditorias de conteúdo, protocolos de revisão, registros de treinamento da equipe, determinará o resultado da defesa tanto quanto os argumentos jurídicos invocados no processo. Documentação produzida depois que o processo se inicia tem muito menos valor probatório do que documentação que existia antes. A governança de IA eleitoral precisa começar agora, não quando a primeira representação for protocolada.

CAPÍTULO 13

LGPD e Proteção de Dados nas Eleições

Tratamento de Dados Pessoais, Consentimento e a Articulação entre ANPD e Justiça Eleitoral

Dados Pessoais, Privacidade e Democracia

A relação entre proteção de dados pessoais e integridade democrática é um dos temas de maior desenvolvimento na literatura jurídica e política contemporânea, e sua relevância para o direito eleitoral brasileiro ainda não foi inteiramente incorporada pela doutrina especializada. A proteção de dados é frequentemente tratada como matéria de direito civil ou regulação tecnológica, separada do debate sobre integridade eleitoral. Essa separação é analiticamente equivocada e praticamente prejudicial: os dados pessoais dos eleitores são a matéria-prima sobre a qual as tecnologias de persuasão eleitoral operam, e a qualidade da proteção que o ordenamento oferece a esses dados determina, em larga medida, a extensão em que eleitores permanecem capazes de exercer autonomia política genuína.

Daniel Solove, em *Nothing to Hide* (2013), fornece o enquadramento teórico mais influente para compreender a privacidade não como interesse individual isolado, mas como condição estrutural da democracia. O argumento central de Solove é que a privacidade não importa apenas porque protege segredos individuais: ela importa porque preserva o espaço necessário para que cidadãos formem e expressem opiniões políticas sem a pressão do monitoramento, desenvolvam identidades políticas sem o risco de exploração das vulnerabilidades que esse desenvolvimento inevitavelmente revela, e participem do debate público sem o peso do conformismo que a vigilância permanente induz. Em contexto de monitoramento intensivo de comportamento digital, a erosão da privacidade não é apenas um dano individual: é uma distorção das condições de possibilidade da participação política autêntica (SOLOVE, 2013, p. 1-8).

Shoshana Zuboff, em *A Era do Capitalismo de Vigilância* (2019), aprofunda essa análise ao demonstrar que o modelo econômico das grandes plataformas digitais não é apenas um negócio que coleta dados como subproduto de sua atividade principal: é um negócio cujo produto principal é precisamente a extração e comercialização de dados comportamentais para fins de modificação comportamental em escala. Nesse modelo, os dados do eleitor não são apenas informação sobre suas preferências: são matéria-prima

para a produção de sistemas preditivos e instrumentos de influência que podem ser vendidos a qualquer ator com recursos suficientes, incluindo campanhas eleitorais, governos e atores estrangeiros com interesse em resultados eleitorais domésticos. A consequência mais perturbadora que Zuboff identifica não é que os dados sejam usados para prever comportamento: é que sejam usados para modificá-lo de formas que os próprios cidadãos não percebem e não consentiram (ZUBOFF, 2019, p. 11-14).

A articulação entre essas análises teóricas e o contexto jurídico brasileiro tem dois marcos normativos centrais que precisam ser compreendidos em conjunto. O primeiro é a Lei Geral de Proteção de Dados, Lei nº 13.709/2018, aprovada num contexto em que o escândalo Cambridge Analytica havia tornado inescapável para o legislador brasileiro a relação entre proteção de dados e integridade democrática. Como documentam Pinheiro (2021) e Bioni (2021), as obras doutrinárias mais completas sobre a interpretação da LGPD no contexto brasileiro, a lei foi construída sobre um fundamento que vai além da proteção de direitos individuais: ela reconhece que o tratamento inadequado de dados pessoais em escala é uma questão de ordem pública, com consequências que se estendem à qualidade do debate democrático e à igualdade de condições na competição política (PINHEIRO, 2021, p. 45-52; BIONI, 2021, p. 89-94).

O segundo marco normativo é a Resolução TSE nº 23.732/2024, que articulou pela primeira vez, de forma expressa e operacional, as obrigações de proteção de dados com as obrigações de integridade eleitoral. Ao estabelecer que candidatos e campanhas devem observar as disposições da LGPD no tratamento de dados pessoais de eleitores, ao vedar o uso de dados obtidos sem consentimento para fins de microtargeting eleitoral e ao determinar a eliminação de dados coletados irregularmente como sanção acessória em casos de violação, a Resolução criou a ponte normativa que transforma a LGPD de lei de proteção individual em instrumento de proteção da integridade eleitoral coletiva.

A articulação específica entre LGPD e direito eleitoral é, contudo, um campo em que a produção doutrinária permanece desproporcionalmente escassa em relação à importância do tema. A maioria das análises disponíveis aborda a LGPD e o direito eleitoral como campos paralelos com pontos de contato ocasionais, em vez de como sistemas normativos que precisam ser lidos de forma integrada para que cada um produza seus efeitos plenos. Esta obra procura contribuir para o preenchimento dessa lacuna ao examinar sistematicamente os principais pontos de intersecção entre os dois campos: as bases legais para o tratamento de dados pessoais em contexto eleitoral, os requisitos específicos que a LGPD impõe ao microtargeting político, as obrigações de transparência

sobre uso de IA no tratamento de dados eleitorais e o regime de responsabilização que resulta da articulação entre as sanções da LGPD e as sanções eleitorais específicas da Resolução TSE nº 23.732/2024.

A tese que orienta essa análise integrada pode ser enunciada de forma direta: a proteção de dados pessoais dos eleitores não é uma obrigação paralela às obrigações eleitorais que campanhas devem observar. É uma condição necessária para que o processo eleitoral cumpra sua função democrática de permitir que cidadãos formem e expressem preferências políticas autênticas. Uma campanha que viola a LGPD para orientar seu microtargeting não está apenas cometendo uma infração administrativa de proteção de dados: está comprometendo a qualidade epistêmica do processo pelo qual o eleitor forma suas preferências, o que é um dano ao próprio fundamento da legitimidade eleitoral que o direito eleitoral existe para proteger. Essa articulação entre os dois sistemas normativos é o que a análise que se segue procura desenvolver com a precisão que o tema exige.

A LGPD como Garantia Democrática no Contexto Eleitoral

A Lei Geral de Proteção de Dados Pessoais, Lei nº 13.709/2018, é frequentemente apresentada como instrumento de proteção da privacidade individual, e essa caracterização é correta enquanto vai. No contexto eleitoral, porém, ela é também, e talvez primariamente, um instrumento de proteção da democracia. A distinção não é retórica: ela tem consequências diretas sobre como as obrigações da LGPD devem ser interpretadas quando o titular dos dados é um eleitor, o responsável pelo tratamento é uma campanha eleitoral e a finalidade do tratamento é orientar estratégias de persuasão política.

A lógica que conecta proteção de dados à proteção democrática pode ser enunciada com precisão. Se a democracia pressupõe cidadãos capazes de formar preferências políticas autênticas mediante deliberação racional e acesso a informação verdadeira, como argumentado nos capítulos anteriores a partir de Habermas e Sunstein, então qualquer sistema que utilize dados pessoais para identificar vulnerabilidades psicológicas, cognitivas ou situacionais dos eleitores e para calibrar mensagens políticas de forma a explorar essas vulnerabilidades está comprometendo o pressuposto democrático fundamental. Não porque a persuasão política seja ilegítima, pois não é: a política é, em sua essência, persuasão. Mas porque há uma diferença moralmente e juridicamente relevante entre a

persuasão que apela à razão e apresenta razões para consideração consciente e a manipulação que explora vulnerabilidades para produzir comportamentos que o eleitor não escolheria em condições de plena informação e autonomia. Ao regular o tratamento de dados comportamentais que tornam a segunda modalidade possível em escala industrial, a LGPD está, funcionalmente, protegendo as condições de possibilidade da democracia deliberativa.

Essa compreensão ampliada da função da LGPD no contexto eleitoral tem respaldo doutrinário em Bioni (2021), que argumenta que a proteção de dados pessoais no Brasil foi construída sobre um fundamento de ordem pública que transcende a proteção de interesses individuais: o artigo 2º da LGPD, ao elencar entre seus fundamentos o respeito à privacidade, o desenvolvimento econômico e tecnológico e a livre iniciativa, inclui também os direitos humanos, o desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais. A referência explícita ao exercício da cidadania é particularmente relevante: ela indica que o legislador reconheceu que a proteção de dados pessoais tem uma dimensão cívica que não se esgota na proteção individual (BIONI, 2021, p. 95-99).

A interseção entre LGPD e direito eleitoral é, portanto, um campo de crescente relevância prática que ainda não foi sistematizado pela doutrina com a profundidade que o tema exige. Candidatos e partidos que utilizam ferramentas de análise de dados em campanhas precisam assegurar conformidade com a LGPD, sob pena de responsabilização tanto pela Autoridade Nacional de Proteção de Dados quanto pela Justiça Eleitoral. Essa dupla exposição regulatória não é acidental: ela reflete o reconhecimento, expresso na Resolução TSE nº 23.732/2024, de que violações de proteção de dados em contexto eleitoral têm dimensão que ultrapassa o âmbito da ANPD e integra o sistema de proteção da integridade eleitoral.

Para compreender o regime jurídico aplicável, é necessário distinguir três categorias de dados pessoais que as campanhas eleitorais frequentemente tratam, cada uma com regras de licitude distintas. A primeira categoria é a dos dados de identificação e contato, como nome, endereço, e-mail e telefone de eleitores, que podem ser obtidos de registros públicos, de cadastros partidários ou por coleta direta com consentimento. O tratamento dessa categoria para fins de comunicação eleitoral é lícito com base no legítimo interesse do controlador ou no cumprimento de obrigação legal, mas exige, em qualquer caso, transparência sobre a finalidade do tratamento e respeito ao direito de oposição do titular. A segunda categoria é a dos dados comportamentais, como histórico de navegação,

padrões de consumo de conteúdo e registros de interação com plataformas, que são o insumo central dos sistemas de microtargeting eleitoral. O tratamento dessa categoria é significativamente mais restrito, especialmente quando os dados revelam, ainda que por inferência, as opiniões políticas do titular. A terceira categoria, e a de regime jurídico mais restritivo, é a dos dados sensíveis, entre os quais a LGPD inclui expressamente as opiniões políticas no artigo 5º, inciso II, com exigências que a seção seguinte analisa em detalhe.

A conformidade com a LGPD em contexto eleitoral não é, portanto, um exercício de verificação de caixas em formulário de compliance: é uma análise de risco que precisa ser conduzida para cada operação de tratamento específica, identificando a base legal aplicável, avaliando os riscos para os direitos dos titulares e implementando as salvaguardas adequadas. A distinção entre o que é permitido, o que é restrito a bases legais específicas e o que é vedado independentemente de qualquer base legal é o que a análise que se segue procura construir com precisão operacional para o contexto das campanhas eleitorais que utilizem ferramentas de IA e microtargeting.

Dados Sensíveis e a Questão da Inferência Política

O elemento de maior tensão jurídica na interseção entre LGPD e marketing eleitoral é o tratamento de dados sensíveis, especialmente a categoria das opiniões políticas, e a questão ainda mais complexa da inferência de opiniões políticas a partir de dados que, individualmente, não são políticos.

A LGPD define dado sensível no artigo 5º, inciso II, de forma que abrange um conjunto heterogêneo de categorias que o legislador identificou como merecedoras de proteção reforçada em razão de sua particular capacidade de gerar discriminação ou dano ao titular: dados sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou organização de caráter religioso, filosófico ou político, dado referente à saúde ou vida sexual, dado genético ou biométrico, quando vinculado a pessoa natural. A inclusão da opinião política nessa lista não é circunstancial: ela reflete a avaliação legislativa de que dados que revelam posicionamento político do cidadão têm potencial de discriminação e perseguição que justifica regime jurídico mais exigente (PINHEIRO, 2021, p. 87-91).

O regime jurídico dos dados sensíveis é significativamente mais restritivo do que o dos dados pessoais comuns em pelo menos dois aspectos centrais. O primeiro é o rol de bases legais disponíveis: enquanto dados pessoais comuns podem ser tratados com base em qualquer das dez hipóteses do artigo 7º da LGPD, incluindo o legítimo interesse do

controlador, os dados sensíveis estão restritos às hipóteses do artigo 11º, que exclui o legítimo interesse e exige, em regra, consentimento específico e destacado ou uma das hipóteses alternativas taxativamente previstas. O segundo aspecto é a exigência de que o tratamento de dados sensíveis seja necessário para atingir a finalidade declarada, o que importa numa análise de proporcionalidade mais rigorosa do que a exigida para dados comuns: não basta que o tratamento seja conveniente ou útil para a campanha, é necessário que seja necessário para a finalidade específica declarada.

Para as campanhas eleitorais, o problema central não é o tratamento de dados que explicitamente revelam opinião política, como a filiação partidária declarada de um eleitor. Esse tipo de dado é claramente sensível e seu regime jurídico é inequívoco. O problema central é o tratamento de dados que, individualmente ou em combinação, permitem inferir opinião política sem que nenhum dos dados individuais seja, por si mesmo, político. Um eleitor que consome determinado tipo de conteúdo de notícias, segue determinadas contas nas redes sociais, reside em determinado bairro, tem determinada faixa etária e determinado nível de renda está fornecendo, por esse conjunto de dados aparentemente neutros, informação suficiente para que sistemas de machine learning construam um perfil psicográfico que permite prever com alta precisão suas preferências políticas. Os dados individuais são comuns; o perfil resultante é sensível.

A questão jurídica que essa situação coloca ainda não tem resposta consolidada na jurisprudência da ANPD ou do TSE: o tratamento do dado inferido deve obedecer ao regime do dado que o compõe ou ao regime do dado que resulta da inferência? A interpretação mais protetora dos direitos dos titulares, e a que mais se alinha com a finalidade da LGPD de preservar a autodeterminação informativa do cidadão, é a de que o dado inferido, quando é sensível, deve seguir o regime dos dados sensíveis independentemente da natureza dos dados que o geraram. Bioni (2021) apoia essa interpretação ao argumentar que a proteção de dados pessoais deve ser orientada pelo risco para os direitos dos titulares, não pela classificação formal dos dados em categorias estanques: se o processamento de dados comuns produz, como resultado, um perfil que revela opinião política, o risco para o titular é o mesmo que se os dados fossem explicitamente políticos, e a proteção deveria ser equivalente (BIONI, 2021, p. 201-207).

Essa interpretação tem implicações práticas imediatas para o microtargeting eleitoral por IA. Sistemas que constroem perfis psicográficos de eleitores a partir de dados comportamentais, mesmo que nenhum dos dados individuais seja explicitamente político, estão potencialmente tratando dados sensíveis por inferência. Isso significa que a base

legal para esse tratamento precisaria ser uma das hipóteses do artigo 11º da LGPD, e não o legítimo interesse do artigo 7º, inciso IX, que é a base mais frequentemente invocada por campanhas para justificar operações de análise de dados eleitoral. A pergunta que o advogado eleitoral precisa fazer não é apenas "que dados estamos coletando?", mas "que dados estamos produzindo por inferência, e qual é o regime jurídico adequado para esses dados inferidos?".

A ANPD ainda não publicou orientação específica sobre dados inferidos em contexto eleitoral, e o protocolo de cooperação entre ANPD e TSE, previsto na Resolução TSE nº 23.732/2024, ainda está em fase de desenvolvimento. Essa lacuna regulatória não é, porém, um espaço de licitude: é uma zona de risco em que a ausência de orientação específica não elimina a aplicação das disposições gerais da LGPD sobre dados sensíveis, que continuam a viger independentemente de orientação complementar. Campanhas que operam nessa zona de risco sem análise jurídica adequada estão assumindo exposição regulatória cujo alcance, quando a ANPD e o TSE construírem sua jurisprudência conjunta, pode ser significativamente maior do que anteciparam.

LGPD — Regime Jurídico Diferenciado para Dados Sensíveis

Art. 5º, II — Dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural. Art. 11, I — O tratamento de dados pessoais sensíveis somente poderá ocorrer nas seguintes hipóteses: quando o titular ou seu responsável legal consentir, de forma específica e destacada, para finalidades específicas. O consentimento genérico para tratamento de dados pessoais, mesmo que previsto em termos de uso aceitos pelo titular, não satisfaz o requisito de consentimento específico para dados sensíveis.

Brasil. Lei nº 13.709/2018 (LGPD), arts. 5º, II e 11, I.

Bases Legais para Tratamento de Dados em Campanhas Eleitorais

A LGPD estabelece, no artigo 7º, que todo tratamento de dados pessoais deve ter uma base legal que o autorize. Essa exigência, aparentemente simples, é o ponto de maior complexidade prática para campanhas eleitorais que utilizam ferramentas de análise de dados e IA: a base legal precisa ser identificada antes do tratamento, ser adequada à natureza dos dados tratados e à finalidade do tratamento, e ser documentada de forma que demonstre conformidade em caso de investigação pela ANPD ou pela Justiça Eleitoral. A

escolha incorreta da base legal não é uma irregularidade formal: é a ausência de licitude do tratamento, com consequências sancionatórias que se acumulam sobre cada operação de tratamento realizada sem amparo legal adequado.

No contexto eleitoral, três bases legais têm maior relevância prática, cada uma com alcance e limitações que precisam ser compreendidos com precisão antes de qualquer decisão sobre arquitetura de dados de campanha.

O consentimento do titular, previsto no artigo 7º, inciso I, é a base legal mais intuitiva e a mais frequentemente mencionada no debate público. Para ser válido no contexto da LGPD, precisa ser livre, informado, inequívoco e específico para cada finalidade de tratamento. A especificidade é o requisito mais frequentemente ignorado na prática: um consentimento genérico para "uso dos meus dados para fins de comunicação" não autoriza o tratamento para microtargeting psicográfico, para compartilhamento com terceiros prestadores de serviços de análise ou para uso em treinamento de modelos de IA. Cada finalidade distinta exige consentimento específico. Para dados sensíveis, incluindo os perfis psicográficos que revelam orientação política por inferência, o artigo 11º, inciso I, exige consentimento específico e destacado: ele não pode estar enterrado em cláusulas gerais de termos de uso, mas deve ser solicitado de forma separada e explícita. O consentimento tem ainda duas fragilidades estruturais que o tornam base legal menos sólida do que aparenta para campanhas de longo prazo: pode ser revogado pelo titular a qualquer momento, com efeito imediato sobre as operações de tratamento em andamento, e a prova de que foi obtido validamente recai integralmente sobre o controlador.

O cumprimento de obrigação legal ou regulatória, previsto no artigo 7º, inciso II, tem aplicação eleitoral relevante mas restrita. Partidos políticos têm obrigações legais de registro de filiados, manutenção de cadastros eleitorais e prestação de contas que envolvem tratamento de dados pessoais com base nessa hipótese. A base legal, contudo, não autoriza o tratamento para finalidades que vão além do cumprimento da obrigação específica que a fundamenta: o dado coletado para cadastro de filiado não pode ser utilizado para microtargeting eleitoral com fundamento em obrigação legal, porque a obrigação de registro de filiados não inclui autorização para tratamento analítico do comportamento político dos titulares. A confusão entre a base legal que autoriza a coleta e a base legal que autoriza usos subsequentes dos mesmos dados é um dos erros mais comuns de compliance em campanhas eleitorais.

O legítimo interesse do responsável pelo tratamento, previsto no artigo 7º, inciso IX, é a base legal mais flexível disponível para dados pessoais comuns e a mais

frequentemente invocada por campanhas para justificar operações de análise de dados. Para ser válido, precisa passar por um teste de três etapas que a LGPD não enuncia explicitamente mas que a doutrina e a ANPD reconhecem como implícito na norma: o interesse do controlador deve ser legítimo; o tratamento deve ser necessário para realizá-lo; e os interesses do controlador não devem prevalecer sobre os direitos e liberdades fundamentais dos titulares. No contexto eleitoral, o interesse de um partido em comunicar suas propostas ao eleitorado é, em abstrato, legítimo. A questão é se o tratamento de dados comportamentais em escala para construção de perfis psicográficos e segmentação psicológica de mensagens passa no segundo e no terceiro testes: é esse tratamento necessário para a comunicação política legítima, ou existem meios menos intrusivos de atingir a mesma finalidade? E os direitos dos titulares à autodeterminação informativa e à proteção contra manipulação prevalecem sobre o interesse da campanha em personalizar sua comunicação?

Como observa Pinheiro (2021), a aplicação do teste de legítimo interesse em contexto político é particularmente sensível porque a finalidade do tratamento, influenciar preferências eleitorais, está na fronteira entre persuasão legítima e manipulação que a doutrina da proteção de dados ainda não delimitou com precisão suficiente. Operações de microtargeting que exploram vulnerabilidades psicológicas documentadas, como os vieses cognitivos identificados por Kahneman, provavelmente não passam no terceiro teste, porque colocam os interesses do controlador em posição de prevalência sobre os direitos fundamentais dos titulares à autodeterminação e à liberdade de escolha política genuína (PINHEIRO, 2021, p. 134-139).

A implicação prática mais urgente dessa análise já foi antecipada na seção anterior: o legítimo interesse não está disponível como base legal para dados sensíveis, incluindo perfis psicográficos que revelam orientação política por inferência. Para essas operações, a única base legal viável na maioria dos cenários práticos é o consentimento específico e destacado do artigo 11º, inciso I. Campanhas que construíram sua arquitetura de dados sobre a premissa de que o legítimo interesse autoriza qualquer análise comportamental de eleitores precisam revisar essa premissa antes de 2026, quando a cooperação entre ANPD e TSE estará mais desenvolvida e a exposição regulatória por bases legais incorretas será mais provável de ser identificada e sancionada.

O quadro a seguir traduz as quatro bases legais mais relevantes para o contexto eleitoral em termos operacionais, especificando para cada uma a aplicação concreta em campanhas e as limitações que determinam quando ela pode e quando não pode ser

invocada. A coluna de limitações é a mais importante para a estratégia de compliance: ela revela que nenhuma das quatro bases tem alcance irrestrito, e que o erro mais frequente na prática não é desconhecer as bases disponíveis, mas superestimar o alcance de uma delas para cobrir operações que ela não autoriza. O consentimento parece amplo mas é frágil pela revogabilidade e pela exigência de especificidade por finalidade. O cumprimento de obrigação legal parece seguro mas é estreito, cobrindo apenas as obrigações específicas que o fundamentam. O legítimo interesse parece flexível mas está vedado para dados sensíveis e exige teste de balanceamento que operações de microtargeting psicográfico frequentemente não passam. O exercício regular de direitos, a base menos conhecida entre as quatro, tem aplicação relevante em contexto de litígio eleitoral mas não autoriza o tratamento estratégico anterior ao processo. A leitura integrada do quadro confirma a conclusão desenvolvida na seção anterior: campanhas que utilizam perfis psicográficos inferidos para microtargeting sem consentimento específico e destacado dos titulares não têm base legal adequada disponível em nenhuma das hipóteses do artigo 7º, o que torna o tratamento ilícito independentemente da sofisticação da arquitetura de dados construída sobre ele.

Tabela 43. Bases Legais para Tratamento de Dados em Campanhas Eleitorais

Base Legal	Previsão Legal	Aplicação em Campanhas	Limitações
Consentimento (art. 7º, I)	Manifestação livre, informada e inequívoca do titular.	Tratamento de dados coletados diretamente de eleitores para fins de microtargeting; dados sensíveis (exige consentimento específico e destacado).	Revogável a qualquer tempo; não pode ser condicionado à participação em evento de campanha.
Cumprimento de obrigação legal (art. 7º, II)	Necessidade de cumprir obrigação legal ou regulatória.	Divulgação de doadores na prestação de contas; transparência de dados de candidatos.	Limitado às obrigações legais específicas; não autoriza tratamento para finalidades eleitorais estratégicas.
Legítimo interesse (art. 7º, IX)	Legítimo interesse do controlador ou de terceiro.	Uso de dados públicos (TSE, registros eleitorais) para análise estatística de tendências; monitoramento de redes sociais para fins de segurança da campanha.	Não se aplica a dados sensíveis; exige teste de balanceamento entre o interesse e os direitos do titular.

Exercício regular de direitos (art. 7º, VI)	Em processo judicial, administrativo ou arbitral.	Uso de dados em investigações eleitorais; produção de prova em AIJE.	Restrito ao contexto do processo; não autoriza uso estratégico anterior.

Fonte: Elaborado pelo autor com base em Bioni (2021), Pinheiro (2021) e LGPD (Lei nº 13.709/2018).

Direitos dos Titulares e suas Implicações para Campanhas

A LGPD confere aos titulares de dados um conjunto robusto de direitos que as campanhas eleitorais são obrigadas a observar e que, na prática, são sistematicamente subestimados por assessores jurídicos que tratam a proteção de dados como exercício de elaboração de política de privacidade e não como conjunto de obrigações operacionais concretas. A distinção é relevante: uma política de privacidade bem redigida que nenhum processo interno é capaz de cumprir não é conformidade, é aparência de conformidade, e a diferença entre as duas torna-se visível precisamente quando um titular exerce seus direitos e a campanha não tem condições de responder adequadamente.

Os direitos dos titulares estão previstos no artigo 18 da LGPD e incluem o direito de confirmação da existência de tratamento, o direito de acesso aos dados tratados, o direito de correção de dados incompletos ou inexatos, o direito de anonimização ou bloqueio de dados desnecessários ou tratados em desconformidade com a lei, o direito de portabilidade, o direito de eliminação dos dados tratados com base no consentimento, o direito de informação sobre as entidades com as quais os dados foram compartilhados, o direito de informação sobre a possibilidade de não fornecer o consentimento e as consequências dessa recusa, e o direito de revogação do consentimento. Para campanhas eleitorais que utilizam dados de eleitores em escala, esses direitos criam obrigações operacionais que precisam ser planejadas antes do período eleitoral, não improvisadas quando o primeiro pedido de exercício de direitos chegar.

O direito de acesso tem implicação particularmente sensível para operações de microtargeting: um eleitor que solicita confirmação da existência de tratamento e acesso aos seus dados tem o direito de saber que a campanha construiu um perfil psicográfico sobre ele, quais dados foram utilizados para isso e de onde vieram. Se a campanha não tem condições de responder a essa solicitação de forma completa e transparente dentro do prazo de quinze dias estabelecido pela ANPD, está em violação da LGPD independentemente de qualquer outra consideração sobre a licitude do tratamento em si.

A impossibilidade de responder ao direito de acesso é, ela própria, uma infração autônoma que se adiciona às infrações decorrentes do tratamento irregular.

O direito de eliminação dos dados tratados com base no consentimento é o que cria o risco operacional mais imediato para campanhas que construíram sua arquitetura de dados sobre essa base legal. Como o consentimento é revogável a qualquer tempo com efeito imediato, a revogação em massa por grupos de eleitores, coordenada por adversários políticos ou por organizações de proteção de dados, pode desmantelar a base de dados de uma campanha em momento crítico do ciclo eleitoral. Campanhas que dependem estruturalmente do consentimento como base legal única precisam de planos de contingência para essa hipótese, incluindo a avaliação de bases legais alternativas para subconjuntos dos dados que possam sustentá-las, e sistemas técnicos capazes de identificar e eliminar os dados de titulares específicos sem comprometer a integridade do conjunto remanescente.

O direito de oposição, que embora não listado explicitamente no artigo 18 decorre da estrutura geral da LGPD e foi reconhecido pela ANPD como direito implícito dos titulares, é especialmente relevante para operações baseadas no legítimo interesse. Quando a campanha invoca o legítimo interesse como base legal, o titular tem o direito de se opor ao tratamento, e a campanha precisa demonstrar que seus interesses legítimos prevalecem sobre os interesses do titular ou interromper o tratamento. Em contexto eleitoral, esse direito cria um mecanismo pelo qual eleitores que não querem ser objeto de microtargeting podem formalmente exigir que a campanha cesse o tratamento de seus dados para esse fim, com obrigação de resposta e cumprimento pela campanha.

A efetivação desses direitos exige que as campanhas mantenham registros de tratamento suficientemente detalhados para responder às solicitações dos titulares com precisão e dentro dos prazos legais. O artigo 37 da LGPD obriga os controladores a manter registro das operações de tratamento de dados pessoais, especialmente quando baseado no legítimo interesse. Para o contexto eleitoral, a Resolução TSE nº 23.610/2019, com a redação dada pelas resoluções subsequentes, incorporou essa obrigação ao determinar que candidatos e partidos mantenham o registro das operações de tratamento de dados pessoais durante o período eleitoral e por prazo adicional que permita a verificação em eventual processo judicial ou administrativo. O registro não é apenas uma obrigação de conformidade: é o instrumento que torna possível o exercício dos direitos dos titulares e a verificação pela ANPD e pela Justiça Eleitoral.

A questão que o advogado eleitoral precisa endereçar com antecedência é a seguinte: se todos os eleitores cujos dados a campanha trata exercessem simultaneamente seus direitos de acesso, confirmação e eliminação, a campanha teria condições operacionais de responder? Na maioria dos casos, a resposta honesta é não. Esse gap entre as obrigações legais e a capacidade operacional real é o risco de compliance mais subestimado pelas campanhas eleitorais brasileiras, e é o primeiro ponto que a ANPD e a Justiça Eleitoral examinarão quando a cooperação entre as duas instituições produzir as primeiras investigações conjuntas sobre uso de dados em campanhas.

O quadro a seguir organiza os cinco direitos dos titulares com maior relevância operacional para o contexto eleitoral, especificando como cada um se manifesta concretamente na relação entre eleitor e campanha e qual consequência operacional gera para a organização que trata seus dados.

A coluna de consequências para a campanha é a que mais diretamente interessa ao advogado eleitoral e ao gestor de compliance: ela revela que os direitos dos titulares não são abstrações jurídicas que a campanha observa declaratóriamente em sua política de privacidade, mas obrigações que exigem infraestrutura técnica e processual específica para serem cumpridas quando exercidas.

O direito de confirmação e acesso exige que a campanha saiba, em tempo real, quais dados de qual eleitor está tratando e seja capaz de comunicá-los de forma inteligível. O direito de eliminação exige sistemas técnicos capazes de localizar e excluir dados de um titular específico sem comprometer a integridade do conjunto remanescente.

O direito de oposição ao tratamento automatizado, o menos conhecido entre os cinco, é o que cria o vetor de accountability mais direto sobre sistemas de microtargeting: ele permite que o eleitor questione por que foi incluído num segmento específico de targeting e exija explicação sobre os critérios utilizados, o que é impossível de responder para campanhas que operam com sistemas de IA opacos sem documentação dos critérios de segmentação.

A leitura do quadro como conjunto confirma o ponto de encerramento da seção anterior: se todos os titulares exercessem simultaneamente os direitos aqui descritos, a maioria das campanhas eleitorais brasileiras não teria condições operacionais de responder adequadamente a nenhum deles.

Tabela 44. Direitos dos Titulares de Dados Relevantes para o Contexto Eleitoral

Direito	Previsão Legal	Exercício em Contexto Eleitoral	Consequência para a Campanha
Confirmação e acesso (art. 18, I e II)	Direito de saber se dados do eleitor estão sendo tratados e de acessá-los.	Eleitor pode solicitar à campanha confirmação sobre o tratamento de seus dados e receber cópia dos dados tratados.	Campanha deve manter registro de dados tratados e estar preparada para responder solicitações em prazo razoável.
Correção (art. 18, III)	Direito de corrigir dados incompletos, inexatos ou desatualizados.	Eleitor pode corrigir dados demográficos ou comportamentais incorretos em cadastros da campanha.	Campanha deve implementar processo de correção de dados.
Eliminação (art. 18, VI)	Direito de solicitar eliminação de dados tratados sem base legal adequada.	Eleitor pode solicitar eliminação de seus dados de sistemas de targeting da campanha.	Campanha deve atender a solicitações de eliminação e demonstrar que o tratamento tinha base legal.
Oposição ao tratamento automatizado (art. 20)	Direito de se opor a decisões tomadas unicamente com base em tratamento automatizado.	Eleitor pode questionar por que foi incluído em determinado segmento de targeting e exigir revisão.	Cria vetor de accountability para sistemas de microtargeting; campanha deve ser capaz de explicar critérios de segmentação.
Revogação do consentimento (art. 8º, §5º)	Direito de revogar consentimento a qualquer tempo.	Eleitor que havia consentido com uso de seus dados para targeting pode revogar esse consentimento.	Campanha deve cessar o tratamento imediatamente após a revogação.

Fonte: Elaborado pelo autor com base em Bioni (2021), Pinheiro (2021) e LGPD, art. 18.

Articulação entre ANPD e Justiça Eleitoral

A conformidade com a LGPD no contexto eleitoral envolve potencialmente duas autoridades com competências parcialmente sobrepostas, distintas em sua natureza institucional, em seus instrumentos de fiscalização e em suas lógicas sancionatórias, mas convergentes no objeto quando o fato ilícito é o uso irregular de dados pessoais de eleitores em campanha. Compreender a articulação entre a ANPD e a Justiça Eleitoral não é exercício teórico: é condição para que o advogado eleitoral antecipe os riscos de responsabilização do cliente com precisão, porque a dupla exposição regulatória pode

multiplicar as consequências de um único fato ilícito de forma que nem a doutrina de proteção de dados nem a doutrina eleitoral, analisadas isoladamente, permitem visualizar.

A ANPD tem competência para fiscalizar e sancionar violações à LGPD independentemente do contexto em que ocorram, incluindo o eleitoral. Sua atuação segue a lógica da regulação de proteção de dados: examina se o tratamento de dados pessoais teve base legal adequada, se os princípios da LGPD foram observados, se os direitos dos titulares foram respeitados e se as obrigações de segurança e governança foram cumpridas. O regime sancionatório da ANPD é gradual e progressivo, incluindo advertência com prazo para adoção de medidas corretivas, multa simples de até dois por cento do faturamento da pessoa jurídica no último exercício no Brasil limitada a cinquenta milhões de reais por infração, multa diária com o mesmo limite, publicização da infração após apuração e confirmação, bloqueio dos dados pessoais objeto da infração e eliminação dos dados pessoais objeto da infração. Para candidatos e partidos, que frequentemente não têm faturamento empresarial como parâmetro de cálculo, a ANPD ainda não consolidou o critério aplicável para o cálculo da multa percentual, o que é uma das lacunas operacionais da regulação que a cooperação com o TSE precisará endereçar.

A Justiça Eleitoral tem competência para fiscalizar e sancionar violações à regulação eleitoral, incluindo as obrigações relacionadas à proteção de dados incorporadas pela Resolução TSE nº 23.732/2024 e pela Resolução TSE nº 23.610/2019. Sua atuação segue a lógica da regulação eleitoral: examina se a conduta comprometeu a igualdade da disputa, se violou as vedações absolutas da Resolução, se o conteúdo irregular foi removido em tempo hábil e se há nexo entre a conduta e o benefício eleitoral do responsável. As sanções disponíveis para a Justiça Eleitoral no campo da proteção de dados incluem as multas específicas da Resolução TSE nº 23.732/2024, a determinação de eliminação de dados coletados irregularmente como sanção acessória, e, em casos que configurem abuso de poder ou uso indevido de meios de comunicação, a cassação do registro ou diploma com inelegibilidade por oito anos.

A sobreposição entre as duas competências não é acidental: ela é estrutural. Quando um candidato utiliza dados pessoais de eleitores sem base legal adequada para orientar microtargeting eleitoral, a conduta viola simultaneamente a LGPD, por ausência de base legal para o tratamento, e a regulação eleitoral, por uso de dados obtidos sem autorização para fins de propaganda política. Os dois fundamentos são distintos, as duas autoridades têm competência própria para agir e as duas sanções podem ser aplicadas cumulativamente, porque tutelam bens jurídicos diferentes: a ANPD tutela o direito

individual do titular à autodeterminação informativa, e a Justiça Eleitoral tutela a igualdade coletiva da disputa eleitoral. A cumulação de sanções por um único fato com dois fundamentos distintos é juridicamente possível e não configura bis in idem quando os bens jurídicos tutelados são diferentes, como a jurisprudência do STJ e do TSE tem reconhecido em outras hipóteses de sobreposição regulatória.

A articulação formal entre as duas autoridades ainda está em desenvolvimento, e essa lacuna tem consequências práticas que merecem reconhecimento. Na ausência de protocolo de cooperação consolidado, investigações paralelas pela ANPD e pela Justiça Eleitoral sobre o mesmo fato podem produzir resultados inconsistentes: uma autoridade pode concluir pela ausência de violação enquanto a outra conclui pela existência da mesma, sem mecanismo de coordenação que harmonize as conclusões. Investigações paralelas também impõem ao investigado o ônus de responder a dois processos com lógicas probatórias distintas, prazos diferentes e exigências de produção de prova que podem ser conflitantes. E a ausência de compartilhamento de inteligência entre as duas autoridades significa que evidências coletadas por uma delas podem não estar disponíveis para a outra, fragmentando o quadro probatório de operações que têm, por design, uma dimensão de proteção de dados e uma dimensão eleitoral indissociáveis.

Um protocolo de cooperação entre TSE e ANPD é, por essas razões, uma das demandas mais urgentes da agenda regulatória eleitoral brasileira para o ciclo de 2026. Os elementos que esse protocolo precisaria endereçar incluem o compartilhamento de informações sobre investigações em curso, a definição de critérios para a autoridade com maior conexão com o fato assumir a liderança da investigação sem prejuízo da competência da outra, a harmonização dos prazos de resposta para que o investigado não seja onerado por dois processos com cronogramas incompatíveis, e a adoção de parâmetros comuns para o cálculo de sanções quando o mesmo fato for sancionado pelas duas autoridades. A experiência europeia do Comitê Europeu para a Proteção de Dados, que coordena a atuação das autoridades nacionais de proteção de dados dos países membros quando uma operação de tratamento tem impacto transfronteiriço, oferece um modelo de coordenação que, adaptado ao contexto institucional brasileiro, poderia orientar a construção do protocolo TSE-ANPD (UNIÃO EUROPEIA, 2016, arts. 60-76).

Para o advogado eleitoral que atua hoje, antes que esse protocolo exista, a orientação prática é a seguinte: qualquer investigação que envolva uso de dados pessoais de eleitores em campanha deve ser tratada como potencialmente dual desde o início, com estratégia de defesa coordenada perante as duas autoridades e com atenção às

inconsistências que decisões independentes podem criar. Uma posição adotada no processo perante a ANPD pode ser usada pela Justiça Eleitoral, e vice-versa. A coordenação entre os advogados que atuam nas duas frentes não é uma comodidade: é uma necessidade estratégica que a ausência de protocolo formal entre as autoridades torna ainda mais urgente.

⚖ JURISPRUDÊNCIA TSE — Processo Administrativo ANPD nº 00261.001048/2023-03

EMENTA:

Investigação sobre o uso de dados pessoais de eleitores por empresa de marketing político digital sem a obtenção de consentimento adequado. O sistema utilizava dados de comportamento de navegação e interação em redes sociais para inferir perfis psicográficos incluindo orientação política dos titulares, sem que esses tivessem consentido especificamente com esse uso.

RESULTADO:

A ANPD concluiu que o tratamento era ilícito por ausência de base legal adequada para dados sensíveis e determinou a cessação do tratamento, a eliminação dos dados coletados e a aplicação de multa. O caso estabeleceu o precedente de que a inferência de opiniões políticas a partir de dados comportamentais não sensíveis configura tratamento de dado sensível para os fins da LGPD, posição que o TSE adotou como referência interpretativa para a aplicação da Resolução nº 23.732/2024.

Segurança da Informação e Vazamentos em Campanhas

Além das questões de licitude do tratamento analisadas nas seções anteriores, campanhas eleitorais que utilizam dados pessoais de eleitores de forma sistemática enfrentam uma dimensão de risco que é independente da conformidade com as bases legais: o risco de incidentes de segurança que comprometam a integridade, a confidencialidade ou a disponibilidade dos dados tratados. Uma campanha pode ter base legal adequada para cada operação de tratamento, ter obtido os consentimentos necessários, ter respeitado os direitos dos titulares e ainda assim ser responsabilizada pela ANPD e pela Justiça Eleitoral se um incidente de segurança expuser os dados de eleitores a acessos não autorizados. A segurança da informação não é consequência automática da conformidade jurídica: é uma dimensão autônoma de compliance que exige investimento técnico específico e independente.

O risco de vazamento de dados em campanhas tem três dimensões que precisam ser compreendidas separadamente porque cada uma exige respostas distintas. A primeira é a dimensão regulatória: a LGPD exige no artigo 46 que os controladores adotem medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou

qualquer forma de tratamento inadequado ou ilícito. O descumprimento dessa obrigação sujeita o controlador a responsabilização civil pelos danos decorrentes do vazamento, independentemente de culpa quando a atividade de tratamento implicar risco para os direitos dos titulares, e a sanções administrativas pela ANPD que se acumulam sobre as sanções decorrentes da violação que o vazamento eventualmente revela. Um vazamento que expõe dados tratados sem base legal adequada gera, portanto, dupla responsabilização: pela ausência de base legal e pela ausência das medidas de segurança que teriam prevenido a exposição.

A segunda dimensão é a reputacional, que em contexto eleitoral tem características específicas que a distinguem de vazamentos em outros setores. Quando dados pessoais de consumidores vazam de uma empresa, o dano reputacional é relevante mas contido: afeta a relação entre a empresa e seus clientes. Quando dados pessoais de eleitores vazam de uma campanha, o dano tem natureza política: ele revela que a campanha coletou e armazenou informações sobre eleitores que esses eleitores frequentemente desconheciam ter fornecido, que essas informações incluem dados comportamentais e psicográficos que os eleitores não saberiam ter sido objeto de análise, e que a campanha não foi capaz de proteger adequadamente informações que havia coletado sem o conhecimento pleno dos titulares. Em ambiente eleitoral polarizado, essa revelação pode ser instrumentalizada por adversários com efeito desproporcional ao dano técnico do vazamento, transformando um incidente de segurança num evento de campanha.

A terceira dimensão, a mais específica ao contexto eleitoral e a menos discutida na literatura de segurança da informação, é a dimensão estratégica da exposição dos dados. Um banco de dados de campanha que vaza pode conter não apenas informações pessoais de eleitores, mas informações sobre a estratégia da campanha: quais segmentos do eleitorado foram identificados como persuadíveis, quais mensagens foram testadas para cada perfil psicográfico, quais vulnerabilidades eleitorais a campanha identificou em suas análises e quais regiões foram priorizadas ou despriorizadas na alocação de recursos. Essas informações têm valor estratégico imenso para adversários que as recebam, seja pelo acesso direto ao banco de dados vazado, seja pela análise dos padrões de targeting que o vazamento revela. A proteção dos dados de eleitores e a proteção da estratégia da campanha são, nessa dimensão, a mesma obrigação técnica.

As boas práticas de segurança da informação aplicáveis a campanhas eleitorais derivam dos padrões gerais de segurança de dados, mas precisam ser calibradas para as características específicas do ambiente de campanha: equipes temporárias com alta

rotatividade, múltiplos prestadores de serviço externos com acesso a sistemas críticos, pressão de tempo que frequentemente leva à priorização de velocidade sobre segurança e ausência de infraestrutura de segurança institucional que organizações permanentes mantêm. Nesse contexto, quatro práticas são prioritárias. A criptografia de dados em repouso e em trânsito é a medida mais básica e a mais frequentemente negligenciada: dados de eleitores armazenados em planilhas não criptografadas em laptops de voluntários ou transmitidos por canais não seguros entre prestadores de serviço são o ponto de vulnerabilidade mais comum em campanhas de médio porte. O controle de acesso por privilégio mínimo, que limita o acesso de cada membro da equipe apenas aos dados necessários para suas funções específicas, reduz drasticamente a superfície de exposição em caso de comprometimento de uma credencial individual. As auditorias regulares de segurança, incluindo testes de penetração nos sistemas que tratam dados de eleitores, identificam vulnerabilidades antes que sejam exploradas por atores externos. E o plano de resposta a incidentes, que estabelece com antecedência os procedimentos a seguir em caso de vazamento, incluindo a obrigação de notificação à ANPD no prazo de setenta e duas horas estabelecido pela LGPD e a comunicação aos titulares afetados, é o instrumento que determina se um incidente se torna uma crise gerenciável ou um escândalo eleitoral.

A obrigação de notificação à ANPD no prazo de setenta e duas horas após o conhecimento do incidente, prevista no artigo 48 da LGPD, tem implicação processual específica em contexto eleitoral: em período de campanha, setenta e duas horas é um intervalo em que a notícia do vazamento pode ter circulado amplamente antes que qualquer resposta institucional da campanha esteja disponível. O plano de resposta a incidentes precisa, portanto, incluir não apenas o protocolo de notificação à ANPD, mas a estratégia de comunicação pública que acompanha essa notificação, de forma que a campanha controle a narrativa sobre o incidente em vez de reagir à narrativa construída por adversários e pela imprensa. Como em todos os aspectos da gestão de crises eleitorais, a resposta improvisada é sistematicamente inferior à resposta planejada, e o planejamento precisa ocorrer antes do incidente, não depois.

■ RESUMO DO CAPÍTULO:

Este capítulo analisou a articulação entre a Lei Geral de Proteção de Dados e o direito eleitoral como campos que precisam ser lidos de forma integrada, não como sistemas paralelos com pontos de contato ocasionais. A tese central que orientou toda a análise é que a proteção de dados pessoais dos eleitores não é obrigação paralela às obrigações eleitorais: é condição necessária para que o processo eleitoral cumpra sua função

democrática de permitir que cidadãos formem preferências políticas autênticas. A análise teórica inicial estabeleceu o fundamento dessa tese a partir de dois autores centrais. Solove demonstrou que a privacidade importa não porque protege segredos individuais, mas porque preserva o espaço necessário para a formação e expressão de identidades políticas sem a pressão do monitoramento. Zuboff demonstrou que o modelo econômico das plataformas tem como produto principal a modificação comportamental em escala, com consequências que se estendem ao campo político. Esses dois argumentos convergem para uma conclusão juridicamente relevante: uma campanha que viola a LGPD para orientar seu microtargeting não comete apenas infração administrativa de proteção de dados, mas compromete a qualidade epistêmica do processo pelo qual o eleitor forma suas preferências, que é dano ao fundamento da legitimidade eleitoral. A análise da LGPD como garantia democrática sistematizou três categorias de dados pessoais tratados por campanhas eleitorais com regimes jurídicos distintos: dados de identificação e contato, tratáveis com base no legítimo interesse ou no cumprimento de obrigação legal com transparência sobre a finalidade; dados comportamentais, significativamente mais restritos especialmente quando revelam opiniões políticas por inferência; e dados sensíveis, sujeitos ao regime mais restritivo do artigo 11º da LGPD que exclui o legítimo interesse como base legal disponível. A questão da inferência política foi identificada como o ponto de maior tensão jurídica do campo. A pesquisa de Kosinski, Stillwell e Graepel demonstrou que dados comportamentais aparentemente neutros são suficientes para construir perfis psicográficos que preveem orientação política com alta acurácia. A posição doutrinária mais sólida, apoiada em Bioni, é que o dado inferido sensível deve seguir o regime dos dados sensíveis independentemente da natureza dos dados que o geraram, porque o risco para o titular é idêntico. Isso significa que sistemas de microtargeting que constroem perfis psicográficos por inferência estão potencialmente tratando dados sensíveis sem base legal adequada quando invocam o legítimo interesse do artigo 7º em vez do consentimento específico do artigo 11º. A análise das bases legais identificou o legítimo interesse como a mais frequentemente invocada e a mais frequentemente mal aplicada em campanhas eleitorais. O teste de três etapas que a valida, legitimidade do interesse, necessidade do tratamento e não prevalência sobre direitos dos titulares, é de difícil passagem para operações de microtargeting psicográfico que exploram vulnerabilidades cognitivas documentadas. O consentimento, embora seja a única base legal viável para dados sensíveis inferidos, tem fragilidades estruturais pela revogabilidade e pela exigência de especificidade por finalidade que o tornam base menos sólida para arquiteturas de dados de longo prazo. Os direitos dos titulares foram analisados não como abstrações jurídicas mas como obrigações operacionais concretas que exigem infraestrutura técnica específica. O teste prático proposto, se todos os titulares exercessem simultaneamente seus direitos de acesso, confirmação e eliminação a campanha teria condições de responder?, foi identificado como o diagnóstico mais preciso do gap entre obrigações legais e capacidade operacional real, que é o risco de compliance mais subestimado pelas campanhas brasileiras. A articulação entre ANPD e Justiça Eleitoral foi analisada como sobreposição estrutural e não acidental: o mesmo fato ilícito viola simultaneamente a LGPD e a regulação eleitoral, porque tutela bens jurídicos distintos, a autodeterminação informativa individual e a igualdade coletiva da disputa. A cumulação de sanções não configura bis in idem. A ausência de protocolo de cooperação consolidado entre as duas autoridades cria três consequências práticas: inconsistência de resultados, duplo ônus processual e fragmentação probatória. O protocolo TSE-ANPD foi identificado como uma das demandas mais urgentes da agenda regulatória eleitoral para 2026. O capítulo encerrou com a análise da segurança da informação como dimensão autônoma de compliance, independente da conformidade com as bases legais. O risco de vazamento foi analisado em três dimensões: regulatória, pela responsabilização pela ausência de medidas de segurança do artigo 46 da LGPD; reputacional, pelo potencial de instrumentalização política do incidente em ambiente polarizado; e estratégica, pela exposição da arquitetura de targeting da campanha a adversários. A obrigação de notificação à ANPD em setenta e duas horas foi articulada com a necessidade de estratégia de comunicação pública integrada ao plano de resposta a incidentes.

QUESTÕES PARA REFLEXÃO E DEBATE

1. A tese central do capítulo é que violações à LGPD em contexto eleitoral comprometem não apenas direitos individuais de titulares, mas a qualidade epistêmica do processo pelo qual eleitores formam suas preferências políticas. Essa tese tem implicações normativas que vão além da responsabilização de campanhas: ela sugere que resultados eleitorais obtidos com microtargeting que violou a LGPD em escala suficiente para influenciar o resultado seriam, em alguma medida, democraticamente ilegítimos. O direito eleitoral brasileiro tem instrumentos para endereçar essa dimensão coletiva do dano, ou ela permanece juridicamente invisível?

2. A posição doutrinária sobre dados inferidos sensíveis, que defende que o dado inferido deve seguir o regime do dado que resulta da inferência, ainda não foi consolidada pela ANPD nem pelo TSE. Enquanto essa consolidação não ocorre, campanhas que operam nessa zona de incerteza não têm orientação clara sobre qual base legal utilizar. Como o direito deveria tratar a responsabilização de atores que agiram durante esse período de incerteza regulatória, especialmente se a interpretação adotada pela ANPD retroagir sobre práticas anteriores à sua consolidação?

3. O teste de legítimo interesse aplicado ao microtargeting eleitoral que explora vieses cognitivos documentados provavelmente não passa no terceiro critério, a não prevalência dos interesses do controlador sobre os direitos dos titulares. Mas toda comunicação política envolve algum grau de apelo a emoções e vieses cognitivos: a distinção entre persuasão legítima e manipulação que explora vulnerabilidades não é nítida. Como o direito deveria construir essa fronteira de forma operacionalizável, sem proibir toda comunicação política que apele a emoções ou sem permitir toda operação de targeting que possa ser descrita como persuasão?

4. A sobreposição de competências entre ANPD e Justiça Eleitoral cria dupla exposição regulatória para o mesmo fato, com sanções cumulativas que não configuram bis in idem porque os bens jurídicos tutelados são distintos. Do ponto de vista da efetividade regulatória, a cumulação de sanções é um incentivo poderoso à conformidade. Do ponto de vista da proporcionalidade, ela pode produzir punições desproporcionais ao dano causado. Como o protocolo TSE-ANPD deveria tratar essa tensão entre efetividade e proporcionalidade na definição das sanções cumuláveis?

5. O direito de oposição ao tratamento automatizado do artigo 20 da LGPD permite ao eleitor questionar por que foi incluído num segmento específico de targeting e exigir revisão. Para responder a essa solicitação, a campanha precisa ser capaz de explicar os critérios de segmentação de seus sistemas de IA, o que é impossível para sistemas de aprendizado de máquina com opacidade inerente. Como o direito deveria tratar a tensão entre o direito à explicação dos titulares e as limitações técnicas de sistemas de IA que produzem resultados não explicáveis em linguagem compreensível?

6. A revogação em massa e coordenada de consentimentos por grupos de eleitores, facilitada por organizações de proteção de dados ou instrumentalizada por adversários políticos, pode desmantelar a base de dados de uma campanha em momento crítico do ciclo eleitoral. Esse cenário representa o exercício legítimo de direitos individuais dos titulares ou uma forma de interferência coordenada no processo eleitoral que o direito deveria regulamentar? A resposta muda se a coordenação for realizada por atores estrangeiros?

7. A segurança da informação em campanhas eleitorais tem três dimensões, regulatória, reputacional e estratégica, sendo a terceira a mais específica ao contexto eleitoral: um vazamento expõe não apenas dados de eleitores mas a arquitetura de targeting da campanha a adversários. Esse tipo de vazamento, quando provocado deliberadamente por atores com acesso interno à campanha ou por invasão externa patrocinada por adversários, configura interferência ilícita no processo eleitoral, crime eleitoral ou apenas violação de direito civil, e quais os instrumentos jurídicos disponíveis para a campanha afetada?

8. O capítulo identificou que a maioria das campanhas eleitorais brasileiras não teria condições operacionais de responder adequadamente ao exercício simultâneo de direitos por todos os titulares de dados que tratam. Isso significa que a maioria das campanhas brasileiras está em violação estrutural da LGPD independentemente de qualquer análise sobre bases legais ou conteúdo dos dados. A ANPD e a Justiça Eleitoral deveriam priorizar a construção de capacidade de conformidade nas campanhas por meio de orientação e safe harbors antes de 2026, ou deveriam aguardar o ciclo eleitoral e sancionar as violações que se manifestem?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Sua campanha provavelmente já está em violação da LGPD. O problema é a base legal, não o conteúdo.

A maioria das campanhas eleitorais brasileiras que utiliza ferramentas de análise de dados e microtargeting invoca o legítimo interesse do artigo 7º, inciso IX, da LGPD como base legal para todas as suas operações de tratamento. Esse caminho é juridicamente inadequado para perfis psicográficos que revelam orientação política por inferência, que são dados sensíveis sujeitos ao artigo 11º e para os quais o legítimo interesse simplesmente não está disponível. A violação não está no conteúdo do targeting nem na ferramenta utilizada: está na ausência de base legal adequada para o tipo de dado que o sistema está produzindo. Antes de 2026, toda arquitetura de dados de campanha precisa ser revisada por advogado com conhecimento simultâneo de LGPD e direito eleitoral.

2. Consentimento específico é a única saída para microtargeting psicográfico. E ele tem fragilidades sérias.

Para sistemas que inferem orientação política a partir de dados comportamentais, a única base legal viável é o consentimento específico e destacado do artigo 11º, inciso I, da LGPD. Isso significa que o eleitor precisa ter consentido explicitamente, de forma separada e claramente identificável, com o uso de seus dados para construção de perfil psicográfico com finalidade eleitoral. Consentimento genérico em termos de uso não satisfaz esse requisito. A fragilidade estrutural do consentimento como base legal é que ele é revogável a qualquer tempo com efeito imediato, o que significa que uma campanha cuja base de dados depende inteiramente de consentimentos pode perder o amparo legal de toda sua operação de targeting se um número significativo de titulares revogar simultaneamente. Toda campanha que usa consentimento como base legal precisa de plano de contingência para esse cenário.

3. Faça o teste das 72 horas agora, não no dia do vazamento.

A LGPD exige notificação à ANPD no prazo de setenta e duas horas após o conhecimento de incidente de segurança. Em período de campanha, esse intervalo é suficiente para que a notícia do vazamento circule amplamente antes que qualquer resposta institucional esteja disponível. O plano de resposta a incidentes precisa existir antes do incidente, não ser improvisado depois. Ele deve incluir: o protocolo de notificação à ANPD com modelo de comunicação já aprovado, a estratégia de comunicação pública que acompanha a notificação, a identificação dos dados afetados e dos titulares a serem comunicados, e o protocolo de coordenação com a Justiça Eleitoral se os dados vazados forem relevantes para processos eleitorais em curso. Campanhas que improvisam em crise de dados perdem o controle da narrativa nas primeiras horas, que são as que mais importam.

4. Qualquer investigação sobre dados de campanha é potencialmente dual: ANPD e Justiça Eleitoral simultaneamente.

O mesmo fato, uso de dados de eleitores sem base legal adequada para microtargeting, pode ser investigado simultaneamente pela ANPD, por violação à LGPD, e pela Justiça Eleitoral, por violação à Resolução TSE 23.732/2024. As duas autoridades têm competências próprias, as duas podem aplicar sanções cumulativamente e a cumulação não configura bis in idem porque os bens jurídicos tutelados são distintos. Isso significa que a estratégia de defesa precisa ser coordenada desde o início entre advogados com expertise nas duas áreas: uma posição adotada perante a ANPD pode ser usada pela Justiça Eleitoral e vice-versa. A falta de coordenação entre as duas frentes de defesa é um erro que a ausência de protocolo formal entre as autoridades torna mais fácil de cometer e mais difícil de corrigir.

5. O direito de acesso do eleitor pode revelar que sua campanha não sabe o que está fazendo com os dados.

Quando um eleitor solicita confirmação da existência de tratamento e acesso aos seus dados, a campanha tem quinze dias para responder de forma completa e transparente. Se o sistema de targeting foi construído por uma agência externa com modelo de IA opaco, a campanha pode simplesmente não saber quais dados de quais eleitores está tratando, de onde eles vieram e como foram usados para segmentação. A impossibilidade de responder ao direito de acesso é infração autônoma que se acumula sobre as infrações decorrentes do tratamento irregular. A perguntar a fazer agora é: se o eleitor mais vigilante do seu eleitorado solicitasse hoje acesso completo a todos os dados que sua campanha tem sobre ele, você conseguiria responder em quinze dias com informação precisa e completa?

6. O vazamento de dados de campanha não expõe só eleitores. Expõe sua estratégia inteira ao adversário.

Um banco de dados de campanha que vaza contém muito mais do que informações pessoais de eleitores. Contém a arquitetura da estratégia: quais segmentos foram identificados como persuadíveis, quais mensagens foram testadas para cada perfil psicográfico, quais regiões foram priorizadas, quais vulnerabilidades do eleitorado foram mapeadas. Essas informações têm valor estratégico imenso para adversários. A criptografia de dados em repouso e em trânsito, o controle de acesso por privilégio mínimo e as auditorias regulares de segurança não são apenas obrigações regulatórias: são proteção da vantagem competitiva da campanha. Tratar segurança da informação como formalidade de compliance é subestimar o risco real, que não é apenas jurídico mas estratégico.

7. Seu fornecedor de dados não é o responsável legal. Você é.

Quando uma campanha contrata uma empresa de dados ou uma agência de marketing político que coleta e trata dados de eleitores em seu nome, a campanha é o controlador para fins da LGPD e a empresa é o operador. A responsabilidade primária pelo tratamento adequado dos dados é do controlador, não do operador. Isso significa que se a agência tratou dados sem base legal adequada, inferiu perfis psicográficos sem consentimento específico ou compartilhou dados com terceiros não autorizados, a campanha responde pela violação mesmo que o contrato atribua a responsabilidade ao fornecedor. A proteção contratual pode gerar ação regressiva contra o fornecedor, mas não afasta a responsabilidade perante a ANPD e a Justiça Eleitoral. Toda contratação de empresa de dados ou agência de marketing político precisa incluir auditoria prévia das práticas de tratamento de dados, cláusulas específicas de conformidade com a LGPD e direito de auditoria durante a execução do contrato.

CAPÍTULO 14

Ferramentas Práticas de Detecção e Monitoramento

Deepfakes, Desinformação, Análise de Sentimento e Auditoria do Ambiente Eleitoral Digital

Ferramentas Digitais e Integridade Eleitoral

O campo das ferramentas digitais para proteção da integridade eleitoral é, ao mesmo tempo, um dos mais dinâmicos e dos menos sistematizados da literatura especializada, e essa combinação cria um risco específico para o profissional que busca orientação prática: a abundância de ferramentas disponíveis, frequentemente promovidas com afirmações de eficácia não verificadas, pode gerar uma confiança desproporcional nos resultados que elas produzem. Este capítulo, portanto, não pretende ser um catálogo exaustivo de ferramentas, que estaria desatualizado antes de concluído, mas sim apresentar as categorias funcionais disponíveis, os princípios de uso responsável de cada categoria e, especialmente, os critérios para avaliação crítica dos resultados produzidos por cada ferramenta. A pergunta mais importante que o profissional precisa ser capaz de responder não é "qual ferramenta devo usar?", mas "como interpreto os resultados desta ferramenta e quais são seus limites de confiabilidade neste caso específico?".

A velocidade de desenvolvimento de novas ferramentas é impulsionada por uma dinâmica que Howard (2020) denomina de corrida armamentista assimétrica: agentes maliciosos têm incentivos econômicos e políticos para desenvolver ferramentas de produção de desinformação mais eficazes, enquanto defensores da integridade eleitoral desenvolvem ferramentas de detecção e verificação em resposta. A assimetria está em que o desenvolvimento ofensivo pode ocorrer de forma privada, sem divulgação pública de métodos, enquanto o desenvolvimento defensivo frequentemente depende de financiamento público ou filantrópico com ciclos de desenvolvimento mais lentos (HOWARD, 2020, p. 155-159). Essa assimetria não é argumento contra o desenvolvimento de ferramentas defensivas: é a premissa que determina como elas devem ser usadas, sempre com consciência de suas limitações e nunca como substituto do julgamento humano especializado.

A literatura de referência para este capítulo combina fontes técnicas com fontes práticas orientadas ao profissional não especializado. Entre as fontes técnicas, os trabalhos de Goodfellow et al. (2016) sobre redes adversárias generativas estabelecem o fundamento

teórico para compreender como deepfakes são produzidos e, por extensão, que tipos de artefatos os sistemas de detecção devem buscar. O trabalho de Vaswani et al. (2017) sobre a arquitetura Transformer fornece a base técnica para compreender como os modelos de linguagem que produzem texto sintético funcionam e por que sua detecção é estruturalmente mais difícil do que a detecção de deepfakes audiovisuais. Entre as fontes práticas, os manuais produzidos pela First Draft e pelo International Fact-Checking Network oferecem orientação operacional validada por profissionais que trabalham com verificação de conteúdo em contextos eleitorais reais. A tensão entre rigor técnico e acessibilidade prática é o desafio central deste capítulo, e sua resolução passa pela distinção clara entre o que o profissional precisa entender sobre como as ferramentas funcionam e o que precisa saber sobre como interpretar seus resultados.

Detecção de Deepfakes: Princípios, Ferramentas e Limitações

A detecção de deepfakes é um campo em permanente evolução caracterizado por uma dinâmica que define seu horizonte de eficácia: modelos de geração mais avançados frequentemente superam os modelos de detecção disponíveis, porque produzir conteúdo sintético convincente é um problema mais fácil do que identificá-lo com confiabilidade. Como documentaram Nightingale e Farid (2022) em pesquisa sobre detecção de faces sintéticas, mesmo sistemas de detecção especializados e observadores humanos treinados têm taxas de erro significativas diante de conteúdo produzido pelos modelos mais avançados. Qualquer profissional que trabalhe com detecção de deepfakes eleitorais deve internalizar essa limitação estrutural antes de confiar nos resultados de qualquer ferramenta específica, porque a consequência de um falso negativo, acreditar que um deepfake é autêntico, pode ser politicamente devastadora, e a consequência de um falso positivo, acusar de deepfake um conteúdo autêntico, pode ser juridicamente grave.

As principais abordagens técnicas de detecção organizam-se em quatro categorias de análise, cada uma com capacidades e limitações que o profissional não técnico precisa compreender para interpretar os resultados adequadamente.

A análise de artefatos visuais examina inconsistências na produção do conteúdo sintético que resultam das limitações dos modelos de geração: borramento nas bordas do rosto substituído, irregularidades na textura da pele que não correspondem à iluminação do ambiente, inconsistências na projeção de sombras, anomalias nos movimentos dos olhos e taxa de piscamento atípica, e descontinuidades entre o movimento da cabeça e o

movimento do corpo. Essa categoria de análise foi altamente eficaz para deepfakes de primeira e segunda geração, mas sua eficácia declinou substancialmente com o avanço dos modelos de difusão, que aprenderam a eliminar progressivamente os artefatos que os classificadores foram treinados para identificar. A análise de artefatos visuais é ainda útil como primeira triagem, mas resultados negativos, ausência de artefatos detectados, não constituem prova de autenticidade em conteúdo produzido por modelos avançados.

A análise biométrica examina indicadores fisiológicos de difícil replicação sintética, especialmente o padrão de variação sutil de coloração da pele facial que resulta do ritmo cardíaco, fenômeno denominado RPPG (remote photoplethysmography). Modelos sintéticos de primeira e segunda geração não replicavam esses padrões de forma convincente, tornando a análise biométrica uma forma de detecção relativamente robusta para esse período. A limitação atual é que modelos mais recentes, treinados com volumes maiores de dados fisiológicos, conseguem replicar padrões biométricos com fidelidade crescente, e que a análise RPPG exige qualidade de vídeo e condições de iluminação que frequentemente não estão disponíveis no conteúdo eleitoral que circula por WhatsApp e redes sociais, onde a compressão de arquivo degrada exatamente os dados que a análise biométrica precisa.

A análise de metadados examina as informações registradas no arquivo audiovisual em busca de inconsistências com o dispositivo de captura, as condições de gravação e o histórico de edição declarados. Um vídeo que afirma ter sido gravado por determinado modelo de smartphone em determinada data mas cujos metadados revelam software de edição profissional, timestamps inconsistentes ou ausência completa de metadados típicos de gravação original apresenta sinais de manipulação que não dependem da análise do conteúdo em si. A limitação da análise de metadados é que metadados são facilmente manipuláveis por atores com conhecimento técnico básico, e que conteúdo amplamente compartilhado frequentemente perde seus metadados originais nas plataformas de distribuição, tornando a análise inconclusiva para conteúdo que já circulou por múltiplos intermediários.

A análise por modelos de aprendizado de máquina utiliza classificadores treinados especificamente para identificar padrões de geração sintética que escapam à percepção humana, como regularidades estatísticas na distribuição de pixels, padrões de frequência na transformada de Fourier do sinal de imagem e outras características de baixo nível que diferem sistematicamente entre conteúdo orgânico e sintético. Essa categoria é a mais poderosa disponível atualmente e a base das ferramentas de detecção comerciais mais

avançadas, mas também a mais vulnerável à degradação por counter-adversarial training: quando os modelos de geração são treinados especificamente para enganar os classificadores disponíveis, produzem conteúdo que sistematicamente evita os padrões que os classificadores identificam. A corrida armamentista é mais visível e mais acelerada nessa categoria do que em qualquer outra.

O princípio que deve orientar o uso de qualquer dessas ferramentas em contexto eleitoral é o da corroboração múltipla: nenhum resultado de ferramenta isolada é suficiente para conclusões definitivas sobre a autenticidade ou a natureza sintética de um conteúdo. A combinação de análise por múltiplas categorias, com resultados convergentes, aumenta significativamente a confiabilidade da conclusão, mas mesmo a convergência de múltiplos indicadores não substitui a análise pericial forense quando o resultado for utilizado como prova em processo eleitoral. A ferramenta é instrumento de triagem e de orientação da investigação: a prova é o laudo pericial produzido por especialista com metodologia documentada e reproduzível.

O quadro a seguir sistematiza seis ferramentas de detecção com diferentes perfis de capacidade, acesso e finalidade, permitindo ao profissional selecionar a combinação adequada a cada situação concreta. A leitura por colunas revela o critério mais importante para essa seleção: não é a sofisticação técnica da ferramenta, mas a adequação entre sua finalidade declarada e o uso que o profissional pretende fazer do resultado. A coluna de uso recomendado é, por essa razão, a mais operacionalmente relevante do quadro. Ferramentas de triagem rápida como o Deepware Scanner e o InVID/WeVerify são adequadas para orientar a decisão sobre se um conteúdo merece investigação mais aprofundada, mas seus resultados não têm validade jurídica e não substituem análise forense especializada.

A análise forense do TSE ou da Polícia Federal produz o único resultado com cadeia de custódia e validade como prova em processo eleitoral, mas com prazo estruturalmente incompatível com a urgência do ambiente eleitoral, o que significa que ela deve ser acionada em paralelo com as ferramentas de triagem, não em substituição a elas. A coluna de limitação principal confirma o argumento desenvolvido na seção anterior: nenhuma ferramenta tem cobertura universal, e a combinação de múltiplas ferramentas com resultados convergentes é sempre mais confiável do que a confiança em qualquer ferramenta isolada. O resultado mais importante que o quadro entrega, porém, não está em nenhuma coluna específica: está na linha da forense especializada, que evidencia que o único instrumento capaz de produzir prova juridicamente válida sobre a natureza sintética

de um conteúdo eleitoral é também o único com prazo incompatível com a urgência do pleito, o que é a contradição estrutural mais relevante do campo e a que mais diretamente determina a estratégia processual do advogado eleitoral.

Tabela 45. Ferramentas de Detecção de Deepfakes: Capacidades e Limitações

Ferramenta	Tipo	O que Detecta	Limitação Principal	Acesso	Uso Recomendado
Microsoft Video Authenticator	Comercial (gratuita)	Artefatos visuais de manipulação por IA; fornece score de probabilidade de autenticidade.	Treinada em modelos específicos; pode não detectar modelos mais recentes.	Portal Azure; integração via API.	Triagem inicial de vídeos suspeitos antes de análise forense aprofundada.
Deepware Scanner	Online (gratuita)	Detecção de deepfakes de vídeo utilizando múltiplos modelos de classificação.	Limitação de tamanho de arquivo; pode gerar falsos negativos para deepfakes de alta qualidade.	deepware.ai	Verificação rápida de vídeos curtos recebidos por WhatsApp ou redes sociais.
FotoForensics	Online (gratuita)	Análise de metadados e padrões de compressão de imagens para identificar manipulações.	Focada em imagens estáticas; não analisa vídeo.	fotoforensics.com	Verificação de imagens estáticas suspeitas em material eleitoral.
InVID/WeVerify	Extensão de navegador (gratuita)	Análise forense rápida de vídeos e imagens em redes sociais; busca reversa de imagens.	Dependente de banco de dados de referências; limitada para conteúdo novo.	Plugin para Chrome e Firefox.	Verificação de contexto e origem de conteúdo eleitoral suspeito.

Sensity AI	Comercial (pago)	Detecção de deepfakes em tempo real em múltiplas plataformas; alertas automáticos.	Custo elevado; adequado para operações de grande porte.	Por contrato.	Monitoramento contínuo do ambiente digital em campanhas presidenciais.
Forense especializada (TSE/Polícia Federal)	Institucional	Análise técnica completa com cadeia de custódia; resultado com validade jurídica.	Disponível apenas para investigações formais; prazo não compatível com urgência eleitoral.	Por requisição formal.	Produção de prova em processos eleitorais após identificação de conteúdo suspeito.

Fonte: Elaborado pelo autor com base em Paris e Donovan (2019), Chesney e Citron (2019) e documentação das ferramentas (2024).

ATENÇÃO PRÁTICA — Limitação Fundamental que Todo Profissional Deve Conhecer

Nenhuma ferramenta de detecção de deepfakes disponível comercialmente oferece certeza absoluta. O resultado de qualquer dessas ferramentas deve ser tratado como indício probabilístico, não como prova. Para fins jurídicos eleitorais, o laudo de forense digital com validade probatória deve ser produzido por especialista credenciado, com documentação adequada da cadeia de custódia do arquivo analisado. A urgência da remoção de conteúdo suspeito pode ser fundamentada em resultado de ferramenta comercial combinado com análise humana qualificada, mas a responsabilização penal ou eleitoral exige prova técnica de padrão superior.

Monitoramento de Redes Sociais: Social Listening Eleitoral

O monitoramento de redes sociais para fins eleitorais, denominado social listening eleitoral, é uma categoria de ferramentas com dupla finalidade estratégica que precisam ser compreendidas separadamente por que implicam usos, responsabilidades e limitações distintos. A primeira finalidade é ofensiva ou competitiva: auxilia campanhas na identificação de tendências de opinião pública em formação, crises de imagem antes que atinjam o pico de disseminação, narrativas adversárias em ascensão e oportunidades de comunicação baseadas em temas que mobilizam organicamente o eleitorado-alvo. A segunda finalidade é defensiva ou institucional: auxilia a Justiça Eleitoral, pesquisadores e organizações de fact-checking na detecção de operações coordenadas de desinformação,

identificação de redes de amplificação artificial e monitoramento do cumprimento de ordens de remoção por plataformas.

A distinção entre as duas finalidades não é apenas descritiva: ela tem implicações jurídicas diretas. Uma campanha que utiliza social listening para identificar tendências de opinião e ajustar sua comunicação está exercendo estratégia de comunicação política legítima. Uma campanha que utiliza social listening para identificar eleitores persuadíveis e orientar microtargeting psicográfico está potencialmente tratando dados pessoais sem base legal adequada, com as implicações analisadas no capítulo anterior. A ferramenta é a mesma; o uso determina o regime jurídico aplicável.

As ferramentas de social listening operam por um processo de três etapas que o profissional precisa compreender para avaliar a confiabilidade dos resultados que recebe. A primeira etapa é a coleta, que agrega em tempo quase real dados de plataformas abertas por meio de APIs (interfaces de programação de aplicativos) que cada plataforma disponibiliza em condições variáveis. A segunda etapa é o processamento, que aplica técnicas de Processamento de Linguagem Natural para identificar temas, sentimentos, entidades mencionadas e padrões de disseminação nos dados coletados. A terceira etapa é a visualização, que apresenta os resultados em painéis que permitem análise de tendências temporais, geográficas e demográficas.

A disponibilidade e a qualidade dos dados variam significativamente entre plataformas, e essa variação é o fator que mais limita a confiabilidade do social listening eleitoral na prática. O Twitter/X foi historicamente a plataforma mais aberta a pesquisadores e ferramentas de monitoramento, com APIs que permitiam acesso a volumes significativos de dados públicos. Após a mudança de gestão em 2022 e a subsequente reformulação do acesso à API com cobrança de tarifas elevadas, o volume de dados do Twitter/X acessível a ferramentas de social listening de menor porte reduziu-se dramaticamente, comprometendo a continuidade de séries históricas de pesquisa e aumentando o custo de monitoramento em tempo real. O Facebook público e o Instagram oferecem acesso restrito por meio da API do Graph, com limitações que tornam impossível o monitoramento sistemático de grupos e páginas sem autorização específica. O TikTok tem API de pesquisa disponível para pesquisadores credenciados, mas com restrições significativas sobre volume e frequência de consultas. O Telegram público, paradoxalmente, oferece acesso mais aberto do que a maioria das grandes plataformas para canais e grupos públicos, o que o torna particularmente relevante para o monitoramento de desinformação eleitoral no

contexto brasileiro, onde o aplicativo tem presença significativa em ecossistemas de comunicação política.

O WhatsApp, o canal de desinformação mais relevante no contexto eleitoral brasileiro como documentado nos capítulos anteriores, é estruturalmente inacessível ao social listening convencional: sua criptografia de ponta a ponta e a natureza privada de suas comunicações impedem qualquer monitoramento sistemático por ferramentas externas. Essa limitação não é técnica no sentido de que poderia ser superada por ferramentas mais sofisticadas: é constitutiva do design de privacidade do aplicativo. O monitoramento de desinformação no WhatsApp depende de metodologias alternativas, como grupos de monitoramento colaborativo em que pesquisadores participam de grupos públicos para coletar amostras, sistemas de denúncia por usuários como o tipline da Agência Lupa e o serviço do Comprova, e análise de conteúdo que migra do WhatsApp para plataformas públicas após circular por grupos privados. Nenhuma dessas metodologias produz cobertura comparável ao monitoramento de plataformas abertas, o que é o argumento mais concreto disponível para a tese de que a desinformação eleitoral mais relevante no Brasil é precisamente a que o sistema de monitoramento menos consegue alcançar.

Três capacidades específicas do social listening eleitoral merecem destaque pela sua relevância para a proteção da integridade eleitoral. A primeira é a detecção de coordenação inautêntica, que identifica padrões de disseminação que não correspondem a compartilhamento orgânico: picos de volume impossíveis sem automação, uniformidade temporal de publicações que sugere disparo coordenado, redes de contas com padrões de comportamento atípicos como criação recente, ausência de histórico ou atividade concentrada em janelas temporais estreitas. Como demonstraram Araújo e Oliveira (2021) em análise de redes coordenadas de amplificação nas eleições brasileiras, a diferença entre compartilhamento orgânico e operação coordenada é frequentemente identificável por análise de padrões temporais e estruturais que escapam à percepção individual mas são detectáveis por ferramentas de análise de redes (ARAÚJO; OLIVEIRA, 2021, p. 112-118).

A segunda capacidade é o monitoramento de narrativas emergentes, que permite identificar desinformação em estágio inicial de disseminação, quando a intervenção de fact-checking ou de direito de resposta ainda pode ser eficaz antes que a narrativa atinja saturação. A terceira é o mapeamento de influenciadores e de estrutura de redes, que identifica os nós de maior influência nos ecossistemas de comunicação política relevantes para determinado eleitorado, permitindo tanto estratégias de comunicação mais eficientes

quanto a identificação de atores que podem estar sendo utilizados como vetores de disseminação de desinformação.

O quadro a seguir organiza cinco ferramentas de social listening por perfil de uso, custo e cobertura, permitindo ao profissional identificar a opção mais adequada ao porte da campanha e aos objetivos específicos do monitoramento. A leitura por linhas revela uma progressão que vai do monitoramento profissional de alto custo ao monitoramento técnico de custo zero, com trade-offs distintos em cada ponto.

A coluna de custo aproximado é a primeira que a maioria dos profissionais consulta, mas a coluna de melhor uso é a que determina se o investimento é justificado: uma ferramenta de alto custo como o Brandwatch é adequada para campanhas presidenciais que precisam de monitoramento profissional em múltiplas plataformas simultaneamente, mas é desproporcional para uma campanha municipal que precisa apenas acompanhar menções ao candidato nas redes sociais locais.

A entrada do CrowdTangle merece atenção especial: sua descontinuação para uso geral em 2024 eliminou o principal instrumento gratuito de acesso a dados do ecossistema Meta para pesquisadores não credenciados, criando uma lacuna relevante no monitoramento de desinformação eleitoral que circula por Facebook e Instagram, os dois canais mais utilizados pelo eleitorado brasileiro acima de quarenta anos.

A última linha, Python com bibliotecas abertas, revela o dado mais operacionalmente relevante do quadro para equipes com capacidade técnica: as ferramentas de maior flexibilidade e menor custo são também as de maior barreira de entrada, porque exigem competência técnica em programação que a maioria das equipes de campanha não tem internamente.

A conclusão prática que o quadro entrega é que o social listening eleitoral eficaz em 2026 exigirá, para a maioria das campanhas, alguma combinação entre ferramenta comercial para cobertura ampla e automatizada e competência técnica interna ou terceirizada para análise aprofundada de padrões que as ferramentas comerciais identificam mas não interpretam.

Tabela 46. Ferramentas de Social Listening Eleitoral

Ferramenta	Tipo	Plataformas Monitoradas	Funcionalidades Principais	Custo Aproximado	Melhor Uso
Brandwatch	Comercial	Twitter/X, Facebook, Instagram, YouTube, Reddit, blogs, notícias.	Análise de sentimento; detecção de tendências; identificação de influenciadores; alertas em tempo real.	A partir de US$ 1.000/mês.	Campanhas presidenciais e estaduais com orçamento para monitoramento profissional.
Mention	Comercial	Redes sociais, blogs, fóruns, notícias.	Monitoramento de menções; análise de sentimento básica; relatórios automatizados.	A partir de US$ 41/mês.	Campanhas municipais e equipes pequenas que precisam de monitoramento básico.
Talkwalker	Comercial	Mais de 150 fontes incluindo redes sociais, TV e rádio.	Análise visual (imagens e vídeos); detecção de crises; analytics avançado.	Sob consulta.	Monitoramento de imagem do candidato incluindo conteúdo visual.
CrowdTangle (Meta)	Institucional (para pesquisadores)	Facebook e Instagram públicos.	Análise de conteúdo viral; rastreamento de grupos públicos; dados de engajamento.	Gratuito para pesquisadores credenciados; descontinuado para uso geral em 2024.	Pesquisa acadêmica sobre desinformação eleitoral no ecossistema Meta.
Python + Tweepy/snscrape	Código aberto (gratuito)	Twitter/X (com limitações após 2023); Telegram público; outras com APIs.	Coleta customizada; análise com bibliotecas de NLP; visualização com matplotlib/seaborn.	Gratuito (requer conhecimento técnico).	Pesquisadores e campanhas com equipe técnica própria.

Fonte: Elaborado pelo autor com base em Matejic (2018) e documentação das ferramentas (2024).

Ferramentas de NLP para Análise de Discurso Político

O Processamento de Linguagem Natural oferece ao profissional de campanha e ao pesquisador eleitoral um conjunto de ferramentas para análise sistemática de grandes volumes de texto político que seriam impossíveis de processar manualmente na escala e na velocidade que o ambiente eleitoral exige. Discursos, publicações em redes sociais, comentários de eleitores, notícias e comunicados de campanha produzem, ao longo de um ciclo eleitoral, volumes de texto da ordem de dezenas de milhões de documentos. A análise qualitativa manual, mesmo com equipes grandes, alcança amostras que representam uma fração mínima desse volume. As ferramentas de NLP permitem que análises quantitativas sistemáticas sejam realizadas sobre o corpus completo, identificando padrões que escapariam à análise amostral e produzindo evidências que podem orientar tanto decisões estratégicas de campanha quanto investigações de integridade eleitoral.

Duas aplicações são particularmente relevantes para o contexto eleitoral e merecem desenvolvimento específico antes da análise das ferramentas disponíveis. A análise de sentimento classifica automaticamente textos por tonalidade emocional, tipicamente em categorias de positivo, negativo e neutro, com variações que incluem emoções específicas como raiva, medo, entusiasmo e tristeza.

No contexto eleitoral, ela permite monitorar em tempo quase real como o eleitorado está respondendo emocionalmente a eventos de campanha, declarações de candidatos e coberturas jornalísticas, identificando crises de imagem antes que atinjam o pico de disseminação e oportunidades de comunicação que mobilizam resposta emocional positiva. A limitação mais relevante da análise de sentimento para o contexto eleitoral é sua sensibilidade ao sarcasmo, à ironia e ao humor político, que são formas de comunicação especialmente frequentes no discurso político brasileiro e que classificadores treinados em textos gerais frequentemente interpretam de forma incorreta. Um tuíte que usa ironia para criticar um candidato pode ser classificado como positivo por um sistema que não captura a inversão semântica, comprometendo a confiabilidade de análises que dependem dessa classificação.

A detecção automatizada de desinformação aplica modelos de classificação para identificar, em grandes volumes de texto, conteúdo com características linguísticas associadas à desinformação: afirmações factuais verificáveis sem fonte citada, linguagem emocionalmente carregada que contrasta com o tom neutro típico de relatos factuais, padrões de amplificação que sugerem coordenação e divergência com fatos estabelecidos em bases de conhecimento verificadas.

Essa aplicação é analiticamente promissora mas operacionalmente mais complexa do que a análise de sentimento, porque a desinformação é uma categoria que exige julgamento sobre a relação entre conteúdo e realidade, não apenas sobre características formais do texto. Modelos que identificam desinformação apenas por características linguísticas sem acesso a bases de conhecimento verificado têm taxas de falso positivo que os tornam inadequados para uso como único instrumento de triagem.

Para o contexto do português brasileiro, o modelo BERTimbau representa o estado da arte disponível para tarefas de análise de texto. Desenvolvido pelo NLP Group do Instituto de Ciências Matemáticas e de Computação da USP e disponibilizado publicamente pela plataforma Hugging Face, o BERTimbau é um modelo de linguagem pré-treinado em grande corpus de texto em português brasileiro que pode ser ajustado para tarefas específicas como análise de sentimento, classificação de conteúdo e detecção de padrões discursivos.

Sua vantagem sobre modelos treinados em inglês e adaptados para o português é que captura as especificidades semânticas, sintáticas e pragmáticas do português brasileiro, incluindo gírias, regionalismos e expressões idiomáticas que modelos multilíngues genéricos frequentemente processam de forma inadequada. A relevância dessa especificidade para o contexto eleitoral é direta: o discurso político brasileiro é rico em expressões regionais, referências culturais específicas e construções sintáticas que modelos não treinados no corpus brasileiro processam com qualidade significativamente inferior.

A utilização do BERTimbau e de modelos similares requer conhecimentos básicos de programação em Python e familiaridade com as bibliotecas Transformers e Hugging Face, mas tutoriais acessíveis e documentação detalhada estão disponíveis nas plataformas oficiais. Para equipes de campanha sem capacidade técnica interna, a alternativa mais acessível é a contratação de consultoria especializada em NLP aplicado ao contexto eleitoral, campo que tem crescido significativamente no Brasil a partir de 2018 e que conta com profissionais com experiência nos ciclos eleitorais anteriores.

A terceira opção, para pesquisadores e organizações com acesso a financiamento adequado, é o uso de modelos de linguagem comerciais de alta capacidade via API, como os disponibilizados pela Anthropic, OpenAI e Google, que oferecem capacidade de análise de texto sem necessidade de conhecimento técnico de implementação, mas com custos que escalam proporcionalmente ao volume de texto processado e com limitações de

privacidade que precisam ser avaliadas quando o texto a analisar contém dados pessoais de eleitores.

O princípio que deve orientar o uso de qualquer ferramenta de NLP em contexto eleitoral é o mesmo que orienta o uso de ferramentas de detecção de deepfakes: os resultados são insumo para julgamento humano especializado, não substituto dele. Um modelo de análise de sentimento que classifica noventa por cento dos textos corretamente ainda classifica dez por cento incorretamente, e em corpus de milhões de documentos isso representa centenas de milhares de classificações erradas que podem comprometer análises baseadas exclusivamente nos resultados automatizados. A combinação de análise automatizada em larga escala com validação humana de amostras representativas é o protocolo que melhor equilibra escala e confiabilidade nos resultados.

O quadro a seguir organiza cinco ferramentas de NLP por tarefa, nível técnico exigido e aplicação eleitoral específica, funcionando como mapa de navegação para profissionais com diferentes perfis de capacidade técnica. A coluna de nível técnico necessário é a primeira que determina quais linhas são operacionalmente acessíveis para cada equipe: as ferramentas de nível básico, como o VADER e a Google Natural Language API, podem ser utilizadas por qualquer profissional com acesso à internet e disposição para seguir documentação, enquanto as de nível intermediário a avançado, como o Hugging Face Transformers, exigem proficiência em Python e familiaridade com conceitos de aprendizado de máquina que a maioria das equipes de campanha não tem internamente.

A leitura por linha do BERTimbau confirma a orientação da seção anterior: para análise de texto político em português brasileiro, ele é a opção de maior acurácia disponível gratuitamente, e sua utilização deve ser priorizada sobre modelos multilíngues genéricos sempre que a equipe tiver capacidade técnica para implementá-lo. A linha do Hugging Face Transformers revela a capacidade mais avançada disponível no quadro e a de maior potencial para o contexto eleitoral de 2026: o fine-tuning de modelos pré-treinados para detecção de desinformação específica do contexto brasileiro permite criar classificadores altamente precisos para narrativas, expressões e padrões discursivos que modelos genéricos não capturam adequadamente.

A linha da Google Natural Language API representa o ponto de entrada mais acessível para equipes sem capacidade técnica, com a ressalva já identificada na seção anterior sobre custos que escalam com o volume e limitações de privacidade que precisam ser avaliadas quando o corpus contém dados pessoais de eleitores.

Tabela 47. Ferramentas de NLP para Análise Eleitoral

Ferramenta	Tarefa Principal	Língua	Nível Técnico Necessário	Acesso	Aplicação Eleitoral
BERTimbau (USP)	Análise de sentimento, classificação de texto, NER em português.	Português brasileiro	Intermediário (Python)	Hugging Face (gratuito)	Análise de sentimento de comentários sobre candidatos; classificação de conteúdo eleitoral.
VADER	Análise de sentimento lexical em inglês.	Inglês	Básico (Python)	pip install vaderSentiment	Análise de conteúdo em inglês; referência metodológica para comparação.
spaCy	Pipeline completo de NLP: tokenização, POS tagging, NER, dependências.	Múltiplas línguas (pt disponível)	Intermediário (Python)	pip install spacy	Extração de entidades (candidatos, partidos, temas) de corpus de texto eleitoral.
Hugging Face Transformers	Acesso a milhares de modelos pré-treinados para diversas tarefas de NLP.	Múltiplas línguas	Intermediário a avançado	pip install transformers	Fine-tuning de modelos para detecção de desinformação eleitoral específica do contexto brasileiro.
Google Natural Language API	Análise de entidades, sentimento, classificação de conteúdo via API.	Múltiplas línguas	Básico (chamadas de API)	Google Cloud (pago por uso)	Análise em escala de grande volume de conteúdo eleitoral sem conhecimento técnico de ML.

Fonte: Elaborado pelo autor com base em Souza, Nogueira e Lotufo (2020) e documentação das ferramentas (2024).

Análise de Dados Eleitorais do TSE com Python

O Portal de Dados Abertos do TSE disponibiliza um dos acervos de dados eleitorais mais ricos e acessíveis do mundo, e essa riqueza é sistematicamente subutilizada tanto por

pesquisadores quanto por profissionais de campanha e advogados eleitorais que poderiam extrair dela evidências analiticamente relevantes para seus trabalhos. O acervo inclui resultados de votação por seção eleitoral desde 1994, candidaturas com dados biográficos e partidários, prestações de contas com receitas e despesas declaradas por candidato e partido, bens declarados por candidatos, dados de filiação partidária e perfil do eleitorado por município. A combinação dessas bases permite análises que nenhuma fonte de dados privada consegue replicar, porque elas capturam o processo eleitoral em sua totalidade, desde o financiamento até o resultado final por seção eleitoral.

A relevância desse acervo para a proteção da integridade eleitoral na era da IA é mais direta do que aparece à primeira vista. Operações de desinformação sofisticadas deixam rastros nos dados eleitorais que análise quantitativa pode identificar: variações anômalas no padrão de votação em municípios específicos que não se explicam por fatores socioeconômicos conhecidos, inconsistências entre despesas declaradas de campanha e o volume estimado de conteúdo digital produzido, padrões de financiamento que sugerem coordenação entre doadores aparentemente independentes, e discrepâncias entre o perfil declarado de candidatos e o perfil de seus doadores que podem indicar operações de caixa dois. Como argumentam Benkler, Faris e Roberts (2018), a análise de dados estruturados é um dos instrumentos mais eficazes para identificar operações coordenadas que, por design, buscam evitar a detecção por monitoramento de conteúdo (BENKLER; FARIS; ROBERTS, 2018, p. 356-361).

O acesso aos dados do TSE em formato CSV permite análises quantitativas sofisticadas com ferramentas de código aberto disponíveis gratuitamente. Python, com suas bibliotecas de análise de dados, é o ambiente mais adequado para esse trabalho por três razões: a biblioteca pandas permite manipulação eficiente de tabelas com milhões de linhas, que é a escala típica dos arquivos de dados eleitorais do TSE; a biblioteca matplotlib e suas extensões permitem visualizações que tornam padrões analiticamente relevantes visualmente identificáveis; e a biblioteca networkx permite análises de redes que identificam estruturas de coordenação entre doadores, candidatos e partidos que não seriam detectáveis por análise tabular convencional.

Quatro análises específicas têm particular relevância para a proteção da integridade eleitoral e para a prática do advogado eleitoral que busca evidências em investigações de abuso de poder econômico ou financiamento irregular.

A primeira é a análise de padrões anômalos de financiamento, que examina a distribuição de doações por doador, candidato e período, buscando desvios

estatisticamente significativos em relação aos padrões históricos e aos padrões de candidatos comparáveis. Doações em valores exatamente abaixo dos limites que disparam obrigações de transparência adicional, doadores que contribuem para múltiplos candidatos de partidos distintos sem relação aparente, e picos de arrecadação concentrados em janelas temporais que coincidem com eventos eleitorais específicos são padrões que análise quantitativa identifica com precisão e que a revisão manual de prestações de contas frequentemente não detecta pela escala do corpus.

A segunda é a análise de inconsistências entre despesas declaradas e volume de atividade digital estimada. Com a democratização das ferramentas de IA generativa, o custo marginal de produção de conteúdo digital em escala caiu drasticamente, como analisado no Capítulo 12 na discussão sobre abuso de poder econômico. Campanhas que declaram despesas mínimas com produção de conteúdo mas apresentam volume de atividade digital em plataformas incompatível com esse nível de investimento estão potencialmente operando com recursos não declarados. A análise quantitativa que combina dados de prestação de contas com dados de atividade em plataformas, obtidos por social listening ou por análise de dados fornecidos pelas plataformas em investigações formais, pode produzir evidências indiciárias relevantes para processos de abuso de poder econômico.

A terceira é a análise de variações suspeitas no padrão de votação, que examina a distribuição de votos por seção eleitoral buscando padrões estatisticamente atípicos que não se explicam por variáveis socioeconômicas, geográficas ou históricas conhecidas. Essa análise é particularmente relevante em contextos de alegações de irregularidades eleitorais, porque permite distinguir variações que têm explicação em características observáveis do eleitorado de variações que permanecem inexplicadas após controle estatístico por essas variáveis, o que é o primeiro passo analítico para investigação mais aprofundada de possíveis irregularidades.

A quarta é a análise de redes de financiamento, que constrói grafos de relações entre doadores e candidatos para identificar estruturas de coordenação que não seriam visíveis na análise individual de cada doação. Quando múltiplos doadores aparentemente independentes contribuem para os mesmos candidatos em padrões temporais e de valor coordenados, a análise de redes pode revelar a estrutura subjacente que conecta essas contribuições, produzindo evidências indiciárias relevantes para investigações de financiamento coordenado não declarado.

A barreira de entrada para essas análises é menor do que a maioria dos profissionais jurídicos antecipa. O Portal de Dados Abertos do TSE disponibiliza os dados em formato

CSV com dicionários de variáveis detalhados. Python com pandas é amplamente documentado e tutoriais específicos para análise de dados eleitorais brasileiros estão disponíveis em repositórios públicos no GitHub. O investimento de aprendizado necessário para realizar análises básicas de prestação de contas e padrões de votação é acessível a qualquer profissional com disposição para dedicar algumas semanas a cursos introdutórios de análise de dados. Para análises mais sofisticadas, a colaboração com pesquisadores de ciência de dados de universidades brasileiras, que frequentemente trabalham com dados eleitorais do TSE e têm interesse em parcerias com profissionais jurídicos que possam traduzir achados quantitativos em argumentos jurídicos, é um modelo de trabalho que tem produzido resultados relevantes nos ciclos eleitorais anteriores.

Principais Bases de Dados do TSE para Análise de Campanha

Candidaturas: nome, partido, cargo, situação do registro, dados biográficos. Disponível em: dados.tse.jus.br/dataset/candidatos. Prestação de contas: receitas e despesas por candidato, por doador, por fornecedor. Disponível em: dados.tse.jus.br/dataset/contas-campanha. Resultados eleitorais: votos por candidato, por zona, por seção, por município. Disponível em: dados.tse.jus.br/dataset/resultados. Impulsionamento: registro de anúncios patrocinados em plataformas digitais. Disponível em: dados.tse.jus.br/dataset/transparencia-nas-eleicoes. Filiação partidária: histórico de filiações por CPF. Disponível em: dados.tse.jus.br/dataset/filiados.

Brasil. TSE. Portal de Dados Abertos: dados.tse.jus.br

Protocolo de Resposta a Incidentes de Deepfake Eleitoral

A velocidade com que um deepfake eleitoral pode se propagar em ambiente de comunicação digital fragmentado e acelerado torna o protocolo de resposta pré-estabelecido não uma boa prática de gestão de crise, mas uma condição necessária para que qualquer resposta seja eficaz. Em período pré-eleitoral, um deepfake disseminado às 22h de uma sexta-feira pode ter alcançado centenas de milhares de eleitores antes que qualquer membro da equipe jurídica da campanha seja notificado na manhã de sábado. A improvisação nesse momento, quando a pressão temporal é máxima, as consequências são imediatas e os tomadores de decisão estão sob estresse, é o cenário mais desfavorável possível para uma resposta eficaz. Protocolos pré-estabelecidos não eliminam a pressão: eles eliminam a necessidade de tomar decisões sobre procedimentos sob pressão,

liberando a capacidade de julgamento da equipe para as decisões que realmente dependem de avaliação situacional.

A estrutura de um protocolo eficaz de resposta a incidentes de deepfake eleitoral organiza-se em cinco fases sequenciais que precisam ser definidas, testadas e internalizadas pela equipe antes do início do período eleitoral.

A primeira fase é a detecção e triagem, que deve ocorrer nos primeiros trinta minutos após o conhecimento do incidente. Nessa fase, a equipe de monitoramento identifica o conteúdo suspeito, documenta sua localização em todas as plataformas onde está circulando por meio de capturas de tela com timestamp, e realiza triagem inicial usando ferramentas de detecção de deepfakes disponíveis, como as descritas na seção anterior deste capítulo. O objetivo da triagem não é produzir conclusão definitiva sobre a natureza sintética do conteúdo, que exige análise forense, mas determinar se o conteúdo tem características suficientes para acionar o protocolo completo. Conteúdo que passa pela triagem inicial é imediatamente escalado para a segunda fase sem aguardar confirmação definitiva: o custo de acionar o protocolo desnecessariamente é incomparavelmente menor do que o custo de atrasar a resposta a um deepfake real.

A segunda fase é a preservação de evidências, que deve ocorrer em paralelo com a triagem e nunca depois dela. Toda evidência sobre o deepfake precisa ser preservada com metodologia que garanta sua admissibilidade como prova em processo eleitoral: capturas de tela autenticadas com serviço de timestamping confiável, arquivos de vídeo ou áudio baixados antes de qualquer remoção, URLs documentadas com data e hora de acesso, e registros de métricas de engajamento disponíveis publicamente no momento da descoberta. A preservação de evidências é o passo que mais frequentemente é negligenciado na urgência da resposta, e é o que mais frequentemente compromete a eficácia do processo judicial posterior: evidências não preservadas no momento do incidente frequentemente não podem ser recuperadas depois que o conteúdo é removido ou modificado.

A terceira fase é a análise forense, que deve ser iniciada dentro da primeira hora e conduzida por perito especializado previamente identificado e contratado. A análise forense produz o laudo com validade jurídica que distingue a suspeita técnica, produzida pelas ferramentas de triagem, da prova pericial admissível em processo eleitoral. O perito forense precisa estar disponível para acionamento imediato durante o período eleitoral, o que significa que o contrato com o perito precisa ser estabelecido antes do início da campanha, com cláusulas específicas sobre prazo de resposta em situações de urgência eleitoral.

Campanhas que chegam ao período eleitoral sem perito forense contratado perdem horas críticas identificando e negociando com um profissional no momento em que cada hora de atraso permite que o deepfake alcance mais eleitores.

A quarta fase é o acionamento jurídico, que deve ocorrer simultaneamente à análise forense sem aguardar seu resultado definitivo. Essa fase envolve três ações paralelas: a notificação formal à plataforma pelo canal estabelecido nos acordos de cooperação com o TSE, acionando o prazo de vinte e quatro horas para remoção e a responsabilidade solidária por descumprimento; o protocolo de representação administrativa perante o TRE competente com pedido de liminar de remoção, que pode ser baseado na suspeita técnica fundamentada mesmo antes do laudo forense definitivo; e o registro da queixa-crime por calúnia eleitoral, que preserva a opção de responsabilização penal independentemente do desfecho do processo administrativo. As três ações são independentes entre si e podem ser realizadas em paralelo por diferentes membros da equipe jurídica, o que exige que os modelos de peças estejam preparados com antecedência e que as responsabilidades de cada membro da equipe estejam claramente definidas no protocolo.

A quinta fase é a resposta comunicacional, que precisa ser coordenada com as fases jurídicas mas não pode aguardar seu desfecho. Eleitores expostos ao deepfake precisam receber a correção antes que a narrativa falsa se consolide como enquadramento de referência, o que como demonstrado por Pennycook e Rand (2019) ocorre com velocidade que processos jurídicos não conseguem acompanhar. A resposta comunicacional inclui a declaração pública do candidato desmentindo o conteúdo com clareza e sem amplificar desnecessariamente o alcance do deepfake, a mobilização de canais de distribuição informal pelos mesmos vetores que o deepfake utilizou para circular, e a coordenação com fact-checkers parceiros que possam publicar verificação com credibilidade editorial independente da campanha. O conteúdo da resposta comunicacional precisa ser preparado em template com antecedência, adaptável ao conteúdo específico do incidente, porque a redação de declarações sob pressão temporal máxima produz consistentemente resultados inferiores aos produzidos com tempo adequado de elaboração.

O protocolo completo precisa ser testado antes do período eleitoral por meio de simulação de incidente que reproduza as condições de pressão temporal do cenário real. Equipes que leram o protocolo mas nunca o executaram sob simulação de pressão frequentemente descobrem na prática que os procedimentos definidos têm lacunas, que as responsabilidades não estão suficientemente claras ou que os contatos de acionamento estão desatualizados. A simulação tem custo marginal baixo e valor de teste alto: ela é o

único instrumento que permite identificar e corrigir essas falhas antes que se manifestem no momento de menor capacidade de correção.

O quadro a seguir traduz as cinco fases do protocolo em formato operacional, especificando para cada uma o tempo disponível, as ações concretas, o responsável e as ferramentas a utilizar. A leitura por coluna de tempo revela a característica mais crítica do protocolo: as fases se sobrepõem deliberadamente. A notificação jurídica começa na segunda hora, quando a análise técnica preliminar ainda está em andamento; a comunicação pública começa na terceira hora, quando o processo jurídico foi apenas iniciado. Essa sobreposição não é imprecisão no design do protocolo: é sua característica mais importante, porque cada hora de atraso em qualquer frente representa exposição adicional de eleitores ao conteúdo falso que nenhuma resposta posterior consegue reverter completamente. A coluna de responsável é o elemento que mais frequentemente revela lacunas quando o protocolo é revisado criticamente: se a campanha não tem especialista técnico identificado para a segunda fase, se o advogado eleitoral não tem modelo de representação preparado para a terceira, ou se a equipe de comunicação não tem template de resposta para a quarta, o protocolo existe no papel mas não é executável na prática. A simulação de incidente recomendada na seção anterior serve precisamente para testar a coluna de responsável contra a realidade da equipe disponível, identificando antes do pleito quais fases têm lacunas que precisam ser preenchidas por contratação, treinamento ou redistribuição de responsabilidades.

Tabela 48. Protocolo de Resposta a Incidente de Deepfake Eleitoral

Fase	Tempo	Ações	Responsável	Ferramentas
Detecção	0-1h	Identificar e documentar o conteúdo suspeito (URL, screenshot, download do arquivo com metadados preservados). Verificar canal de origem e volume de disseminação.	Equipe de monitoramento digital da campanha.	InVID/WeVerify; screenshot tools; ferramentas de download com preservação de metadados.

Análise técnica preliminar	1-3h	Aplicar ferramentas de detecção disponíveis (Deepware, Microsoft Video Authenticator). Identificar artefatos visuais e inconsistências. Documentar resultado com evidências.	Especialista técnico da campanha ou assessor contratado.	Deepware Scanner; Microsoft Video Authenticator; FotoForensics.
Notificação jurídica	2-4h	Protocolar representação no TSE/TRE com pedido de remoção liminar e direito de resposta. Anexar evidências técnicas e documentação da disseminação.	Advogado eleitoral da campanha.	Portal e-CAC Eleitoral; jurisprudencia.tse.jus.br (precedentes).
Comunicação pública	3-6h	Publicar resposta clara e direta nas redes sociais do candidato, identificando o conteúdo como falso e apresentando o material autêntico correspondente. Notificar veículos de imprensa.	Comunicação da campanha.	Redes sociais oficiais; releases para imprensa; WhatsApp dos grupos de apoiadores.
Acompanhamento	24-72h	Monitorar a eficácia da remoção pelas plataformas. Acompanhar o processo no TSE/TRE. Documentar danos para eventual pedido de indenização.	Equipe jurídica e de monitoramento.	Ferramentas de social listening; portal do TSE.

Fonte: Elaborado pelo autor com base em Paris e Donovan (2019) e jurisprudência do TSE (2024).

■ **RESUMO DO CAPÍTULO:**

Este capítulo analisou as categorias funcionais de ferramentas digitais disponíveis para proteção da integridade eleitoral, organizadas pelo princípio de que a pergunta mais importante não é qual ferramenta usar, mas como interpretar seus resultados e quais são seus limites de confiabilidade em cada caso específico. Esse princípio orientou toda a análise: ferramentas são instrumentos de triagem e orientação da investigação, nunca substitutos do julgamento humano especializado. A análise das ferramentas de detecção de deepfakes estabeleceu quatro categorias técnicas com capacidades e limitações distintas: análise de artefatos visuais, eficaz para deepfakes de primeira e segunda geração mas com eficácia declinante diante de modelos de difusão avançados; análise biométrica por rPPG, robusta para modelos anteriores mas degradada pela compressão de arquivo que ocorre na circulação por WhatsApp; análise de metadados, útil mas vulnerável à manipulação e à perda de dados na cadeia de distribuição; e análise por modelos de aprendizado de máquina, a mais poderosa disponível e a mais vulnerável ao counter-adversarial training. O princípio da corroboração múltipla, que exige convergência de múltiplas categorias de análise antes de conclusões definitivas, foi identificado como o protocolo que melhor equilibra confiabilidade e praticidade. A distinção entre ferramenta de triagem e prova pericial com cadeia de custódia foi estabelecida como a mais operacionalmente relevante para o advogado eleitoral. O social listening eleitoral foi analisado em sua dupla finalidade, estratégica para campanhas e defensiva para a Justiça Eleitoral e pesquisadores, com a implicação jurídica de que a mesma ferramenta tem regime jurídico diferente conforme o uso. A limitação estrutural do WhatsApp como canal inacessível ao social listening convencional foi identificada como a contradição mais relevante do campo no contexto brasileiro: a desinformação mais impactante circula precisamente pelo canal que os sistemas de monitoramento menos conseguem alcançar. Três capacidades específicas foram desenvolvidas: detecção de coordenação inautêntica por análise de padrões temporais e estruturais, monitoramento de narrativas emergentes em estágio inicial de disseminação e mapeamento de influenciadores e estrutura de redes. As ferramentas de NLP foram analisadas em duas aplicações eleitoralmente relevantes, análise de sentimento e detecção automatizada de desinformação, com ênfase nas limitações de cada uma para o contexto específico do discurso político brasileiro. O BERTimbau foi identificado como o modelo de estado da arte para análise de texto em português brasileiro, com vantagem específica sobre modelos multilíngues pela captura de especificidades semânticas e pragmáticas do português brasileiro. Três pontos de entrada foram apresentados para diferentes perfis de capacidade técnica: implementação própria com Python, consultoria especializada e APIs comerciais, cada um com trade-offs de custo, flexibilidade e privacidade. A análise de dados eleitorais do TSE com Python foi apresentada como o instrumento mais subutilizado por profissionais jurídicos eleitorais. Quatro análises específicas foram desenvolvidas com sua lógica de detecção: padrões anômalos de financiamento, inconsistências entre despesas declaradas e volume de atividade digital, variações suspeitas no padrão de votação e análise de redes de financiamento. Para profissionais sem formação técnica, Tableau Public e Google Looker Studio foram identificados como pontos de entrada acessíveis. O modelo de parceria entre advogado eleitoral e analista de dados foi apresentado como o caminho mais eficaz para transformar dados quantitativos em argumentos processuais. O protocolo de resposta a incidentes de deepfake eleitoral organizou a resposta em cinco fases com prazos e responsáveis definidos: detecção e triagem nos primeiros trinta minutos, preservação de evidências em paralelo, análise forense dentro da primeira hora, acionamento jurídico simultâneo em três frentes paralelas e resposta comunicacional coordenada. A sobreposição deliberada das fases foi identificada como característica central do design, não como imprecisão. A simulação pré-eleitoral foi apresentada como o único instrumento capaz de testar a executabilidade real do protocolo antes que falhas se manifestem no momento de menor capacidade de correção

QUESTÕES PARA REFLEXÃO E DEBATE

1. O capítulo estabeleceu que nenhuma ferramenta de detecção de deepfakes tem cobertura universal e que a corrida armamentista entre modelos de geração e detecção tende a favorecer estruturalmente os sistemas de geração. Se essa assimetria é irredutível no médio prazo, qual deveria ser a estratégia regulatória prioritária: investir em melhoria das ferramentas de detecção, investir em sistemas de proveniência criptográfica como o padrão C2PA que tornam a autenticação independente da detecção, ou investir em literacia midiática que reduza a vulnerabilidade dos eleitores ao conteúdo sintético?

2. A limitação estrutural do WhatsApp ao social listening eleitoral cria uma assimetria: a desinformação que circula no canal de maior impacto eleitoral é precisamente a que o sistema de monitoramento menos consegue alcançar. Essa assimetria é inerente ao design de privacidade do aplicativo e não pode ser superada por ferramentas mais sofisticadas sem comprometer a criptografia de ponta a ponta. Como o direito eleitoral deveria endereçar essa assimetria sem comprometer o direito à privacidade das comunicações, que é um direito fundamental constitucionalmente protegido?

3. A análise de dados eleitorais do TSE com Python pode identificar padrões anômalos de financiamento e inconsistências entre despesas declaradas e volume de atividade digital que são indícios relevantes de abuso de poder econômico. Qual é o valor probatório que a Justiça Eleitoral deveria atribuir a análises quantitativas produzidas por peritos particulares contratados pelas partes, e quais padrões metodológicos deveriam ser exigidos para que essas análises tenham validade como prova em processos de investigação judicial eleitoral?

4. O protocolo de resposta a incidentes de deepfake eleitoral pressupõe que a campanha tem perito forense pré-contratado, advogado eleitoral com modelo de representação preparado e equipe de comunicação com template de resposta disponível. Essa infraestrutura de resposta é acessível para campanhas presidenciais e estaduais bem financiadas, mas está fora do alcance da maioria das campanhas municipais de menor porte. Como o sistema eleitoral deveria democratizar o acesso a essa infraestrutura de proteção sem criar assimetrias adicionais entre candidatos com e sem recursos?

5. A análise de sentimento aplicada ao discurso político tem limitações significativas para o contexto brasileiro por sua dificuldade em capturar ironia, sarcasmo e humor político. Essas limitações técnicas têm implicação jurídica direta: um sistema de triagem de desinformação baseado em análise de sentimento pode gerar falsos positivos que levam à remoção de conteúdo politicamente crítico mas legítimo. Como o direito eleitoral deveria regular o uso de sistemas automatizados de moderação de conteúdo por plataformas, garantindo que falsos positivos não se tornem instrumentos de censura política?

6. O capítulo identificou que a análise de redes de financiamento pode revelar estruturas de coordenação entre doadores aparentemente independentes que não seriam detectáveis por análise individual de cada doação. Essa capacidade analítica é valiosa para investigações de abuso de poder econômico, mas também pode ser utilizada por campanhas para identificar e pressionar doadores de adversários. Como o acesso público aos dados de prestação de contas do TSE deveria ser estruturado para maximizar a transparência e a fiscalização cidadã sem criar instrumentos de intimidação de financiadores de candidatos?

7. O protocolo de resposta a incidentes estabelece que a comunicação pública desmentindo o deepfake deve ocorrer na terceira a sexta hora, sem amplificar desnecessariamente o conteúdo falso. Essa orientação entra em tensão com a pesquisa de Pennycook e Rand, que sugere que correções precisam alcançar o mesmo público exposto ao conteúdo original para serem eficazes, o que frequentemente exige amplificação ativa. Como campanhas deveriam navegar essa tensão entre o risco de amplificar o deepfake e o risco de deixar eleitores sem a correção necessária?

8. Este capítulo analisou ferramentas para proteção da integridade eleitoral disponíveis em 2026. Para as eleições de 2030, quais categorias de ferramentas, hoje inexistentes ou em estágio inicial de

desenvolvimento, seriam mais necessárias para enfrentar os riscos de IA eleitoral que os sistemas atuais não conseguem endereçar adequadamente?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Resultado de ferramenta de detecção não é prova. É triagem. Nunca confunda os dois.

Nenhuma ferramenta de detecção de deepfakes disponível atualmente produz resultado com validade jurídica como prova em processo eleitoral. O Deepware Scanner, o Microsoft Video Authenticator e o FotoForensics são instrumentos de triagem: eles orientam a decisão sobre se um conteúdo merece investigação aprofundada, mas seus resultados não substituem o laudo pericial com cadeia de custódia produzido por especialista forense. A confusão entre triagem e prova é o erro mais comum no uso dessas ferramentas em contexto eleitoral e o que mais frequentemente compromete processos judiciais posteriores. Use as ferramentas de triagem para decidir se aciona o perito; use o perito para produzir a prova.

2. Contrate o perito forense antes do período eleitoral, não no dia do incidente.

O laudo pericial é o único resultado com validade jurídica sobre a natureza sintética de um conteúdo eleitoral. Produzir esse laudo exige um perito especializado contratado e disponível para acionamento imediato. Uma campanha que chega ao período eleitoral sem perito forense identificado e pré-contratado perderá horas críticas identificando e negociando com um profissional no momento em que cada hora de atraso permite que o deepfake alcance mais eleitores. O contrato com o perito precisa incluir cláusula de prazo de resposta em situações de urgência eleitoral, com penalidade pelo descumprimento. Esse é o investimento de preparação com maior retorno esperado no protocolo de resposta a deepfakes.

3. Preserve as evidências antes de acionar qualquer outra resposta. O conteúdo removido não volta.

A preservação de evidências é o passo mais frequentemente negligenciado na urgência da resposta a deepfakes e o que mais frequentemente compromete a eficácia do processo judicial posterior. Quando um deepfake é identificado, a primeira ação é documentar: captura de tela autenticada com timestamp, download do arquivo com metadados preservados, registro das URLs em todas as plataformas onde o conteúdo está circulando e documentação das métricas de engajamento disponíveis publicamente naquele momento. Essas evidências precisam ser coletadas antes de qualquer notificação à plataforma para remoção, porque a remoção elimina o acesso ao conteúdo original. Evidências não preservadas no momento do incidente frequentemente não podem ser recuperadas depois.

4. Social listening no WhatsApp não existe. Sua estratégia de monitoramento precisa reconhecer isso.

O WhatsApp é o canal de desinformação mais relevante no contexto eleitoral brasileiro e é estruturalmente inacessível ao social listening convencional. Sua criptografia de ponta a ponta e a natureza privada das comunicações impedem qualquer monitoramento sistemático por ferramentas externas. Isso não é limitação técnica superável por ferramentas mais sofisticadas: é constitutiva do design de privacidade do aplicativo. Estratégias de monitoramento que assumem cobertura do WhatsApp estão operando com uma lacuna crítica. A resposta prática é combinar monitoramento de plataformas abertas com redes de informantes voluntários em grupos de WhatsApp relevantes, sistemas de denúncia como o tipline da Agência Lupa, e monitoramento do conteúdo que migra do WhatsApp para plataformas públicas após circulação privada.

5. Os dados eleitorais do TSE são públicos e gratuitos. Usá-los pode fazer diferença num processo de abuso de poder.

O Portal de Dados Abertos do TSE disponibiliza resultados de votação por seção eleitoral, prestações de contas completas e dados de filiação partidária em formato CSV, gratuitamente, sem necessidade de qualquer credencial. Para o advogado eleitoral que investiga abuso de poder econômico, esses dados permitem análises que nenhuma fonte privada consegue replicar: padrões anômalos de financiamento, inconsistências entre despesas declaradas e volume de atividade digital e variações suspeitas no padrão de votação por seção eleitoral. Se você não tem capacidade técnica para analisar esses dados, a parceria com pesquisador universitário de ciência de dados é o caminho mais rápido. Se tem disposição para aprender, Tableau Public e Google Looker Studio permitem visualizações analíticas sem nenhuma linha de código.

6. O protocolo de resposta só funciona se existir antes do incidente. Simule agora.

Um protocolo de resposta a deepfakes que existe apenas no papel não é um protocolo: é uma intenção. A única forma de saber se o protocolo é executável na prática é testá-lo sob simulação de pressão antes do período eleitoral. A simulação revela lacunas que a leitura do documento não revela: responsabilidades que não estão claras, contatos de acionamento desatualizados, fases que se sobrepõem de forma inviável com a equipe disponível, e etapas que assumem recursos ou capacidades que a campanha não tem. Corrigir essas lacunas em simulação tem custo zero. Descobri-las no momento do incidente real tem custo potencialmente decisivo para o resultado eleitoral.

7. Notificação à plataforma e resposta comunicacional são ações paralelas, não sequenciais.

Um erro frequente no gerenciamento de crises de deepfake eleitoral é tratar as respostas como sequenciais: primeiro notificar a plataforma, depois acionar a Justiça Eleitoral, depois comunicar publicamente. O protocolo eficaz trata essas três frentes como paralelas desde a segunda hora. A notificação à plataforma ativa o prazo de vinte e quatro horas e a responsabilidade solidária. O protocolo jurídico ao TRE busca a liminar de remoção independentemente da cooperação da plataforma. A resposta comunicacional alcança os eleitores que já foram expostos, que não esperam pelo resultado do processo judicial. Cada hora de atraso em qualquer das três frentes é exposição adicional de eleitores ao conteúdo falso que nenhuma resposta posterior consegue reverter completamente.

CAPÍTULO 15

O que Candidatos, Assessores e Partidos Precisam Saber

Guia Operacional para Conformidade, Estratégia e Gestão de Riscos na Era da IA Eleitoral

Da Teoria à Prática: O que este Capítulo Oferece

Os quatorze capítulos anteriores desta obra construíram um arcabouço teórico, jurídico e técnico abrangente sobre o uso da inteligência artificial no processo eleitoral brasileiro. Este capítulo tem propósito distinto e complementar: traduzir esse arcabouço em protocolos operacionais concretos para os atores que atuam diretamente nas campanhas eleitorais, candidatos, assessores, consultores de marketing político, advogados eleitorais e dirigentes partidários. A distância entre compreender o campo e saber operar nele é real e relevante: um profissional que leu os capítulos anteriores conhece a Resolução TSE nº 23.732/2024, compreende o regime de responsabilização, entende as limitações das ferramentas de detecção e conhece os precedentes jurisprudenciais do TSE, mas ainda pode chegar ao período eleitoral sem saber exatamente o que fazer na segunda-feira de manhã para garantir que sua campanha está em conformidade.

A estrutura deste capítulo segue a lógica de um manual de boas práticas orientado por perfil de ator e por momento do ciclo eleitoral. Para cada perfil, os riscos específicos são identificados antes dos protocolos, porque protocolos sem diagnóstico preciso de risco tendem a ser genéricos demais para orientar decisões concretas. Para cada momento do ciclo eleitoral, a sequência de ações é apresentada na ordem em que precisam ser tomadas, porque a ordem importa: medidas preventivas que chegam depois do incidente não são preventivas. O objetivo não é substituir a consultoria jurídica especializada, que é indispensável em contextos de maior complexidade, mas fornecer o mapa de navegação que permite ao profissional identificar quando precisa de ajuda especializada e chegar preparado a essa consulta com as perguntas certas já formuladas.

Uma observação metodológica sobre o uso dos instrumentos deste capítulo é necessária antes de sua apresentação. Checklists, protocolos e matrizes de risco são instrumentos de redução de complexidade: eles simplificam situações multidimensionais em categorias manejáveis para permitir ação em tempo útil. Essa simplificação é sua vantagem operacional e sua limitação analítica. Casos que não se encaixam nas categorias previstas pelo instrumento, que são frequentemente os casos mais importantes, exigem

julgamento que nenhum checklist consegue fornecer. O profissional que usa os instrumentos deste capítulo como substituto do julgamento está usando-os incorretamente; o que os usa como estrutura que organiza o julgamento está usando-os da forma pretendida.

Checklist de Conformidade: IA e a Resolução TSE nº 23.732/2024

O primeiro instrumento operacional que todo profissional de campanha precisa ter é uma checklist de conformidade com a Resolução TSE nº 23.732/2024. A ausência de conformidade com a Resolução não é apenas risco jurídico: é risco reputacional, porque violações identificadas por adversários ou pela imprensa durante o período eleitoral podem se tornar narrativas de campanha mais prejudiciais do que as sanções formais aplicadas depois do pleito.

A checklist funciona como dois instrumentos simultaneamente: como instrumento preventivo, verificando antes do início do período eleitoral se a campanha está estruturalmente conforme; e como instrumento de auditoria corrente, verificando periodicamente durante a campanha se as práticas operacionais estão mantendo a conformidade que a estrutura inicial estabeleceu.

A checklist a seguir foi elaborada a partir da análise das principais obrigações da Resolução e das condutas que têm gerado as primeiras representações e decisões do TSE nas eleições municipais de 2024, o primeiro ciclo de aplicação da norma. Ela não é exaustiva: casos específicos podem exigir verificações adicionais que dependem das características particulares de cada campanha. Ela é, contudo, suficiente para identificar as violações mais frequentes e as lacunas de governança mais comuns que transformam risco potencial em responsabilização concreta.

☑ CHECKLIST DE CONFORMIDADE

Resolução TSE nº 23.732/2024 · IA Eleitoral · Eleições 2026

Nº	OBRIGAÇÃO	COMO VERIFICAR	STATUS
BLOCO 1 — CONTEÚDO GERADO POR IA			

1	Todo conteúdo gerado com auxílio de IA está rotulado de forma clara e proeminente.	*Revisar cada peça eleitoral e verificar se a identificação está presente, legível e em posição de destaque.*	☐ Conforme ☐ Pendente ☐ N/A
2	Nenhum conteúdo simula declarações ou posições de candidatos adversários por deepfake ou síntese de voz.	*Verificar com a equipe criativa se alguma ferramenta de síntese de voz ou face swap foi usada com imagens de terceiros.*	☐ Conforme ☐ Pendente ☐ N/A
3	Os chatbots da campanha se identificam imediatamente como sistemas automatizados em toda interação.	*Testar o chatbot como usuário externo e verificar se a identificação é a primeira mensagem enviada.*	☐ Conforme ☐ Pendente ☐ N/A
BLOCO 2 — PROTEÇÃO DE DADOS E LGPD			
4	Os dados pessoais de eleitores utilizados para targeting têm base legal adequada documentada.	*Verificar com o DPO ou advogado de dados quais bases legais sustentam cada tipo de tratamento realizado.*	☐ Conforme ☐ Pendente ☐ N/A
5	Dados sensíveis (opinião política inferida por perfil psicográfico) têm consentimento específico e destacado.	*Solicitar à empresa de marketing digital a documentação das bases legais para tratamento de dados psicográficos.*	☐ Conforme ☐ Pendente ☐ N/A
6	O registro de operações de tratamento de dados está sendo mantido conforme art. 37 da LGPD.	*Verificar existência e atualização do registro com o responsável de compliance da campanha.*	☐ Conforme ☐ Pendente ☐ N/A
BLOCO 3 — FINANCIAMENTO E PRESTAÇÃO DE CONTAS			
7	A contratação de ferramentas e empresas de IA está declarada na prestação de contas eleitorais.	*Verificar com o tesoureiro se todos os contratos com empresas de tecnologia e marketing digital estão registrados.*	☐ Conforme ☐ Pendente ☐ N/A
8	O impulsionamento pago de conteúdo em plataformas digitais está registrado no sistema do TSE.	*Acessar o portal de transparência de impulsionamento do TSE e verificar o registro das campanhas pagas.*	☐ Conforme ☐ Pendente ☐ N/A
9	Não há contratos com agências de disparo em massa pelo WhatsApp sem declaração como despesa de campanha.	*Verificar com o gestor financeiro se há contratos com empresas de disparos não declarados.*	☐ Conforme ☐ Pendente ☐ N/A
BLOCO 4 — PROTOCOLO DE RESPOSTA E GOVERNANÇA			
10	A campanha tem protocolo de resposta a incidentes de deepfake eleitoral estabelecido e testado.	*Verificar existência do protocolo e data do último teste de simulação com a equipe responsável.*	☐ Conforme ☐ Pendente ☐ N/A
11	Perito forense especializado em análise de deepfakes está pré-contratado para o período eleitoral.	*Verificar contrato e cláusula de prazo de resposta em situação de urgência eleitoral.*	☐ Conforme ☐ Pendente ☐ N/A
12	Conteúdo sintético não será publicado nos 3 dias anteriores a cada turno (1º/out e 22/out de 2026).	*Verificar calendário de publicação programada e desativar*	☐ Conforme ☐ Pendente

		automações para os períodos vedados.	☐ **N/A**

LEGENDA: ☐ **Conforme** = item verificado e em ordem ☐ **Pendente** = requer ação imediata ☐ **N/A** = não aplicável a esta campanha

Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024 e jurisprudência do TSE (2024).

Guia Estratégico: Usos Lícitos e Eficazes de IA em Campanhas

A conformidade regulatória não é obstáculo ao uso estratégico e eficaz da inteligência artificial em campanhas eleitorais: é sua condição de sustentabilidade. Campanhas que utilizam IA de forma transparente e juridicamente fundamentada não estão apenas evitando sanções; estão construindo uma vantagem competitiva real sobre adversários que operam em violação e que carregam o risco permanente de ter suas operações interrompidas por liminares de remoção, de ter suas estratégias expostas em processos eleitorais e de ter seus resultados contestados por abuso de poder. O risco jurídico não é apenas custo de conformidade: é desvantagem estratégica que pode decidir eleições.

Os usos mais eficazes de IA em campanhas são frequentemente os que apresentam menor risco jurídico, e essa correlação não é coincidência: ela reflete o fato de que as tecnologias mais poderosas disponíveis hoje operam com e não contra a autonomia do eleitor. A produção de conteúdo em escala com qualidade consistente, a personalização de mensagens por perfil de público, a análise de dados eleitorais para identificação de oportunidades e a automação de atendimento com identificação transparente são todas capacidades que amplificam a efetividade da comunicação política legítima sem exigir o engano que gera responsabilização.

A distinção que orienta o mapeamento a seguir é a mesma que estrutura toda a regulação eleitoral de IA analisada nos capítulos anteriores: a diferença entre persuasão que apela à razão e à emoção de forma transparente, que é o núcleo da comunicação política legítima em qualquer democracia, e manipulação que explora vulnerabilidades para produzir comportamentos que o eleitor não escolheria em condições de plena informação e autonomia. Ferramentas que operam no primeiro campo são não apenas lícitas, mas estrategicamente superiores a longo prazo, porque constroem a relação de confiança com o eleitorado que é o ativo político mais durável disponível. Ferramentas que operam no segundo campo podem produzir ganhos táticos no ciclo eleitoral corrente ao custo de

responsabilização jurídica, exposição reputacional e erosão da legitimidade que comprometem trajetórias políticas inteiras.

O quadro a seguir organiza os usos lícitos de IA em seis categorias funcionais, cada uma com indicação de ferramentas disponíveis, cuidados jurídicos específicos e nível de risco regulatório. A coluna de nível de risco não significa que os usos de risco moderado devam ser evitados: significa que exigem atenção jurídica adicional antes da implementação, especialmente em relação às bases legais de tratamento de dados e às obrigações de rotulagem.

GUIA ESTRATÉGICO DE USOS LÍCITOS DE IA EM CAMPANHAS ELEITORAIS

Usos permitidos · Ferramentas · Cuidados jurídicos · Nível de risco regulatório

CATEGORIA	USO ESTRATÉGICO	FERRAMENTAS	CUIDADOS JURÍDICOS	RISCO
PRODUÇÃO DE CONTEÚDO	Geração de textos para posts, e-mails e materiais de campanha com revisão humana.	*ChatGPT, Claude, Gemini, Jasper*	Rotulagem obrigatória. Revisão humana do conteúdo antes da publicação.	✓ BAIXO
	Criação de imagens e artes de campanha com identidade visual do candidato.	*Midjourney, DALL-E 3, Adobe Firefly*	Rotulagem obrigatória. Não usar imagens de terceiros sem autorização.	✓ BAIXO
	Síntese de voz do próprio candidato para conteúdo acessível ou multilíngue.	*ElevenLabs, Resemble AI*	Exige autorização expressa e documentada do candidato. Rotulagem obrigatória.	⚠ MOD.
ANÁLISE E DADOS	Análise de sentimento e tendências de opinião a partir de redes sociais públicas.	*Brandwatch, BERTimbau, Mention*	Dados públicos: legítimo interesse aplicável. Não armazenar dados individuais sem base legal.	✓ BAIXO
	Análise de dados eleitorais do TSE para identificação de padrões e oportunidades.	*Python (pandas), Tableau Public, Google Looker Studio*	Dados públicos: sem restrição de uso. Cuidado com inferências sobre eleitores individuais.	✓ BAIXO
	Microtargeting por perfil demográfico e geográfico com dados declarados.	*Meta Ads, Google Ads, TikTok Ads*	Base legal: consentimento ou legítimo interesse. Não usar dados sensíveis inferidos sem consentimento específico.	⚠ MOD.
ATENDIMENTO E RELACIONAMENTO	Chatbot de atendimento eleitoral com identificação obrigatória como sistema automatizado.	*ManyChat, Typebot, GPT-4 via API*	Identificação imediata como IA é obrigação legal. Multa de R$ 10k–50k pelo descumprimento.	✓ BAIXO

	Automatização de respostas a perguntas frequentes em redes sociais com supervisão humana.	*Zapier + LLM, Make, n8n*	Identificar como sistema automatizado. Garantir supervisão humana para casos complexos.	✓ BAIXO
MONITORAMENTO E DEFESA	Monitoramento de deepfakes e conteúdo irregular sobre o candidato em tempo real.	*Deepware, InVID, Microsoft Video Authenticator*	Resultado de ferramenta é triagem, não prova. Acionar perito forense para produção de prova em processo.	✓ BAIXO
	Social listening para detecção precoce de crises de imagem e narrativas adversárias.	*Brandwatch, Mention, Talkwalker*	Dados públicos: legítimo interesse aplicável. Atenção à privacidade de usuários identificáveis.	✓ BAIXO

RISCO REGULATÓRIO: ✓ **BAIXO** = uso permitido sem restrições especiais ⚠ **MODERADO** = requer atenção jurídica prévia ! **ALTO** = consultar advogado eleitoral antes de implementar

Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024, LGPD e jurisprudência do TSE (2024).

Gestão de Riscos: Mapeamento e Mitigação

A gestão de riscos jurídicos e reputacionais no uso de IA em campanhas eleitorais não começa com a escolha das ferramentas nem com a redação de contratos: começa com o mapeamento estruturado dos riscos específicos de cada tipo de atividade, porque sem diagnóstico preciso qualquer estratégia de mitigação será genérica demais para proteger o que precisa ser protegido e restritiva demais para permitir o que pode e deve ser feito. O profissional que trata todos os usos de IA como igualmente arriscados está, na prática, desincentivando usos lícitos e estrategicamente valiosos enquanto deixa sem atenção suficiente os usos que concentram o risco real de responsabilização.

A matriz de riscos que se segue classifica os principais riscos do uso de IA eleitoral em duas dimensões que, combinadas, determinam a prioridade de atenção gerencial e jurídica. A primeira dimensão é a probabilidade de ocorrência, que estima a frequência com que o risco se materializa em campanhas com o perfil descrito, com base nos padrões identificados nos ciclos eleitorais de 2022 e 2024. A segunda dimensão é o impacto potencial, que avalia as consequências da materialização do risco sobre a campanha em três planos simultâneos: o jurídico, medido pela severidade das sanções aplicáveis; o reputacional, medido pelo potencial de instrumentalização do incidente por adversários e imprensa; e o eleitoral, medido pelo potencial de influência sobre o resultado do pleito.

A combinação dessas duas dimensões produz quatro quadrantes que determinam a estratégia de resposta adequada para cada risco. Riscos de alta probabilidade e alto impacto exigem protocolo de resposta pré-estabelecido e recursos dedicados de

monitoramento: são os que mais frequentemente decidem eleições quando se materializam sem resposta adequada. Riscos de baixa probabilidade e alto impacto exigem plano de contingência documentado mesmo que nunca seja acionado: são os que produzem as crises mais difíceis de gerenciar precisamente porque a campanha não estava preparada. Riscos de alta probabilidade e baixo impacto exigem procedimentos padronizados que os tratem como rotina sem consumir atenção gerencial desproporcional. E riscos de baixa probabilidade e baixo impacto podem ser monitorados sem recursos dedicados, com revisão periódica para verificar se sua classificação mudou com a evolução tecnológica ou regulatória.

Uma ressalva metodológica é necessária antes da leitura da matriz: as classificações de probabilidade e impacto são estimativas baseadas nos padrões dos ciclos eleitorais recentes e na análise doutrinária e jurisprudencial desenvolvida nos capítulos anteriores. Elas não são previsões determinísticas e precisam ser calibradas para as características específicas de cada campanha, do cargo disputado, do perfil do candidato, do contexto político local e do nível de sofisticação dos adversários. Uma campanha presidencial enfrenta um perfil de risco substancialmente diferente de uma campanha para vereador em município de pequeno porte, e a matriz deve ser lida como ponto de partida para análise específica, não como diagnóstico definitivo aplicável a qualquer contexto.

MATRIZ DE RISCOS — IA ELEITORAL

Probabilidade · Impacto · Estratégia de Mitigação · Prioridade

RISCO	PROBABILIDADE	IMPACTO JURÍDICO	IMPACTO REPUTACIONAL	ESTRATÉGIA DE MITIGAÇÃO	PRIORIDADE
RISCOS DE CONTEÚDO SINTÉTICO					
Deepfake de adversário circulando contra o candidato	ALTA	ALTO	ALTO	Protocolo de resposta pré-estabelecido. Perito forense pré-contratado. Monitoramento em tempo real. Notificação formal à plataforma em até 2h da identificação.	P1
Conteúdo de IA da própria campanha sem rotulagem adequada	ALTA	MÉDIO	MÉDIO	Checklist de conformidade antes de cada publicação. Cláusula contratual com agências. Auditoria semanal de conteúdo publicado.	P1

Chatbot sem identificação como sistema automatizado	**MÉDIA**	**ALTO**	**MÉDIO**	Testar chatbot como usuário externo antes do lançamento. Verificar identificação na primeira mensagem. Monitoramento de reclamações.	**P2**
Conteúdo sintético publicado nos 3 dias antes do turno	**BAIXA**	**ALTO**	**ALTO**	Desativar automações de publicação para 1-3/out e 22-24/out. Verificar agenda de conteúdo programado. Briefing da equipe de comunicação.	**P2**
RISCOS DE PROTEÇÃO DE DADOS					
Microtargeting com dados sensíveis sem base legal adequada	**ALTA**	**ALTO**	**MÉDIO**	Auditoria de base legal com DPO antes de cada campanha de targeting. Substituir legítimo interesse por consentimento específico para perfis psicográficos.	**P1**
Vazamento de banco de dados de eleitores	**BAIXA**	**ALTO**	**ALTO**	Criptografia obrigatória. Controle de acesso por privilégio mínimo. Plano de resposta a incidentes com protocolo de notificação à ANPD em 72h.	**P2**
Fornecedor de dados operando sem base legal	**MÉDIA**	**ALTO**	**MÉDIO**	Auditoria prévia de práticas do fornecedor. Cláusula contratual de conformidade com LGPD. Direito de auditoria durante execução.	**P2**
RISCOS DE FINANCIAMENTO E PRESTAÇÃO DE CONTAS					
Gastos com IA não declarados na prestação de contas	**ALTA**	**ALTO**	**ALTO**	Registrar todos os contratos com fornecedores de tecnologia e marketing digital. Verificação quinzenal com tesoureiro da campanha.	**P1**
Valor de impacto de IA desproporcional ao custo declarado	**MÉDIA**	**MÉDIO**	**BAIXA**	Documentar critério de valoração das despesas com IA. Consultar advogado sobre interpretação do TSE sobre valor	**P3**

				equivalente de mercado.	
Disparo em massa pelo WhatsApp sem declaração	**MÉDIA**	**ALTO**	**ALTO**	Verificar com gestor financeiro existência de contratos não declarados. Due diligence sobre todas as agências de comunicação contratadas.	**P1**
RISCOS REPUTACIONAIS E ESTRATÉGICOS					
Adversário produz deepfake do candidato nos 3 dias pré-turno	**MÉDIA**	**MÉDIO**	**ALTO**	Protocolo de resposta rápida com perito disponível 24h. Template de declaração pública preparado. Rede de disseminação da resposta ativada.	**P1**
Exposição de estratégia de targeting por vazamento ou processo	**BAIXA**	**BAIXA**	**ALTO**	Classificar documentos de estratégia como confidenciais. Contratos de confidencialidade com todos os fornecedores. Segmentação do acesso por equipe.	**P3**
Investigação de abuso de poder econômico por uso de IA	**BAIXA**	**ALTO**	**ALTO**	Manter registro detalhado de todas as operações de IA com base legal documentada. Relatório de impacto com critério de valoração. Advogado eleitoral acompanhando prestação de contas.	**P2**

PROBABILIDADE / IMPACTO: ALTA | MÉDIA | BAIXA | MÍNIMO
PRIORIDADE: P1 = Protocolo pré-estabelecido obrigatório P2 = Plano de contingência documentado P3 = Procedimento padrão P4 = Monitoramento periódico

Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024, jurisprudência do TSE (2024) e análise dos ciclos eleitorais de 2022 e 2024.

Contratos com Empresas de Tecnologia: Cláusulas Essenciais

A responsabilidade solidária estabelecida pela Resolução TSE nº 23.732/2024 entre candidatos e seus prestadores de serviço de tecnologia não é apenas uma disposição normativa que o advogado eleitoral precisa conhecer: é o argumento mais poderoso disponível para convencer candidatos e dirigentes partidários de que a gestão contratual com fornecedores de tecnologia merece a mesma atenção que a gestão contratual com

fornecedores de qualquer outro serviço crítico da campanha. Um contrato mal redigido com a agência de marketing digital não é apenas um problema jurídico abstrato: é um vetor concreto pelo qual práticas irregulares de terceiros podem gerar responsabilização do candidato, processos eleitorais, multas e, nos casos mais graves, cassação de registro ou diploma.

Um contrato bem redigido com a empresa de marketing digital ou de tecnologia da campanha cumpre três funções simultâneas que precisam ser compreendidas separadamente para que o instrumento seja usado com sua capacidade plena. A primeira função é a de compliance: ao exigir do fornecedor declarações expressas de conformidade com a Resolução TSE nº 23.732/2024 e com a LGPD, o contrato cria o registro documental que demonstra que a campanha adotou medidas razoáveis de governança, o critério que a jurisprudência do TSE identifica como atenuante ou excludente na dimensão administrativa da responsabilização. A segunda função é a de gestão de risco: ao estabelecer vedações expressas e obrigações de notificação, o contrato cria incentivos para que o fornecedor opere dentro dos limites regulatórios sem que a campanha precise monitorar cada decisão técnica individualmente. A terceira função é a de defesa jurídica: em caso de violação pelo fornecedor, o contrato documenta que a campanha não apenas desconhecia a conduta irregular, mas havia expressamente proibido e criado mecanismo de notificação que o fornecedor descumpriu, o que é o argumento mais sólido disponível para mitigar a responsabilização administrativa e excluir a penal.

As cláusulas essenciais identificadas a seguir não são lista exaustiva e precisam ser adaptadas às características específicas de cada relação contratual. Um contrato com empresa de análise de dados tem perfil de risco distinto de um contrato com agência de produção de conteúdo, que por sua vez difere de um contrato com fornecedor de chatbot. O advogado que adapta as cláusulas ao perfil específico do fornecedor está produzindo instrumento mais eficaz do que o que aplica cláusulas genéricas indiscriminadamente. A especificidade é o que transforma o contrato de formalidade burocrática em instrumento de proteção real.

Um ponto que merece destaque antes da apresentação das cláusulas: a negociação contratual com fornecedores de tecnologia eleitoral é frequentemente conduzida sob pressão de tempo, com o início do período eleitoral próximo e a campanha ansiosa para operacionalizar suas ferramentas. Essa pressão produz contratos assinados sem leitura adequada, com cláusulas genéricas copiadas de modelos anteriores e sem as disposições específicas que o marco regulatório atual exige. O custo dessa negligência contratual pode

ser substancialmente maior do que o tempo investido numa negociação cuidadosa. A orientação prática mais simples disponível é a seguinte: nenhum contrato com fornecedor de tecnologia eleitoral deve ser assinado sem revisão de advogado com conhecimento simultâneo de direito eleitoral e proteção de dados, e essa revisão deve ocorrer antes do início das negociações, não depois da apresentação da proposta final.

CLÁUSULAS ESSENCIAIS — CONTRATOS COM FORNECEDORES DE TECNOLOGIA ELEITORAL

Resolução TSE nº 23.732/2024 · LGPD · Responsabilidade Solidária

Nº	CLÁUSULA	TEXTO RECOMENDADO / CONTEÚDO MÍNIMO	FINALIDADE
BLOCO 1 — CONFORMIDADE REGULATÓRIA			
1	**Declaração de Conformidade**	O CONTRATADO declara expressamente conhecer e cumprir integralmente as disposições da Resolução TSE nº 23.732/2024 e da Lei nº 13.709/2018 (LGPD), obrigando-se a informar imediatamente o CONTRATANTE de qualquer alteração regulatória que impacte as atividades contratadas.	COMPLIANCE
2	**Rotulagem de Conteúdo Gerado por IA**	Todo conteúdo gerado ou significativamente modificado por inteligência artificial produzido no âmbito deste contrato deverá ser entregue ao CONTRATANTE já com identificação clara e ostensiva, na forma exigida pelo art. 15 da Resolução TSE nº 23.732/2024, sendo vedada a publicação de conteúdo sem essa identificação.	COMPLIANCE
3	**Declaração de Ferramentas Utilizadas**	O CONTRATADO obriga-se a declarar, por escrito e com antecedência mínima de 48 horas, toda ferramenta de inteligência artificial a ser utilizada na execução do contrato, incluindo nome, fornecedor, versão e finalidade de uso, mantendo registro atualizado disponível ao CONTRATANTE a qualquer tempo.	COMPLIANCE
BLOCO 2 — VEDAÇÕES EXPRESSAS			
4	**Vedação de Deepfakes de Terceiros**	É expressamente vedado ao CONTRATADO criar, modificar, encomendarem ou disseminar conteúdo audiovisual sintético que represente candidatos, autoridades ou pessoas públicas que não o próprio CONTRATANTE, sob pena de rescisão imediata e sem prejuízo da responsabilização civil e eleitoral correspondente.	DEFESA
5	**Vedação de Síntese de Voz Não Autorizada**	É vedado ao CONTRATADO utilizar tecnologia de clonagem ou síntese de voz para produzir conteúdo atribuído a qualquer pessoa que não tenha dado consentimento específico, expresso e documentado para essa finalidade.	DEFESA

6	**Vedação de Chatbot Não Identificado**	Toda interação automatizada com eleitores conduzida por sistemas de inteligência artificial deverá identificar-se imediatamente, na primeira mensagem, como sistema automatizado, sendo vedada qualquer simulação de interação humana sem essa identificação prévia.	🛡 DEFESA
BLOCO 3 — RESPONSABILIDADE E GESTÃO DE INCIDENTES			
7	**Notificação de Irregularidades**	O CONTRATADO obriga-se a notificar o CONTRATANTE, no prazo máximo de 2 (duas) horas, ao identificar qualquer conteúdo irregular envolvendo a campanha, qualquer investigação ou contato de autoridade eleitoral ou de proteção de dados relacionado ao contrato, e qualquer incidente de segurança que afete dados de eleitores.	⚠ GESTÃO DE RISCO
8	**Responsabilidade Solidária Reversa**	O CONTRATADO responde solidariamente pelos danos eleitorais, reputacionais e financeiros decorrentes de violações à Resolução TSE nº 23.732/2024 ou à LGPD praticadas no âmbito deste contrato, comprometendo-se a indenizar o CONTRATANTE por multas, custas processuais e danos comprovados resultantes de sua conduta irregular.	🛡 DEFESA
9	**Direito de Auditoria**	O CONTRATANTE reserva-se o direito de auditar, a qualquer tempo e mediante aviso prévio de 24 horas, as práticas de uso de inteligência artificial e de tratamento de dados pessoais do CONTRATADO relacionadas à execução deste contrato, podendo contratar auditores independentes para essa finalidade às expensas do CONTRATANTE.	⚠ GESTÃO DE RISCO
BLOCO 4 — PROTEÇÃO DE DADOS E ENCERRAMENTO			
10	**Base Legal para Tratamento de Dados**	O CONTRATADO declara que todo tratamento de dados pessoais de eleitores realizado no âmbito deste contrato tem base legal adequada nos termos da LGPD, especificando a base legal aplicável a cada operação de tratamento no documento de registro de atividades que deverá ser mantido e disponibilizado ao CONTRATANTE.	🔒 PROTEÇÃO DE DADOS
11	**Retenção e Destruição de Dados**	Ao término do contrato ou mediante solicitação do CONTRATANTE, o CONTRATADO obriga-se a devolver ou destruir, no prazo máximo de 15 (quinze) dias, todos os dados pessoais de eleitores tratados no âmbito do contrato, fornecendo certificado de destruição com especificação dos dados eliminados e do método utilizado.	🔒 PROTEÇÃO DE DADOS
12	**Notificação de Incidente de Segurança**	Em caso de incidente de segurança que afete dados pessoais de eleitores, o CONTRATADO obriga-se a notificar o CONTRATANTE no prazo máximo de 12 (doze) horas após a identificação do incidente, com informações suficientes para que o CONTRATANTE cumpra sua obrigação de notificação à ANPD no prazo de 72 horas previsto no art. 48 da LGPD.	🔒 PROTEÇÃO DE DADOS

⚠ **ATENÇÃO:** Nenhum contrato com fornecedor de tecnologia eleitoral deve ser assinado sem revisão de advogado com conhecimento simultâneo de direito eleitoral e proteção de dados. A revisão deve ocorrer antes das negociações, não após a apresentação da proposta final.

Fonte: Elaborado pelo autor com base na Resolução TSE nº 23.732/2024, art. 27; LGPD, arts. 37, 46 e 48; e jurisprudência do TSE sobre responsabilidade solidária (2024).

ATENÇÃO PRÁTICA — Para o Advogado Eleitoral

A ausência de cláusula de conformidade com a Resolução TSE nº 23.732/2024 em contratos com fornecedores de tecnologia eleitoral é, por si só, um fator agravante na eventual responsabilização do candidato por condutas irregulares praticadas pelo fornecedor. O TSE tem interpretado a ausência de due diligence contratual como indicativo de omissão culposa do candidato, incompatível com a alegação de desconhecimento total da conduta irregular. A cláusula de conformidade não elimina a responsabilidade solidária, mas demonstra que o candidato adotou as medidas preventivas razoavelmente exigíveis.

O Papel do DPO na Campanha Eleitoral

A figura do Data Protection Officer, denominado pela LGPD como Encarregado de Proteção de Dados no artigo 41, é frequentemente negligenciada em campanhas eleitorais brasileiras, especialmente nas de menor porte, e essa negligência é compreensível por uma razão prática: o DPO é percebido como uma exigência de grandes organizações com estruturas corporativas permanentes, não como uma necessidade de campanhas eleitorais com duração limitada e equipes temporárias. Essa percepção é equivocada, e o equívoco tem consequências regulatórias que precisam ser antecipadas antes do ciclo eleitoral de 2026.

A obrigação de nomear DPO na LGPD não é universal: ela se aplica a controladores que realizam tratamento de dados pessoais em larga escala, conforme critérios que a ANPD está progressivamente consolidando em orientações e regulamentos. O que é relevante para o contexto eleitoral é que campanhas que utilizam sistemas de microtargeting com bases de dados de eleitores, CRM eleitoral com registro de interações individualizadas, ferramentas de análise comportamental ou plataformas de automação de comunicação política enquadram-se na categoria de tratamento em larga escala independentemente do porte da campanha. O critério não é o tamanho do orçamento: é o volume e a natureza do tratamento de dados. Uma campanha para prefeito de cidade de médio porte que contrata agência de marketing com acesso a banco de dados de cinquenta mil eleitores pode estar realizando tratamento em larga escala para fins da LGPD.

A ausência de DPO não é apenas uma infração isolada: ela é o indicador mais direto de que a campanha não tem a estrutura de governança de dados que a LGPD pressupõe. Quando a ANPD ou a Justiça Eleitoral iniciam uma investigação sobre uso irregular de dados em campanha, a primeira pergunta é quem é o responsável pelo programa de proteção de dados da organização. Uma campanha sem DPO identificado não tem resposta para essa pergunta, o que sinaliza ausência de governança que agrava a avaliação de todas as outras condutas sob investigação.

As atribuições do DPO em contexto eleitoral precisam ser compreendidas em sua especificidade, porque o papel padrão descrito na LGPD foi pensado para organizações permanentes e precisa ser adaptado às características do ambiente de campanha. A primeira atribuição é a verificação contínua da conformidade do tratamento de dados com a LGPD, o que em contexto eleitoral inclui especificamente a verificação das bases legais de cada operação de microtargeting, a conformidade das cláusulas contratuais com fornecedores de tecnologia e a adequação dos formulários de coleta de dados com as exigências de transparência da lei.

A segunda é o atendimento às solicitações de titulares de dados no prazo de quinze dias estabelecido pela ANPD: o DPO precisa ter acesso aos sistemas de dados da campanha para ser capaz de responder a pedidos de acesso, correção e eliminação dentro do prazo, o que exige que esse acesso seja configurado antes do início do período eleitoral. A terceira é a comunicação de incidentes de segurança à ANPD dentro do prazo de setenta e duas horas, função que como analisado na seção sobre segurança da informação precisa estar integrada ao plano de resposta a incidentes da campanha.

A quarta é a orientação da equipe de campanha sobre boas práticas de proteção de dados, incluindo treinamentos mínimos para os membros da equipe que têm acesso a dados de eleitores. E a quinta é atuar como canal de comunicação entre a campanha e a ANPD em caso de investigação, papel que em contexto de dupla exposição regulatória precisará ser coordenado com a estratégia de defesa perante a Justiça Eleitoral.

Para campanhas de menor porte, a exigência de DPO não precisa ser atendida por profissional com dedicação exclusiva. A LGPD permite que o encarregado seja pessoa física ou jurídica, o que significa que um advogado com especialização em proteção de dados pode exercer a função de DPO para a campanha no âmbito de um contrato de consultoria que inclua as atribuições específicas do cargo. Essa solução é tanto juridicamente adequada quanto economicamente acessível para campanhas que não têm estrutura para contratar DPO em regime de dedicação exclusiva. A condição é que o

profissional contratado tenha efetivamente as competências técnicas e jurídicas necessárias para exercer as atribuições do cargo, o que exige verificação antes da contratação, e que sua nomeação seja formalizada e publicada nos termos do artigo 41 da LGPD, porque a nomeação informal sem publicação não produz os efeitos jurídicos que a lei associa à figura.

A identidade e os dados de contato do DPO precisam ser divulgados publicamente, o que para campanhas eleitorais significa incluir essa informação na política de privacidade disponível no site da campanha e nos formulários de coleta de dados utilizados. Essa divulgação não é apenas obrigação formal: ela é o instrumento que permite ao eleitor que quiser exercer seus direitos saber a quem se dirigir, e que permite à ANPD e à Justiça Eleitoral identificar o responsável pelo programa de proteção de dados em caso de investigação. Campanhas que omitem essa informação estão dificultando o exercício de direitos dos titulares e sinalizam para as autoridades que a estrutura de governança de dados não está implementada de forma operacional.

O que Este Capítulo Construiu: Uma Arquitetura de Conformidade

Os instrumentos apresentados ao longo deste capítulo, o checklist de conformidade, o guia de usos lícitos, a matriz de riscos, o modelo de cláusulas contratuais e o protocolo de resposta a incidentes, não são peças independentes que o profissional escolhe conforme a conveniência do momento. São componentes de uma arquitetura de conformidade que só funciona quando operada de forma integrada e, crucialmente, quando implementada antes do período eleitoral e não em resposta a crises já instaladas.

A lógica que conecta esses instrumentos é simples e precisa ser internalizada como princípio operacional: o checklist identifica o que precisa estar em ordem; o guia de usos lícitos define o espaço dentro do qual a campanha pode operar com segurança; a matriz de riscos prioriza onde concentrar atenção e recursos; as cláusulas contratuais transferem parte do risco para os fornecedores que têm capacidade técnica de mitigá-lo; e o protocolo de resposta a incidentes garante que quando algo der errado, e em algum momento algo dará, a campanha não precisará improvisar sob pressão máxima.

Há um ponto que este capítulo procurou demonstrar ao longo de todas as suas seções e que merece ser enunciado explicitamente na conclusão: conformidade regulatória e efetividade estratégica não são objetivos em tensão no uso de IA eleitoral. São objetivos convergentes. Os usos de IA que geram maior responsabilização jurídica são frequentemente os que produzem menor retorno eleitoral sustentável: deepfakes de

adversários que podem ser desmentidos, chatbots não identificados que geram desconfiança quando descobertos, microtargeting com dados sem base legal que expõe a campanha a investigações que paralisam operações em momento crítico. Os usos de IA que geram menor responsabilização jurídica são frequentemente os de maior potencial estratégico: análise de dados eleitorais que identifica oportunidades invisíveis, produção de conteúdo em escala com qualidade consistente, monitoramento em tempo real que permite resposta antes que crises se consolidem.

A campanha que chega a 2026 com arquitetura de conformidade implementada não está apenas protegida juridicamente: está estruturalmente em vantagem competitiva sobre adversários que chegam sem esse preparo. Quando um deepfake surge nos três dias antes do primeiro turno, a campanha com protocolo pré-estabelecido, perito forense contratado e equipe de resposta treinada reage em horas; a campanha que improvisa reage em dias, quando o dano já é irreversível. Quando a ANPD ou a Justiça Eleitoral investigam uso de dados, a campanha com DPO nomeado, base legal documentada e registros de operações atualizados tem defesa organizada desde o primeiro dia; a campanha sem essa estrutura começa a construir sua defesa enquanto responde a questionamentos que já acumulam pressão processual.

O direito eleitoral na era da inteligência artificial não é um campo que penaliza quem usa tecnologia: é um campo que penaliza quem usa tecnologia sem governança. A distinção entre os dois perfis é precisamente o que este capítulo procurou construir em termos operacionais. Os quatorze capítulos anteriores forneceram o arcabouço teórico, jurídico e técnico. Este capítulo forneceu os instrumentos para transformar esse arcabouço em prática. O que vem a seguir é decisão de cada profissional: usar esses instrumentos antes do problema, ou lembrar deles depois.

Uma nota final sobre o calendário.

As eleições gerais de 2026 ocorrem em 4 de outubro, primeiro turno, e 25 de outubro, segundo turno. O período de campanha oficial começa em 16 de agosto. A vedação de conteúdo sintético, mesmo rotulado, começa em 1º de outubro para o primeiro turno e em 22 de outubro para o segundo. A cooperação entre ANPD e TSE estará mais desenvolvida em 2026 do que estava em 2024. Os precedentes do TSE das eleições municipais de 2024 estarão disponíveis como parâmetro interpretativo. E as ferramentas

de IA generativa disponíveis em outubro de 2026 serão substancialmente mais poderosas do que as disponíveis quando este livro foi escrito.

Há tempo para preparar a arquitetura de conformidade. Mas esse tempo é finito, e cada semana que passa sem implementação é uma semana de exposição que o calendário eleitoral não devolve.

■ RESUMO DO CAPÍTULO:

Este capítulo teve propósito distinto dos quatorze anteriores: não construir arcabouço teórico, jurídico ou técnico, mas traduzir esse arcabouço em instrumentos operacionais concretos para os atores que atuam diretamente nas campanhas eleitorais. O argumento central que atravessou todas as seções é que conformidade regulatória e efetividade estratégica são objetivos convergentes no uso de IA eleitoral, não objetivos em tensão. A abertura estabeleceu a distinção fundamental entre compreender o campo e saber operar nele, identificando a segunda como o problema específico que este capítulo endereça. A observação metodológica sobre o uso correto dos instrumentos apresentados, como organizadores do julgamento e não substitutos dele, foi estabelecida como premissa que orienta toda a leitura posterior. O checklist de conformidade com a Resolução TSE nº 23.732/2024 organizou as principais obrigações em doze itens verificáveis, distribuídos em quatro blocos temáticos: conteúdo gerado por IA, proteção de dados e LGPD, financiamento e prestação de contas, e protocolo de resposta e governança. A dupla função do checklist, preventiva antes do período eleitoral e de auditoria corrente durante a campanha, foi identificada como o que o distingue de mera formalidade burocrática. O guia estratégico de usos lícitos mapeou dez aplicações de IA em quatro categorias funcionais, produção de conteúdo, análise e dados, atendimento e relacionamento, e monitoramento e defesa, com ferramentas concretas, cuidados jurídicos específicos e nível de risco regulatório para cada uso. O argumento subjacente ao guia é que os usos mais eficazes de IA são frequentemente os de menor risco jurídico, porque operam na transparência e no respeito à autonomia do eleitor. A matriz de riscos classificou doze riscos em quatro blocos, conteúdo sintético, proteção de dados, financiamento e riscos reputacionais, em duas dimensões combinadas de probabilidade e impacto, com estratégia de mitigação específica e classificação de prioridade P1 a P4 para cada risco. O argumento estruturante da matriz é que sem diagnóstico preciso de risco qualquer estratégia de mitigação será genérica demais para proteger o que precisa ser protegido e restritiva demais para permitir o que pode ser feito. As cláusulas essenciais para contratos com fornecedores de tecnologia eleitoral foram organizadas em doze cláusulas distribuídas em quatro blocos, conformidade regulatória, vedações expressas, responsabilidade e gestão de incidentes, e proteção de dados e encerramento. As três funções simultâneas do contrato bem redigido, compliance, gestão de risco e defesa jurídica, foram desenvolvidas para demonstrar que a negligência contratual não é apenas problema jurídico abstrato mas vetor concreto de responsabilização. O papel do DPO na campanha eleitoral foi analisado como a exigência regulatória mais frequentemente negligenciada, com desenvolvimento do critério de larga escala aplicado ao contexto eleitoral, das cinco atribuições específicas do cargo adaptadas ao ambiente de campanha e da solução prática do DPO por contrato de consultoria para campanhas de menor porte. A finalização integrou os cinco instrumentos como componentes de uma arquitetura de conformidade que só funciona de forma integrada e que precisa ser implementada antes do período eleitoral. A vantagem competitiva da campanha preparada sobre a que improvisa foi demonstrada com o exemplo concreto dos três dias pré-turno. A nota final sobre o calendário de 2026 transformou o argumento abstrato em urgência temporal com datas concretas: 16 de agosto para o início das obrigações de rotulagem, 1º de outubro para a vedação de conteúdo sintético no primeiro turno e 22 de outubro para o segundo.

QUESTÕES PARA REFLEXÃO E DEBATE

1. O capítulo argumenta que conformidade regulatória e efetividade estratégica são objetivos convergentes no uso de IA eleitoral. Essa convergência é estrutural, resultado do design regulatório que penaliza os usos menos eficazes a longo prazo, ou é circunstancial, válida apenas enquanto o sistema de fiscalização não tem capacidade de detectar e sancionar violações em escala? O que acontece com essa convergência quando a probabilidade de detecção é percebida pelos atores como baixa?

2. O checklist de conformidade funciona como organizador do julgamento, não como substituto dele. Casos que não se encaixam nas categorias previstas pelo instrumento exigem avaliação que nenhuma checklist fornece. Que mecanismos institucionais, além da consultoria jurídica individualizada, poderiam ser desenvolvidos para apoiar campanhas de menor porte na navegação de situações que os instrumentos padronizados não cobrem adequadamente?

3. A matriz de riscos classifica o microtargeting com dados sensíveis sem base legal adequada como risco de alta probabilidade e alto impacto, com prioridade P1. Isso implica que a maioria das campanhas brasileiras que utilizam ferramentas de análise comportamental está operando com o risco mais crítico da matriz sem protocolo de resposta estabelecido. Que intervenção regulatória, da ANPD, do TSE ou de ambas em conjunto, seria mais eficaz para reduzir a probabilidade desse risco antes de 2026: orientação e safe harbors, fiscalização preventiva com alertas ou sanção exemplar de casos identificados?

4. As cláusulas contratuais essenciais apresentadas no capítulo transferem parte do risco para os fornecedores de tecnologia, mas a responsabilidade administrativa do candidato perante a Justiça Eleitoral persiste independentemente do que o contrato estabeleça. Essa estrutura de responsabilidade cria o incentivo correto para que fornecedores operem dentro dos limites regulatórios, ou ela reduz o incentivo dos fornecedores à conformidade ao desonerar suas consequências mais graves para o candidato?

5. O DPO por contrato de consultoria é apresentado como solução acessível para campanhas de menor porte. Essa solução pressupõe que advogados com especialização simultânea em direito eleitoral e proteção de dados estejam disponíveis no mercado em número suficiente para atender a demanda das campanhas de 2026. Essa premissa é realista para o contexto brasileiro atual, e o que poderia ser feito pelos órgãos de classe, pelas faculdades de direito e pelo próprio TSE para desenvolver essa capacidade antes do ciclo eleitoral?

6. A nota final do capítulo estabelece que há tempo para preparar a arquitetura de conformidade, mas que esse tempo é finito. Para campanhas que chegarem ao período eleitoral sem essa arquitetura implementada, qual é a estratégia de conformidade mais eficaz disponível quando o tempo de preparação já se esgotou: priorização dos itens de maior risco, busca de safe harbors regulatórios, suspensão das operações de maior risco ou aceitação calculada da exposição com reserva de recursos para defesa?

7. O protocolo de resposta a incidentes de deepfake pressupõe que a campanha tem perito forense pré-contratado, advogado eleitoral com modelo de representação preparado e sistema de monitoramento ativo. Essa infraestrutura tem custo que coloca campanhas presidenciais e estaduais bem financiadas em vantagem estrutural sobre campanhas municipais de menor porte. O sistema eleitoral deveria criar algum mecanismo de acesso coletivo a essa infraestrutura, como um serviço centralizado do TSE de resposta a deepfakes, ou a responsabilidade pela preparação deve permanecer inteiramente com cada campanha individualmente?

8. A convergência entre conformidade e efetividade estratégica argumentada no capítulo funciona para atores que têm horizonte de longo prazo em política, candidatos com carreiras em construção e partidos institucionais. Para atores com horizonte de curto prazo, candidatos em única disputa ou operadores sem compromisso com a continuidade da campanha, a lógica de convergência pode não se aplicar da

mesma forma. Como o design regulatório deveria endereçar especificamente esse subconjunto de atores, para os quais as sanções de longo prazo como inelegibilidade têm menor poder deterrente?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. Este capítulo é um mapa, não um algoritmo. Use-o para organizar o julgamento, não para substituí-lo.

Os instrumentos deste capítulo, checklist, guia de usos lícitos, matriz de riscos, cláusulas contratuais e protocolo de resposta, foram projetados para reduzir a complexidade a categorias manejáveis que permitem ação em tempo útil. Mas casos que não se encaixam nas categorias previstas são frequentemente os mais importantes, e nenhum instrumento padronizado os resolve. O profissional que usa o checklist como lista de verificação burocrática está usando-o incorretamente. O que usa os instrumentos como estrutura que organiza seu julgamento sobre as especificidades de cada campanha está usando-os da forma pretendida. A consultoria jurídica especializada não é opcional nos casos de maior complexidade: é a camada que os instrumentos não conseguem fornecer.

2. Três datas do calendário de 2026 que não podem ser ignoradas por nenhuma campanha.

16 de agosto de 2026: início da propaganda eleitoral e vigência plena das obrigações de rotulagem da Resolução TSE 23.732/2024 sobre todo conteúdo publicado. 1º de outubro de 2026: vedação total de qualquer novo conteúdo sintético, mesmo rotulado, para o primeiro turno de 4 de outubro. 22 de outubro de 2026: a mesma vedação para o segundo turno de 25 de outubro. Essas três datas precisam estar no calendário operacional da campanha com antecedência suficiente para que automações sejam desativadas, agendamentos revisados e a equipe de comunicação briefada. Descobrir essas datas no dia anterior é descobri-las tarde demais.

3. A arquitetura de conformidade tem que existir antes do incidente, não depois.

O checklist identifica o que precisa estar em ordem. O guia define o espaço seguro de operação. A matriz prioriza onde concentrar recursos. As cláusulas contratuais transferem parte do risco aos fornecedores. O protocolo garante que quando algo der errado a campanha não improvise sob pressão máxima. Esses cinco instrumentos só funcionam como arquitetura quando implementados de forma integrada e antes do período eleitoral. Um checklist preenchido na semana do primeiro turno não é conformidade preventiva: é documentação retrospectiva que chega tarde para prevenir e cedo demais para defender.

4. Conformidade não é obstáculo ao uso estratégico de IA. É sua condição de sustentabilidade.

Os usos de IA que geram maior responsabilização jurídica, deepfakes de adversários, chatbots não identificados, microtargeting com dados sem base legal, são frequentemente os de menor retorno eleitoral sustentável: podem ser desmentidos, geram desconfiança quando descobertos e paralisam operações em momento crítico. Os usos de menor risco jurídico, análise de dados do TSE, produção de conteúdo com rotulagem, monitoramento em tempo real, são os de maior potencial estratégico de longo prazo. A campanha que opera dentro dos limites regulatórios não está operando com menos ferramentas: está operando com as ferramentas que não virarão contra ela quando o adversário, a imprensa ou a Justiça Eleitoral identificarem a irregularidade.

5. Nome seu DPO antes de precisar dele. Quando precisar, não haverá tempo para fazê-lo.

A ausência de DPO não é apenas infração isolável: é o sinal mais direto de que a campanha não tem estrutura de governança de dados. Quando a ANPD ou a Justiça Eleitoral investigam uso irregular de dados, a primeira pergunta é quem é o responsável pelo programa de proteção de dados. Uma campanha sem DPO identificado não tem resposta, o que agrava a avaliação de todas as outras condutas. Para campanhas de menor porte, um advogado com especialização em proteção de dados pode exercer a função de DPO por contrato de consultoria, sem necessidade de dedicação exclusiva. A condição é que a nomeação seja formal, publicada no site da campanha, e que o profissional tenha acesso real aos sistemas de dados para poder exercer as atribuições do cargo.

6. A vantagem competitiva da campanha preparada é real e mensurável em horas.

Quando um deepfake surge nos três dias antes do primeiro turno, a campanha com protocolo pré-estabelecido, perito forense contratado e equipe de resposta treinada reage em horas. A que improvisa reage em dias, quando o dano já é irreversível. Quando a ANPD ou a Justiça Eleitoral investigam uso de dados, a campanha com DPO nomeado, base legal documentada e registros atualizados tem defesa organizada desde o primeiro contato. A que não tem essa estrutura começa a construir sua defesa enquanto já responde a questionamentos que acumulam pressão processual. A diferença entre as duas situações é precisamente o tempo de preparação investido antes do período eleitoral. Esse tempo existe agora. Depois de agosto de 2026, não existirá mais.

7. Revise seus contratos com fornecedores de tecnologia antes de 16 de agosto de 2026.

A responsabilidade administrativa do candidato pelo conteúdo produzido por fornecedores de tecnologia é objetiva: ele responde independentemente de conhecimento ou autorização específica da conduta irregular. A única proteção eficaz é o contrato que documenta a proibição expressa, a obrigação de conformidade e a responsabilidade do fornecedor pelas violações que praticar. Contratos assinados antes da Resolução TSE 23.732/2024, ou antes das atualizações regulatórias de 2026, precisam ser revisados e aditados. Essa revisão precisa ser feita por advogado com conhecimento simultâneo de direito eleitoral e proteção de dados, e precisa ocorrer antes do início do período de campanha, não durante ele.

CAPÍTULO 16

Literacia Midiática como Defesa Democrática

Educação, Cidadania Digital e a Construção de Resiliência contra a Manipulação Algorítmica

Literacia Midiática e Resiliência Democrática

A literacia midiática, entendida como a capacidade de acessar, analisar, avaliar e criar conteúdo midiático de forma crítica e reflexiva, é reconhecida pela literatura especializada como um dos fatores mais determinantes da resiliência de uma população à desinformação e à manipulação algorítmica. Sua relevância para o contexto eleitoral não é periférica: ela é estrutural. Como demonstrado ao longo desta obra, as tecnologias de desinformação eleitoral exploram lacunas na capacidade crítica do eleitor, e nenhuma regulação ou ferramenta de detecção consegue fechar essas lacunas de forma comparável ao que a literacia midiática pode produzir quando desenvolvida em escala populacional.

A correlação entre níveis de literacia midiática e resistência à desinformação eleitoral foi documentada em múltiplos estudos comparativos internacionais com metodologias distintas e resultados convergentes. O que essa convergência metodológica indica não é apenas que literacia midiática reduz a vulnerabilidade individual à desinformação: é que populações com maiores níveis de literacia midiática criam ecossistemas informacionais estruturalmente menos hospitaleiros à desinformação em escala, porque a camada de amplificação orgânica, aquela composta por cidadãos que redistribuem conteúdo de boa-fé sem verificação, é substancialmente menor. Como argumentado no Capítulo 10, essa camada não pode ser alcançada por sanção jurídica porque seus participantes não cometem ato ilícito: ela só pode ser reduzida por investimento em capacidade crítica. A literacia midiática é, nessa perspectiva, o único instrumento que protege o eleitorado onde a regulação não chega.

Hobbs (2010), em *Digital and Media Literacy*, oferece o enquadramento conceitual mais influente do campo ao argumentar que a literacia midiática no século XXI vai além da capacidade de operar dispositivos tecnológicos. Ela abrange o conjunto de competências críticas necessárias para participar de forma autônoma e responsável no ecossistema informacional digital: a capacidade de identificar a fonte e os interesses que orientam a produção de um conteúdo, de avaliar a evidência que sustenta uma afirmação, de reconhecer as técnicas retóricas e emocionais utilizadas para persuadir, e de produzir conteúdo próprio com consciência de seu impacto sobre o ambiente informacional

compartilhado (HOBBS, 2010, p. 17-22). Essa concepção ampla é analiticamente relevante porque situa a literacia midiática como competência democrática, não apenas como habilidade técnica: ela é condição necessária para o exercício pleno da autonomia política que a democracia deliberativa pressupõe.

Potter (2016), em *Media Literacy*, aprofunda a dimensão cognitiva do conceito ao descrever os processos mentais pelos quais leitores críticos avaliam a credibilidade e a relevância de conteúdo midiático. O argumento central de Potter é que a literacia midiática não é uma capacidade binária, presente ou ausente, mas um continuum de competências que se desenvolvem com exposição, prática e reflexão. Indivíduos com maior literacia midiática não são imunes à desinformação: são mais lentos em aceitá-la sem questionamento e mais propensos a buscar verificação antes de redistribuí-la. Essa diferença de velocidade de processamento é, em escala populacional, o mecanismo que reduz a velocidade de disseminação de desinformação e aumenta a janela temporal disponível para intervenções de fact-checking e correção (POTTER, 2016, p. 45-52).

A pesquisa de Pennycook e Rand (2019), já referenciada em capítulos anteriores desta obra, fornece a evidência experimental mais diretamente aplicável ao contexto eleitoral: intervenções breves de literacia midiática, como a simples solicitação para que o participante avalie a precisão de uma notícia antes de compartilhá-la, produzem reduções mensuráveis na intenção de compartilhar conteúdo desinformativo sem reduzir a intenção de compartilhar conteúdo verdadeiro. O mecanismo identificado pelos autores é a reativação da atenção à acurácia, que o design das plataformas sistematicamente suprime em favor do engajamento emocional (PENNYCOOK; RAND, 2019, p. 2522-2524). Essa evidência tem uma implicação regulatória direta que ainda não foi suficientemente explorada: intervenções de literacia midiática são escaláveis pelas plataformas a custo marginal próximo de zero, o que as torna candidatas naturais a obrigações regulatórias que o DSA europeu já começou a mapear.

A evidência comparada mais citada sobre o impacto da literacia midiática na resiliência democrática vem da Finlândia. O país integrou a educação midiática ao currículo nacional desde o ensino fundamental, com progressão de competências que começa na identificação de fontes confiáveis no ensino primário e avança para a análise de técnicas de propaganda e manipulação algorítmica no ensino médio. Os resultados mensuráveis desse investimento são consistentes: a Finlândia apresenta os menores índices de vulnerabilidade à desinformação na Europa no Media Literacy Index anual produzido pelo Open Society Institute, resultado que se mantém estável ao longo de múltiplas edições do

índice e que é atribuído explicitamente ao programa nacional de educação midiática pela literatura comparada (OPEN SOCIETY INSTITUTE, 2023, p. 14-18).

A experiência finlandesa demonstra três pontos que a análise comparada permite afirmar com segurança. O primeiro é que os retornos do investimento em literacia midiática são mensuráveis e acumulam-se ao longo de décadas, o que significa que o horizonte de avaliação adequado não é o ciclo eleitoral corrente mas a geração que estará votando nos próximos vinte anos. O segundo é que a literacia midiática produzida por intervenções educacionais formais é significativamente mais robusta do que a produzida por campanhas pontuais de conscientização, porque desenvolve processos cognitivos que se aplicam a novas formas de desinformação que não existiam quando o indivíduo foi treinado. O terceiro é que o investimento em literacia midiática beneficia democraticamente todo o espectro político, porque reduz a vulnerabilidade à desinformação independentemente de qual candidato ou posição a desinformação favorece, o que o torna o instrumento mais neutro e mais legitimo disponível para o fortalecimento da integridade eleitoral.

O Déficit de Literacia Midiática no Brasil e suas Implicações Eleitorais

O Brasil apresenta um déficit estrutural de literacia midiática que é simultaneamente produto de desigualdades históricas no sistema educacional e fator amplificador da vulnerabilidade à desinformação eleitoral. Esse déficit não é uniforme: ele se distribui de forma assimétrica entre regiões, faixas etárias e estratos de escolaridade, criando padrões de vulnerabilidade que têm implicações diretas para a análise eleitoral e que nenhuma intervenção de curto prazo consegue endereçar de forma abrangente.

O estudo do Reuters Institute Digital News Report de 2023, que inclui o Brasil em sua amostra comparada, identificou que apenas 28% dos brasileiros afirmam conseguir identificar com segurança quando uma notícia é falsa, um dos índices mais baixos entre os países pesquisados e substancialmente inferior à média europeia. Esse dado precisa ser interpretado com cuidado metodológico: a autopercepção de capacidade de identificação de desinformação não equivale à capacidade efetiva, e pesquisas experimentais consistentemente demonstram que pessoas tendem a superestimar sua própria resistência à desinformação, fenômeno denominado third-person effect na literatura de comunicação. Isso significa que o percentual real de brasileiros vulneráveis à desinformação é provavelmente superior ao que o dado de autopercepção captura.

O GPOPAI da USP documentou, em análise das eleições de 2018, padrões de exposição e vulnerabilidade à desinformação que confirmam a distribuição assimétrica do déficit. Os grupos demográficos com menor escolaridade e maior dependência do WhatsApp como fonte primária de informação política eram proporcionalmente mais expostos a conteúdo desinformativo e menos capazes de verificar sua autenticidade antes de redistribuí-lo. Esse dado tem uma implicação que vai além do diagnóstico: ele indica que a desinformação eleitoral no Brasil opera com eficácia seletiva, afetando de forma desproporcional os segmentos do eleitorado com menor capital informacional para avaliá-la criticamente. Essa seletividade não é aleatória: é o resultado de operações de targeting que, como documentado no Capítulo 4, identificam e exploram deliberadamente as populações mais vulneráveis.

A dimensão geracional do déficit merece análise específica porque contraria a intuição dominante. Adultos maduros que chegaram ao ambiente digital pela primeira vez por meio do WhatsApp, fenômeno documentado por Araújo e Oliveira (2021) como um dos fatores centrais das eleições de 2018, combinam baixa familiaridade com os marcadores visuais que usuários experientes associam a fontes não confiáveis com alta confiança no conteúdo recebido de pessoas próximas. Esse perfil cria vulnerabilidade específica que intervenções de literacia midiática projetadas para jovens digitalmente nativos frequentemente não alcançam, porque pressupõem competências básicas de navegação no ecossistema digital que parte do eleitorado adulto brasileiro ainda não desenvolveu (ARAÚJO; OLIVEIRA, 2021, p. 45-51).

A distribuição regional do déficit acompanha, com correlação elevada, a distribuição das desigualdades educacionais históricas: regiões com menores índices de conclusão do ensino médio apresentam sistematicamente maiores índices de exposição a desinformação e menores índices de verificação de conteúdo antes do compartilhamento. Essa correlação tem implicação estratégica para qualquer programa de literacia midiática eleitoral que pretenda ter impacto antes de 2026: intervenções que não alcançam especificamente as regiões e os grupos demográficos de maior vulnerabilidade estão concentrando recursos onde o impacto marginal é menor.

O déficit de literacia midiática tem ainda uma dimensão institucional que completa o diagnóstico: as instituições com maior potencial de escala para programas de literacia midiática, escolas, universidades, igrejas, sindicatos e associações comunitárias, não têm, na maioria dos casos, currículos, formadores ou recursos específicos para esse fim. A literacia midiática não está integrada de forma sistemática ao currículo da educação básica

brasileira, não é objeto de formação continuada de professores em escala nacional e não tem programa governamental com recursos e metas definidos comparável ao que a Finlândia desenvolveu ao longo de três décadas. Essa ausência institucional é o obstáculo mais difícil de superar no curto prazo e o que mais claramente distingue a situação brasileira da finlandesa: não é falta de evidência sobre a importância da literacia midiática, mas falta de estrutura institucional para desenvolvê-la em escala.

O quadro a seguir sistematiza o déficit de literacia midiática em quatro dimensões analiticamente distintas, permitindo uma leitura que o relato narrativo tende a unificar indevidamente. A distinção entre dimensões é operacionalmente relevante porque cada uma exige intervenções diferentes: a baixa capacidade de verificação de fontes responde a treinamentos sobre técnicas de checagem; a incapacidade de identificar deepfakes responde a exposição supervisionada a exemplos e a treinamento perceptivo específico; a ausência de compreensão sobre algoritmos responde a educação sobre o funcionamento das plataformas; mas o déficit de ceticismo adequado, a dimensão mais complexa da tabela, não responde a nenhuma dessas intervenções isoladamente porque é um problema de calibração, não de ausência de competência. A linha do ceticismo adequado merece leitura especialmente cuidadosa porque introduz uma assimetria que a narrativa dominante sobre literacia midiática frequentemente ignora: o excesso de ceticismo é tão prejudicial ao discernimento informacional quanto a credulidade excessiva. Um eleitor que nega a autenticidade de registros verdadeiros por desconfiança generalizada das instituições está tão vulnerável à desinformação quanto o que aceita deepfakes sem questionamento, só que pela via oposta. O dividendo do mentiroso, o benefício que atores desonestos extraem da desconfiança generalizada que suas próprias mentiras ajudaram a criar, é o fenômeno que essa linha captura e que qualquer programa de literacia midiática precisa endereçar explicitamente para não produzir o efeito indesejado de aumentar o ceticismo indiscriminado em vez de desenvolvê-lo de forma calibrada.

Tabela 55. Déficit de Literacia Midiática no Brasil: Dimensões e Implicações Eleitorais

Dimensão	Caracterização do Déficit	Grupo Mais Vulnerável	Implicação Eleitoral
Verificação de fontes	Baixa capacidade de identificar fontes confiáveis e distingui-las	Eleitores com escolaridade fundamental incompleta; usuários de	Alta receptividade a desinformação distribuída por grupos de WhatsApp com aparência de notícia.

	de fontes não verificadas.	WhatsApp como fonte primária.	
Identificação de deepfakes	Incapacidade de identificar vídeos e áudios sintéticos sem treinamento específico.	Eleitores com mais de 60 anos; populações com menor exposição a conteúdo digital sofisticado.	Vulnerabilidade a deepfakes de candidatos distribuídos em período pré-eleitoral.
Compreensão de algoritmos	Ausência de compreensão sobre como algoritmos de recomendação personalizam o feed de notícias.	Usuários intensivos de redes sociais sem reflexão sobre o funcionamento das plataformas.	Susceptibilidade a bolhas de filtro e câmaras de eco sem percepção de que estão ocorrendo.
Ceticismo adequado	Tanto excesso de credulidade quanto excesso de ceticismo prejudicam o discernimento informacional.	Extremos: eleitores muito crédulos (vulneráveis à desinformação) e muito céticos (vulneráveis ao dividendo do mentiroso).	Eleitores muito crédulos acreditam em deepfakes; muito céticos negam autenticidade de registros verdadeiros.

Fonte: Elaborado pelo autor com base em Reuters Institute Digital News Report (2023) e GPOPAI/USP (2018).

Competências de Literacia Midiática no Contexto da IA Eleitoral

A literacia midiática para o contexto específico da IA eleitoral requer um conjunto de competências que vai além das habilidades tradicionais de avaliação crítica de fontes jornalísticas. O eleitor que aprende a verificar se um site de notícias tem endereço físico registrado e corpo editorial identificável está adquirindo competência necessária mas insuficiente para navegar num ecossistema informacional em que o conteúdo mais perigoso frequentemente não vem de sites de notícias mas de vídeos sintéticos enviados por pessoas de confiança, de áudios que imitam a voz de candidatos com precisão suficiente para enganar familiares próximos e de chatbots que simulam conversas com representantes de campanhas sem qualquer indicação de que são sistemas automatizados.

A distinção entre literacia midiática tradicional e literacia midiática para o contexto de IA eleitoral não é uma questão de grau mas de natureza. A primeira foi desenvolvida num ambiente em que o conteúdo desinformativo era produzido por humanos com capacidade de escala limitada e deixava marcadores identificáveis de sua origem, como redação descuidada, ausência de corroboração por fontes independentes e padrões visuais reconhecíveis de produção amadora. A segunda precisa operar num ambiente em que o conteúdo desinformativo é produzido por sistemas que não erram por descuido, que podem

ser corroborados por outros conteúdos sintéticos e que atingem níveis de qualidade visual e auditiva indistinguíveis, para o olho e o ouvido não treinados, de registros autênticos. Os marcadores que a literacia midiática tradicional treinava o eleitor a identificar foram progressivamente eliminados pela evolução tecnológica dos sistemas de geração de conteúdo.

As competências específicas que o eleitor do século XXI precisa desenvolver para navegar criticamente nesse ambiente podem ser organizadas em quatro categorias que correspondem às principais formas de conteúdo sintético que circulam no ecossistema eleitoral digital.

A primeira categoria é a competência de identificação de conteúdo sintético audiovisual, que inclui o conhecimento dos principais artefatos visuais e auditivos que sistemas de geração de deepfakes ainda produzem, a capacidade de aplicar técnicas básicas de verificação como a análise de movimentos oculares e taxa de piscamento, e a compreensão de que a ausência de artefatos detectáveis não é prova de autenticidade para conteúdo produzido por modelos avançados. Como demonstrado no Capítulo 14, essa competência tem limitações estruturais que os sistemas de detecção profissional também enfrentam: não se trata de desenvolver no eleitor a capacidade de detectar deepfakes com certeza, mas de desenvolver o hábito de questionar antes de aceitar e de buscar verificação independente antes de redistribuir.

A segunda categoria é a competência de compreensão de algoritmos de recomendação, que inclui a compreensão básica de como plataformas selecionam e priorizam o conteúdo exibido para cada usuário, o reconhecimento dos sinais de que o conteúdo está sendo apresentado por sua capacidade de gerar engajamento emocional e não por sua relevância ou precisão factual, e a capacidade de deliberadamente buscar perspectivas que o algoritmo não apresentaria espontaneamente. Pariser (2011), em *The Filter Bubble*, foi o primeiro a documentar sistematicamente o fenômeno de personalização algorítmica do ecossistema informacional e suas consequências para a formação de opiniões políticas em ambiente democrático. A atualização dessa análise para o contexto de IA generativa, em que os algoritmos de recomendação interagem com conteúdo produzido por outros sistemas de IA numa cadeia de amplificação totalmente automatizada, é o desafio conceitual mais recente da literacia midiática para o contexto eleitoral (PARISER, 2011, p. 9-15).

A terceira categoria é a competência de verificação de informações em contexto de urgência, que é a mais praticamente relevante e a mais difícil de desenvolver. A

desinformação eleitoral opera estrategicamente no tempo: ela é projetada para circular no período em que o eleitor tem menos tempo e menos disposição para verificar, nos dias imediatamente anteriores ao pleito e nos momentos de maior agitação emocional do ciclo eleitoral. Desenvolver a competência de verificação em contexto de urgência significa internalizar rotinas de pausa antes do compartilhamento que sejam suficientemente automáticas para serem aplicadas mesmo sob pressão temporal e emocional. Clayton et al. (2020) demonstraram experimentalmente que intervenções que induzem os usuários a pausar e considerar a acurácia do conteúdo antes de compartilhá-lo produzem reduções mensuráveis na disseminação de desinformação sem reduzir o compartilhamento de conteúdo preciso (CLAYTON et al., 2020, p. 2039-2042). O desafio pedagógico é transformar essa pausa de comportamento induzido por intervenção experimental em hábito cognitivo estável que opere sem necessidade de gatilho externo.

A quarta categoria é a competência de compreensão dos modelos de negócio das plataformas e suas implicações para a qualidade do ecossistema informacional. Como demonstrado no Capítulo 8 desta obra, as plataformas têm incentivo estrutural para amplificar conteúdo emocionalmente ativador independentemente de sua autenticidade, porque o engajamento emocional é o mecanismo que mantém os usuários na plataforma e que valoriza os dados comportamentais que constituem o produto principal do modelo de negócio. Um eleitor que compreende essa lógica estrutural não está apenas mais bem equipado para avaliar o conteúdo que recebe nas plataformas: está em posição de questionar por que determinado conteúdo está sendo amplificado e quais incentivos governam essa amplificação, o que é uma forma de metacognição sobre o ecossistema informacional que Zuboff (2019) identifica como condição necessária para a resistência ao capitalismo de vigilância em sua dimensão política.

O desenvolvimento dessas quatro categorias de competência não é tarefa que se conclui com uma campanha de conscientização ou com um módulo de treinamento pontual. É processo contínuo que exige exposição repetida, prática em contextos variados e atualização constante à medida que as tecnologias de geração e distribuição de desinformação evoluem. Essa característica processual da literacia midiática é o argumento mais forte disponível para a integração curricular sistemática em vez de intervenções pontuais: campanhas de conscientização pré-eleitorais podem sensibilizar, mas apenas a educação continuada desenvolve as competências cognitivas que resistem às formas de desinformação ainda não existentes quando o treinamento ocorreu.

Experiências Internacionais de Referência

A análise de experiências internacionais bem-sucedidas de educação para a literacia midiática oferece ao Brasil não apenas modelos a serem considerados, mas evidência de que o problema é solucionável com escolhas políticas específicas e investimento sustentado ao longo do tempo. A relevância comparada dessas experiências não está na possibilidade de transplante direto de modelos desenvolvidos em contextos institucionais, culturais e tecnológicos distintos do brasileiro: está na identificação dos princípios que as tornaram eficazes e que podem ser adaptados às condições específicas do sistema educacional e do ecossistema informacional brasileiro.

A experiência finlandesa é a mais documentada e a mais frequentemente citada na literatura comparada, e sua análise cuidadosa revela que seu sucesso não repousa em nenhum elemento singular mas na combinação de três características estruturais que se reforçam mutuamente. A primeira é a integração curricular transversal: a Finlândia não criou uma disciplina de literacia midiática, o que tenderia a marginalizar o tema como matéria adicional a ser cumprida formalmente, mas integrou as competências de literacia midiática ao ensino de língua materna, história, estudos sociais e ciências como ferramenta de cada disciplina. Essa integração garante que as competências sejam desenvolvidas em contextos de uso real, não em exercícios abstratos desconectados das práticas informacionais cotidianas dos estudantes. A segunda característica é a progressão de complexidade ao longo dos anos escolares, que começa na identificação de fontes confiáveis no ensino primário e avança para a análise de técnicas de propaganda, viés algorítmico e desinformação coordenada no ensino médio. Essa progressão garante que as competências se desenvolvam em ordem de complexidade crescente, com cada etapa construindo sobre a anterior. A terceira é o investimento contínuo na formação de professores, que são os implementadores do currículo e sem cuja capacitação específica qualquer programa de literacia midiática permanece como intenção curricular não realizada na sala de aula.

A experiência estoniana é analiticamente relevante por uma razão distinta: ela demonstra que investimento em literacia midiática pode ser enquadrado como política de segurança nacional sem que esse enquadramento comprometa sua dimensão democrática. A Estônia, que sofreu ataques de desinformação coordenados relacionados à sua transição pós-soviética e às suas relações com a Rússia, desenvolveu um programa nacional de educação digital e literacia midiática que reconhece explicitamente a desinformação como ameaça à segurança democrática do Estado. Esse enquadramento

produziu consequências institucionais relevantes: alocação de recursos de defesa para programas educacionais, envolvimento de agências de segurança na identificação de padrões de desinformação que informam o currículo de literacia midiática e parceria com empresas de tecnologia para desenvolvimento de ferramentas educacionais acessíveis a populações com diferentes níveis de capacidade técnica. A lição estoniana para o Brasil não é que o TSE deva ser militarizado, mas que o enquadramento da literacia midiática como questão de segurança democrática pode mobilizar recursos e atenção institucional que o enquadramento puramente educacional raramente consegue.

Taiwan implementou, após ataques de desinformação documentados associados a eleições e a tensões com a China continental, um programa de educação midiática que se distingue das experiências finlandesa e estoniana por sua velocidade de implementação e por sua ênfase na parceria entre governo, sociedade civil e plataformas digitais. O modelo taiwanês combina três componentes que operam em paralelo: currículo formal nas escolas, com integração da literacia midiática ao ensino de cívica e estudos sociais; campanhas de comunicação pública que traduzem conceitos de literacia midiática para linguagem acessível ao eleitorado geral; e parceria ativa com plataformas de fact-checking que produzem verificações rápidas de conteúdo viral, disponibilizadas em formato compatível com os canais pelos quais a desinformação circula. O elemento mais notável da experiência taiwanesa é o papel do governo como curador ativo de contra-narrativas: em vez de apenas financiar educação midiática passiva, o governo taiwanês desenvolveu capacidade institucional de resposta rápida a campanhas de desinformação específicas, publicando verificações em tempo suficientemente curto para alcançar o mesmo ciclo de atenção que o conteúdo desinformativo ocupa (TANG, 2020, p. 44-52).

A análise comparada das três experiências permite extrair cinco princípios que transcendem os contextos específicos em que foram desenvolvidos e que podem orientar a construção de uma política nacional de literacia midiática brasileira com resultados mensuráveis antes e durante o ciclo eleitoral de 2026. O primeiro é que integração curricular é mais eficaz que disciplina isolada. O segundo é que formação de professores é condição necessária, não opcional, para qualquer programa que pretenda ter escala. O terceiro é que progressão de complexidade garantida ao longo dos anos escolares produz competências mais robustas do que intervenções pontuais de qualquer intensidade. O quarto é que a parceria com plataformas e com organizações de fact-checking multiplica o alcance de programas com recursos limitados. E o quinto é que enquadramento institucional adequado,

seja como política educacional, de segurança ou democrática, determina os recursos e a atenção que o tema recebe nos ciclos orçamentários que decidem sua implementação real.

O Papel do TSE, dos Partidos e da Sociedade Civil

A educação para a literacia midiática no contexto eleitoral não é responsabilidade exclusiva do sistema educacional formal, e essa distribuição de responsabilidades não é apenas uma questão de divisão de tarefas: é uma questão de alcance. O sistema educacional formal chega a crianças e jovens em processo de formação, mas o eleitorado de 2026 inclui dezenas de milhões de adultos que concluíram sua formação escolar antes de qualquer programa sistemático de literacia midiática e que são precisamente os mais vulneráveis aos padrões de desinformação documentados nos ciclos eleitorais anteriores. Alcançar esse eleitorado adulto exige instrumentos que o sistema educacional formal não tem, e é nesse espaço que o TSE, os partidos políticos e a sociedade civil têm papel insubstituível.

O TSE tem investido progressivamente em programas de educação para a democracia digital que representam o avanço mais concreto disponível no curto prazo para o eleitorado adulto. O Programa de Enfrentamento à Desinformação com Foco nas Eleições, as parcerias com o Ministério da Educação para inclusão do tema em materiais curriculares e o canal Fato ou Boato, que permite ao eleitor verificar conteúdo suspeito diretamente com as agências de fact-checking parceiras, constituem uma arquitetura de literacia midiática aplicada com alcance direto junto ao eleitorado. A limitação dessa arquitetura é a mesma que afeta toda comunicação institucional eleitoral: ela alcança prioritariamente eleitores que já estão engajados com a Justiça Eleitoral como fonte de informação, que são frequentemente os menos vulneráveis à desinformação. O desafio do TSE, que a literatura de comunicação para mudança de comportamento identifica como um dos mais difíceis disponíveis, é alcançar eleitores que não buscam o TSE como fonte e que consomem informação política exclusivamente por canais informais.

Os partidos políticos têm responsabilidade específica nessa agenda que deriva diretamente de sua função constitucional de intermediação entre cidadãos e Estado. Por seu papel de formação política e de mobilização de militantes e apoiadores, os partidos têm acesso a redes de comunicação que o TSE e as organizações de sociedade civil raramente alcançam: grupos de WhatsApp de militância local, reuniões de diretório municipal, eventos de campanha e redes de lideranças comunitárias com raízes nos segmentos do eleitorado

mais vulneráveis à desinformação. Partidos que treinam seus militantes e apoiadores para verificar conteúdo antes de compartilhá-lo não estão apenas cumprindo obrigação ética: estão exercendo influência sobre a camada de amplificação orgânica que é o principal vetor de disseminação de desinformação eleitoral no ecossistema do WhatsApp.

Há, contudo, uma tensão estrutural que o papel dos partidos na literacia midiática precisa reconhecer explicitamente: partidos que instrumentalizam a desinformação para fins eleitorais, como documentado nos ciclos de 2018 e 2022, não podem simultaneamente ser promotores credíveis de literacia midiática junto a seus eleitores. A credibilidade de qualquer iniciativa de literacia midiática partidária depende da coerência entre o que o partido ensina e o que pratica, e eleitores com capacidade crítica suficiente para beneficiar-se dessas iniciativas têm também capacidade suficiente para identificar essa incoerência quando ela existe. Partidos que querem ser atores relevantes na agenda de literacia midiática precisam começar pela conformidade regulatória com a Resolução TSE nº 23.732/2024 e pela renúncia às práticas de desinformação que comprometem a legitimidade de qualquer iniciativa educacional subsequente.

A sociedade civil organizada, incluindo organizações de fact-checking como Agência Lupa, Aos Fatos e Comprova, institutos de pesquisa como o NetLab UFRJ e o FGV DAPP e organizações de direitos digitais como o Instituto Brasileiro de Defesa do Consumidor e o Coletivo Intervozes, constitui o elo mais ágil e mais especializado da cadeia de literacia midiática eleitoral. Sua agilidade deriva de não estar sujeita aos ciclos orçamentários e aos processos decisórios que tornam as respostas institucionais do governo e do TSE mais lentas do que a velocidade da desinformação exige. Sua especialização deriva de anos de prática em verificação de conteúdo e em comunicação de resultados para audiências não especializadas. Sua limitação é o alcance: as organizações de fact-checking brasileiras, apesar de seu crescimento no período pós-2018, ainda alcançam uma fração minoritária do eleitorado, frequentemente aquela com maior literacia midiática prévia. Expandir esse alcance para os segmentos do eleitorado mais vulneráveis exige financiamento sustentável, distribuição por canais não convencionais e parcerias com as redes de comunicação informal que o sistema de fact-checking formal não alcança organicamente.

■ **RESUMO DO CAPÍTULO:**

Este capítulo analisou a literacia midiática como condição estrutural da resiliência democrática na era da inteligência artificial, argumentando que nenhuma regulação ou ferramenta de detecção consegue substituir o que a capacidade crítica desenvolvida em escala populacional pode produzir: a redução da camada de amplificação orgânica que é o principal vetor de disseminação de desinformação eleitoral e que

a sanção jurídica não alcança. A análise teórica estabeleceu os fundamentos a partir de três autores centrais. Hobbs (2010) situou a literacia midiática como competência democrática que vai além da capacidade técnica de operar dispositivos, abrangendo as competências críticas necessárias para participar de forma autônoma no ecossistema informacional digital. Potter (2016) descreveu o caráter processual e contínuo da literacia midiática, demonstrando que é um continuum de competências, não uma capacidade binária, e que a diferença entre indivíduos com maior e menor literacia é a velocidade de processamento crítico. Pennycook e Rand (2019) forneceram a evidência experimental mais diretamente aplicável ao contexto eleitoral, demonstrando que intervenções que reativam a atenção à acurácia produzem reduções mensuráveis na disseminação de desinformação sem reduzir o compartilhamento de conteúdo preciso. O diagnóstico do déficit brasileiro identificou sua distribuição assimétrica entre regiões, faixas etárias e estratos de escolaridade, com três dimensões específicas: o dado de autopercepção do Reuters Institute, qualificado pelo third-person effect que indica que o percentual real de vulnerabilidade é superior ao que a autopercepção captura; a dimensão geracional específica de adultos que chegaram ao digital pelo WhatsApp com perfil de vulnerabilidade que intervenções projetadas para jovens digitalmente nativos não alcançam; e a dimensão institucional como o obstáculo mais difícil de superar no curto prazo, distinguindo claramente o Brasil da Finlândia não pela falta de evidência sobre a importância da literacia midiática mas pela falta de estrutura institucional para desenvolvê-la em escala. As competências específicas para o contexto de IA eleitoral foram organizadas em quatro categorias analiticamente distintas, cada uma com suas limitações e intervenções adequadas: identificação de conteúdo sintético audiovisual, compreensão de algoritmos de recomendação, verificação de informações em contexto de urgência e compreensão dos modelos de negócio das plataformas como metacognição sobre o ecossistema informacional. A distinção entre literacia midiática tradicional e literacia para IA eleitoral foi estabelecida como diferença de natureza, não de grau, porque os marcadores que a literacia tradicional treinava o eleitor a identificar foram progressivamente eliminados pela evolução tecnológica dos sistemas de geração de conteúdo. A análise comparada das experiências finlandesa, estoniana e taiwanesa extraiu cinco princípios transferíveis: integração curricular transversal é mais eficaz que disciplina isolada; formação de professores é condição necessária, não opcional; progressão de complexidade ao longo dos anos escolares produz competências mais robustas que intervenções pontuais; parceria com plataformas e fact-checkers multiplica o alcance com recursos limitados; e enquadramento institucional adequado determina os recursos que o tema recebe nos ciclos orçamentários decisivos. O papel do TSE, dos partidos e da sociedade civil foi analisado com suas contribuições específicas e suas limitações estruturais. O TSE alcança prioritariamente eleitores já engajados com a Justiça Eleitoral, que são frequentemente os menos vulneráveis. Os partidos têm acesso a redes informais que o TSE não alcança, mas enfrentam a tensão estrutural entre ensinar e praticar literacia midiática, que compromete a credibilidade de iniciativas educacionais quando há incoerência entre discurso e conduta. A sociedade civil tem agilidade e especialização que as instituições formais não conseguem replicar, mas alcance limitado aos segmentos do eleitorado com maior literacia midiática prévia.

QUESTÕES PARA REFLEXÃO E DEBATE

1. O capítulo argumenta que a literacia midiática é o único instrumento que protege o eleitorado onde a regulação não chega, especificamente na camada de amplificação orgânica composta por cidadãos que redistribuem desinformação de boa-fé. Essa afirmação pressupõe que literacia midiática e regulação são instrumentos complementares e não substitutos. Em que circunstâncias concretas eles poderiam entrar em conflito, por exemplo quando a regulação restringe conteúdo que cidadãos com maior literacia midiática seriam capazes de avaliar criticamente por conta própria?

2. A evidência de Pennycook e Rand demonstra que intervenções que reativam a atenção à acurácia reduzem a disseminação de desinformação. Essa evidência foi produzida em contextos experimentais controlados. Que desafios de validade externa, de generalização dos resultados experimentais para o

ambiente eleitoral real, precisam ser endereçados antes que políticas públicas de escala sejam desenhadas exclusivamente com base nessa evidência?

3. O diagnóstico do déficit brasileiro identificou que adultos que chegaram ao digital pelo WhatsApp têm perfil de vulnerabilidade específico que intervenções para jovens digitalmente nativos não alcançam. Que formatos, canais e linguagens de programas de literacia midiática seriam mais eficazes para esse segmento específico do eleitorado, e quais instituições, igrejas, sindicatos, associações comunitárias, têm o acesso e a credibilidade para entregá-los?

4. A análise da experiência estoniana demonstrou que enquadrar a literacia midiática como política de segurança nacional mobilizou recursos que o enquadramento puramente educacional raramente consegue. Esse enquadramento tem riscos específicos para o contexto democrático: pode ser instrumentalizado para justificar restrições à liberdade de expressão sob o argumento de combate à desinformação. Como o Brasil deveria navegar essa tensão entre eficácia do enquadramento de segurança e preservação das garantias democráticas?

5. O capítulo identificou uma tensão estrutural no papel dos partidos na literacia midiática: partidos que instrumentalizam desinformação não podem ser promotores credíveis de educação midiática junto a seus eleitores. Que mecanismo de accountability, legal, regulatório ou reputacional, seria mais eficaz para criar incentivo para que partidos renunciem às práticas de desinformação como condição de legitimidade para atuar na agenda de literacia midiática?

6. As organizações de fact-checking brasileiras, apesar do crescimento pós-2018, ainda alcançam uma fração minoritária do eleitorado, frequentemente a de maior literacia midiática prévia. Que modelo de distribuição de verificações de fact-checking, incluindo parceria com WhatsApp, integração com sistemas de recomendação das plataformas ou distribuição por redes religiosas e comunitárias, teria maior potencial de alcançar os segmentos do eleitorado mais vulneráveis?

7. A linha do ceticismo adequado no Quadro 16.1 identificou que excesso de ceticismo é tão prejudicial quanto credulidade excessiva, pelo mecanismo do dividendo do mentiroso. Programas de literacia midiática que desenvolvem ceticismo sem calibrá-lo adequadamente podem inadvertidamente aumentar a vulnerabilidade dos participantes à estratégia de negação sistemática de toda evidência. Como o design pedagógico de programas de literacia midiática deveria endereçar especificamente esse risco?

8. O capítulo argumenta que os retornos do investimento em literacia midiática acumulam-se ao longo de décadas, com horizonte de avaliação de gerações, não de ciclos eleitorais. Essa característica temporal torna o investimento em literacia midiática estruturalmente desfavorecido em processos orçamentários que operam em horizontes anuais e em processos políticos que operam em horizontes eleitorais de quatro anos. Que mecanismos institucionais, fundos de longo prazo, metas constitucionais ou acordos interpartidários, poderiam proteger o investimento em literacia midiática dos ciclos de curto prazo que sistematicamente subinvestem em políticas de retorno lento?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. A literacia midiática do seu eleitorado é ativo estratégico da sua campanha. Invista nela.

Eleitores com maior capacidade crítica para identificar desinformação são mais difíceis de ser mobilizados contra o seu candidato por operações de fake news e deepfakes. Isso não é argumento altruísta: é argumento estratégico. Campanhas que distribuem conteúdo educativo sobre como identificar deepfakes, como verificar áudios políticos e como questionar mensagens de WhatsApp antes de repassá-las estão construindo imunidade eleitoral no seu eleitorado. Esse investimento produz retorno não apenas para a eleição em curso: eleitores mais críticos são eleitores mais fiéis a quem os tratou com respeito e os equipou para navegar um ambiente informacional hostil.

2. Treinar sua militância para verificar antes de compartilhar é mais eficaz do que pedir para não compartilhar desinformação.

A camada de amplificação orgânica da desinformação é composta por pessoas que redistribuem conteúdo de boa-fé, acreditando genuinamente no que compartilham. Essa camada não pode ser alcançada por instrução de não compartilhar: ela só é reduzida por desenvolvimento de capacidade crítica que atua antes do ato de compartilhamento. Partidos e campanhas que treinam sua militância e apoiadores com ferramentas práticas de verificação, como busca reversa de imagens, uso do Fato ou Boato do TSE e análise de metadados básica, estão intervindo no único ponto onde a desinformação pode ser interrompida antes de causar dano: na decisão individual de redistribuir ou não.

3. O excesso de ceticismo é tão perigoso quanto a credulidade. Cuide da calibração, não apenas do volume.

Um eleitorado treinado para desconfiar de tudo é tão vulnerável à desinformação quanto um eleitorado que aceita tudo sem questionar: ele nega a autenticidade de registros verdadeiros, o que é exatamente o que produtores de desinformação sofisticados querem quando constroem narrativas de que toda evidência contra eles é fabricada. O dividendo do mentiroso, o benefício que atores desonestos extraem da desconfiança generalizada que suas próprias mentiras ajudaram a criar, é real e documentado. Programas de literacia midiática que sua campanha promova precisam calibrar o ceticismo, ensinando quando e como questionar, não apenas que se deve questionar.

4. O TSE tem ferramentas de verificação gratuitas que sua militância provavelmente não conhece.

O canal Fato ou Boato do TSE permite ao eleitor verificar conteúdo suspeito diretamente com as agências de fact-checking parceiras da Justiça Eleitoral, de forma gratuita e acessível por WhatsApp. As agências Lupa, Aos Fatos e Comprova têm tiplines de verificação que funcionam no mesmo formato. Distribuir esses recursos para redes de militância, grupos de apoiadores e comunidades locais é uma das intervenções de literacia midiática com maior custo-benefício disponíveis: não requer produção de conteúdo próprio, não gera risco jurídico e produz impacto imediato sobre a capacidade de verificação dos grupos que mais precisam dela.

5. Se você é partido, não pode ensinar literacia midiática e praticar desinformação. A incoerência é detectável.

A credibilidade de qualquer iniciativa de literacia midiática partidária depende da coerência entre o que o partido ensina e o que pratica. Eleitores com capacidade crítica suficiente para beneficiar-se de treinamentos de verificação têm também capacidade suficiente para identificar quando o partido que os ensina a verificar conteúdo produz ou amplifica desinformação em suas próprias comunicações. A conformidade com a Resolução TSE 23.732/2024 e a renúncia às práticas de desinformação documentadas não são apenas obrigações jurídicas: são pré-condições de credibilidade para qualquer papel na agenda de literacia midiática eleitoral.

6. Adultos acima de 50 anos que usam WhatsApp como fonte principal de informação são o grupo mais vulnerável. E os mais difíceis de alcançar pelos canais convencionais.

O perfil demográfico mais vulnerável à desinformação eleitoral no Brasil não é o jovem sem educação formal: é o adulto maduro com redes relacionais extensas, credibilidade nas suas comunidades e dependência do WhatsApp como fonte de informação política, que chegou ao ambiente digital sem o treinamento crítico que usuários mais jovens desenvolveram por imersão. Esse perfil concentra-se em faixas etárias acima de 50 anos, em regiões com menor acesso ao ensino superior e em comunidades religiosas e profissionais com comunicação interna intensa. Alcançá-lo exige canais que não são os convencionais de comunicação política: lideranças religiosas, associações de bairro, grupos profissionais e redes de aposentados que têm credibilidade nesses ambientes e que podem ser parceiros de programas de literacia midiática com muito mais eficácia do que materiais institucionais distribuídos por canais formais.

7. Literacia midiática não é gasto de campanha: é investimento em capital cívico que dura décadas.

Os investimentos em educação midiática que uma campanha ou partido realiza não são despesas eleitorais de curto prazo que se esgotam no dia do pleito: são investimentos em capital cívico que produzirão dividendos democráticos por décadas, independentemente de quem vencer as eleições em que forem aplicados. Um eleitor que aprendeu a verificar conteúdo antes de compartilhar continuará fazendo isso nos ciclos eleitorais seguintes. Uma comunidade que desenvolveu práticas de verificação coletiva tornou-se estruturalmente mais resistente à desinformação. E uma campanha que tratou seus eleitores com o respeito de equipá-los para navegar criticamente o ambiente informacional construiu um ativo relacional que nenhum deepfake de adversário consegue destruir rapidamente.

CAPÍTULO 17

Inteligência Artificial de Próxima Geração

Tendências Tecnológicas e seus Impactos sobre o Processo Eleitoral

Prospectiva Tecnológica e Democracia

A prospectiva tecnológica aplicada ao campo eleitoral é um exercício metodologicamente delicado que a maioria dos analistas evita por razões compreensíveis: o histórico de previsões tecnológicas é suficientemente embaraçoso para dissuadir qualquer estudioso que preze sua credibilidade acadêmica. O chamado ciclo de expectativas de Gartner, que descreve a trajetória típica de tecnologias emergentes do pico de expectativas infladas pelo entusiasmo inicial ao vale do desencantamento que precede a adoção estável em escala produtiva, documenta sistematicamente que previsões tecnológicas tendem a ser excessivamente otimistas no curto prazo e excessivamente pessimistas no longo prazo.

No contexto específico da inteligência artificial generativa, porém, o padrão histórico de subavaliação do impacto é ainda mais pronunciado do que em outras tecnologias. O lançamento do ChatGPT em novembro de 2022 não apenas confirmou as capacidades que a literatura técnica já havia documentado: ele as tornou acessíveis a centenas de milhões de usuários num intervalo de semanas que não tem precedente na história da difusão tecnológica. Nenhuma projeção publicada antes de outubro de 2022, incluindo as mais otimistas, havia antecipado esse ritmo. E desde então, cada ciclo de lançamentos de novos modelos tem superado sistematicamente as capacidades que o ciclo anterior tornava plausíveis. O analista que projeta o futuro da IA eleitoral a partir das capacidades disponíveis hoje está cometendo o mesmo erro metodológico que um analista de 2020 cometeria ao projetar 2024 a partir das capacidades do GPT-3.

Russell e Norvig (2022), na quarta edição de *Artificial Intelligence: A Modern Approach*, a obra de referência mais citada da área, oferecem o enquadramento mais rigoroso disponível para a análise das trajetórias de desenvolvimento. Sua contribuição mais relevante para a análise eleitoral não é a descrição técnica dos sistemas existentes, mas a identificação dos princípios que governam a progressão de capacidades: sistemas de IA melhoram não linearmente quando escalados em dados e computação, produzindo saltos de capacidade que frequentemente surpreendem os próprios pesquisadores que os desenvolvem. Essa não-linearidade é o mecanismo que torna a prospectiva baseada em

extrapolação linear sistematicamente inadequada para o campo (RUSSELL; NORVIG, 2022, p. 1-28).

Brynjolfsson e McAfee (2014), em *The Second Machine Age*, articularam o mecanismo econômico pelo qual tecnologias exponenciais produzem impactos não lineares que surpreendem analistas e reguladores acostumados a progressões lineares. O argumento central, que permanece analiticamente poderoso uma década depois, é que tecnologias digitais seguem curvas de melhoria que dobram regularmente em capacidade por unidade de custo, produzindo numa fase inicial melhorias que parecem modestas e numa fase posterior melhorias que parecem impossíveis. A IA generativa atingiu, entre 2020 e 2024, a inflexão da curva exponencial em que o progresso começa a parecer impossível para quem o observa de fora. O que vem a seguir não é previsível em seus detalhes específicos, mas é estruturalmente certo em sua direção: mais capacidade, menor custo, maior acessibilidade (BRYNJOLFSSON; MCAFEE, 2014, p. 8-15).

Dafoe (2020), em *AI Governance: A Research Agenda*, sistematizou as perguntas fundamentais que o desenvolvimento da IA coloca para a governança democrática, identificando um conjunto de tensões estruturais que nenhum marco regulatório atual, incluindo os mais avançados analisados nos capítulos anteriores desta obra, consegue resolver de forma definitiva. A mais relevante para o contexto eleitoral é a tensão entre a velocidade de inovação tecnológica e a velocidade de adaptação institucional: sistemas democráticos foram projetados para produzir estabilidade e previsibilidade, qualidades que entram em conflito com a adaptação à velocidade que o ecossistema de IA exige. A pergunta que Dafoe coloca, e que este capítulo procura endereçar, é se os instrumentos institucionais disponíveis para a democracia são suficientemente ágeis para governar tecnologias que evoluem mais rapidamente do que qualquer processo legislativo ou regulatório consegue acompanhar (DAFOE, 2020, p. 1-18).

Para os propósitos deste capítulo, a prospectiva tecnológica não é tratada como futurologia especulativa, mas como análise de trajetórias em curso com grau identificável de confiança. Existem tecnologias que já existem em estágio inicial de desenvolvimento cujo aprimoramento progressivo é altamente provável, cujas trajetórias técnicas são compreendidas com razoável precisão e cujas implicações eleitorais podem ser antecipadas com fundamento empírico suficiente para orientar preparação regulatória e institucional. A distinção entre o que é altamente provável no horizonte de dois a cinco anos, o que é plausível no horizonte de cinco a dez anos e o que é especulativo além desse horizonte é o que este capítulo procura manter com consistência metodológica ao longo de toda a análise.

IA Generativa de Próxima Geração: O que Vem Depois do ChatGPT

Os modelos de linguagem de grande escala que dominam o ecossistema de IA generativa em 2026 representam a terceira geração de sistemas com capacidades emergentes genuínas. A primeira geração, representada pelo GPT-2 e por seus contemporâneos, demonstrou que modelos de linguagem podiam gerar texto coerente mas frequentemente incoerente factualmente. A segunda geração, representada pelo GPT-3 e pelos modelos de linguagem de grande escala da mesma era, demonstrou que escala era o mecanismo principal de melhoria de capacidades, produzindo sistemas capazes de realizar tarefas que não estavam nos dados de treinamento. A terceira geração, representada pelos modelos disponíveis em 2026, integrou capacidades multimodais, raciocínio estendido e alinhamento com preferências humanas de formas que tornaram a interação com esses sistemas acessível a usuários sem qualquer formação técnica.

A quarta geração, cujos primeiros exemplares estão disponíveis em versões de pesquisa e em implementações comerciais restritas em 2026, apresenta três avanços que têm implicações eleitorais diretas e que os ciclos regulatórios disponíveis não foram projetados para endereçar.

O primeiro avanço é a expansão radical das janelas de contexto, que permitem ao modelo processar e raciocinar sobre volumes de informação que modelos anteriores não conseguiam manter coerentemente em memória operacional. Modelos da quarta geração são capazes de processar simultaneamente documentos com dezenas de milhões de tokens, o equivalente a bibliotecas inteiras de textos, mantendo coerência de raciocínio ao longo de toda a extensão.

Para o contexto eleitoral, essa capacidade viabiliza análises estratégicas de complexidade e escala antes impossíveis: um modelo capaz de processar simultaneamente toda a base de dados eleitoral do TSE desde 1994, o arquivo histórico completo de um candidato incluindo declarações, votações e posicionamentos ao longo de décadas, e o ecossistema informacional em tempo real de uma campanha, gerando análises estratégicas instantâneas e identificando padrões que analistas humanos levam meses para descobrir. A vantagem competitiva que esse tipo de análise confere é da mesma ordem de grandeza que a vantagem que sistemas de alta frequência conferiram a traders financeiros sobre investidores que operavam manualmente: ela não é apenas uma diferença de velocidade, é uma diferença de universo de padrões detectáveis.

O segundo avanço é a integração genuinamente multimodal em um único modelo, em contraste com as arquiteturas de 2024 que combinavam modelos especializados por modalidade. Modelos multimodais de quarta geração processam e geram texto, imagem, áudio e vídeo com o mesmo sistema, o que significa que o pipeline de produção de conteúdo sintético que em 2024 exigia múltiplas ferramentas especializadas, conhecimento técnico intermediário e horas de processamento colapsa para uma interface de prompt único acessível a qualquer usuário com acesso à internet. A redução de custo e complexidade que esse colapso representa não é incremental: é a diferença entre deepfakes como arma de atores sofisticados com recursos significativos e deepfakes como commodity disponível a qualquer campanha eleitoral de qualquer porte em qualquer nível.

A Resolução TSE nº 23.732/2024 foi calibrada para um mundo em que a produção de conteúdo sintético de alta qualidade ainda exigia recursos não triviais. Esse mundo está desaparecendo.

O terceiro avanço, e o de maior implicação para a integridade eleitoral de longo prazo, é a autonomia operacional crescente dos sistemas de IA, manifestada no desenvolvimento de agentes de IA capazes de executar sequências complexas de tarefas sem intervenção humana contínua. Agentes de IA de quarta geração podem ser instruídos com um objetivo de alto nível, como maximizar o engajamento positivo com um candidato em determinado segmento demográfico, e executar autonomamente todas as etapas necessárias para atingir esse objetivo: pesquisa de informações, criação de conteúdo personalizado, identificação de canais de distribuição adequados, publicação, monitoramento de resultados e ajuste de estratégia com base em feedback. Esse tipo de sistema não viola necessariamente nenhuma vedação específica da Resolução TSE nº 23.732/2024, porque a Resolução regula o conteúdo produzido e não o processo pelo qual ele é produzido, mas cria uma camada de automação eleitoral que torna a atribuição de responsabilidade jurídica estruturalmente mais difícil: quando um agente autônomo decide, sem instrução humana específica, produzir um conteúdo que viola as vedações da Resolução, a cadeia de responsabilização entre a instrução original e a violação específica é suficientemente longa e opaca para que a defesa de desconhecimento seja mais plausível do que seria num sistema com supervisão humana contínua.

A combinação desses três avanços cria o cenário regulatório mais desafiador que o direito eleitoral brasileiro já enfrentou: modelos acessíveis a qualquer ator, produzindo conteúdo sintético de qualidade indistinguível de registros autênticos, distribuído por agentes autônomos em escala e velocidade que nenhum sistema de fiscalização manual consegue monitorar, com cadeias de responsabilização suficientemente opacas para que

a atribuição jurídica seja sistematicamente difícil. A Resolução TSE nº 23.732/2024 é um marco regulatório relevante e pioneiro para o mundo de 2024. Para o mundo de 2028, ela precisará de uma revisão substantiva que a maioria dos atores institucionais ainda não começou a planejar.

Agentes Autônomos de IA e Campanhas de Influência: O Horizonte de 2028

Se os modelos multimodais de quarta geração representam uma evolução das capacidades de conteúdo, os agentes autônomos de IA representam uma ruptura na natureza das operações de influência que precisam ser governadas. A distinção é mais do que técnica: é uma distinção sobre o que significa atribuir responsabilidade por conduta eleitoral irregular num mundo em que parte dessa conduta é decidida por sistemas algorítmicos sem instrução humana específica para cada ato.

Um agente de IA eleitoral autônomo de segunda geração, que já existe em protótipos de pesquisa em 2026 e cuja disponibilização comercial é esperada antes das eleições de 2030, tem capacidades que diferem qualitativamente das ferramentas de IA analisadas nos capítulos anteriores desta obra. Ele não produz conteúdo mediante instrução explícita de um operador humano que especifica o conteúdo desejado: ele recebe um objetivo estratégico e determina autonomamente quais conteúdos, em quais formatos, para quais públicos, em quais momentos e por quais canais produzirão os resultados mais próximos do objetivo instruído. Ele monitora os resultados de suas intervenções em tempo real, aprende com o que funciona e ajusta suas estratégias sem supervisão humana. Ele pode coordenar redes de personas sintéticas que simulam debate orgânico entre cidadãos reais, criando a aparência de formação espontânea de opinião que é na verdade o produto de uma operação de influência totalmente automatizada.

Esse tipo de sistema coloca questões jurídicas para as quais o ordenamento eleitoral vigente não tem resposta adequada. A primeira é a da atribuição de autoria: quando um agente autônomo decide, sem instrução humana específica, criar um conteúdo que viola as vedações da Resolução TSE nº 23.732/2024, quem é o autor da violação? O candidato que instruiu o agente com um objetivo estratégico legítimo? A empresa que desenvolveu o agente? O operador que o configurou? A ausência de uma doutrina de atribuição de responsabilidade para sistemas de IA autônomos em direito eleitoral não é uma lacuna teórica menor: é a vulnerabilidade estrutural que operadores sofisticados explorarão para criar negação plausível em escala industrial.

A segunda questão é a da detecção: agentes de IA autônomos que coordenam campanhas de influência por meio de redes de personas sintéticas produzem padrões de comportamento que se aproximam progressivamente do comportamento orgânico à medida que os sistemas aprendem quais padrões acionam os sistemas de detecção das plataformas. A corrida armamentista entre sistemas de geração de influência e sistemas de detecção de influência inautêntica está, no horizonte de 2028, chegando a um ponto em que a distinção técnica entre comportamento orgânico e comportamento sintético não será mais confiável como critério de intervenção regulatória. Como argumentam Chesney e Citron (2019), a erosão da confiança na autenticidade de conteúdo digital produz um efeito político que persiste mesmo quando os sistemas de detecção melhoram: o dividendo do mentiroso em escala sistêmica, em que a dúvida sobre a autenticidade de qualquer conteúdo beneficia estruturalmente os atores que produzem desinformação, porque contamina retroativamente a credibilidade de registros autênticos (CHESNEY; CITRON, 2019, p. 1793-1797).

A terceira questão é a mais profunda e a que mais diretamente toca os fundamentos democráticos que esta obra procurou articular desde o primeiro capítulo: se campanhas de influência totalmente automatizadas forem capazes de simular com suficiente fidelidade o processo de formação espontânea de opinião pública, o conceito de vontade popular que fundamenta a legitimidade eleitoral democrática torna-se operacionalmente ininteligível. Não porque a manipulação da opinião pública seja nova, ela não é: a propaganda política existe desde que a política existe. Mas porque a escala, a personalização e a imperceptibilidade das operações de influência automatizadas criam uma distância qualitativa entre a democracia que os fundamentos constitucionais pressupõem e a democracia que o ambiente tecnológico torna possível que nenhum ajuste regulatório incremental consegue fechar.

Deepfakes em Tempo Real: A Fronteira Técnica que Muda Tudo

Entre as trajetórias técnicas com maior impacto sobre a regulação eleitoral no horizonte de cinco a dez anos, nenhuma é mais disruptiva do que o desenvolvimento de sistemas de síntese de vídeo em tempo real que eliminam o processo de pós-produção que hoje ainda distingue deepfakes de conversas autênticas. Os sistemas disponíveis em 2026 para síntese de vídeo de alta qualidade ainda operam em modo de processamento diferido: o operador fornece material de treinamento, aguarda o processamento e distribui o resultado. Esse processo, mesmo que levando apenas horas, cria uma janela técnica de

intervenção entre a intenção de produzir o deepfake e sua distribuição que sistemas de monitoramento podem explorar.

Sistemas de síntese de vídeo em tempo real, que já existem em versões de baixa qualidade e cuja progressão para alta qualidade é tecnicamente previsível no horizonte de três a cinco anos, eliminam essa janela. Eles permitem que qualquer conversa por videoconferência seja sintética em tempo real, com a face e a voz de qualquer pessoa com material de treinamento suficiente substituídas pela face e a voz de qualquer outra pessoa sem que o interlocutor perceba a substituição por análise visual ou auditiva. As implicações eleitorais são imediatas e múltiplas: debates políticos transmitidos ao vivo em que a identidade dos participantes não pode ser verificada com certeza; conversas privadas entre líderes políticos cujo conteúdo pode ser fabricado retroativamente com alta fidelidade; e declarações públicas de candidatos produzidas em tempo real por operadores que nunca precisam acessar o candidato real.

A regulação de deepfakes em tempo real exige instrumentos que a Resolução TSE nº 23.732/2024 não foi projetada para fornecer, porque pressupõe a existência de um produto acabado que pode ser analisado, rotulado e removido. Deepfakes em tempo real não têm produto acabado no momento em que causam dano: eles são o próprio canal de comunicação, não o conteúdo distribuído pelo canal. A resposta regulatória para esse problema precisa migrar do modelo de controle de conteúdo para o modelo de controle de autenticação de identidade: em vez de regular o que pode ser dito ou mostrado, regular a verificação da identidade de quem fala ou aparece em transmissões com relevância eleitoral. Esse é o horizonte para o qual a Iniciativa C2PA e sistemas similares de proveniência criptográfica apontam, e é o tipo de solução técnica que os reguladores eleitorais precisam começar a estudar agora para estar prontos quando a necessidade se tornar urgente.

Microtargeting de Precisão Psicológica: Quando os Algoritmos Conhecem o Eleitor Melhor do que Ele Mesmo

A trajetória de desenvolvimento do microtargeting político segue uma progressão cujo ponto de chegada é quantitativamente diferente do ponto de partida em grau suficiente para constituir uma diferença qualitativa. O Cambridge Analytica de 2016, que causou escândalo global ao combinar dados de Facebook com modelos psicográficos para personalizar mensagens políticas, operava com modelos que previam cinco traços de personalidade a partir de dezenas de curtidas. Era tecnologia impressionante para 2016,

mas com limitações de precisão e de cobertura que os escândalos subsequentes tenderam a obscurecer na narrativa de todo-poderoso.

O microtargeting disponível em 2026, alimentado por modelos de linguagem de grande escala treinados em históricos comportamentais de escala muito superior ao que a Cambridge Analytica tinha acesso, opera em dimensão de precisão psicológica que a pesquisa de Kosinski e colaboradores de 2013 apenas esboçou. Modelos atuais são capazes de inferir não apenas traços de personalidade amplos mas estados emocionais em tempo quase real, vulnerabilidades cognitivas específicas ativadas por contextos situacionais particulares e momentos de máxima receptividade a mensagens políticas identificados por padrões comportamentais que o próprio eleitor não reconheceria como politicamente relevantes.

A compra de combustível numa determinada região às 17h de uma segunda-feira, o padrão de músicas ouvidas na semana anterior a uma eleição, a sequência de buscas realizadas depois de uma notícia específica: esses dados, combinados por modelos suficientemente treinados, permitem inferências sobre o estado psicológico eleitoral de um indivíduo com precisão que ultrapassa a capacidade de qualquer consultor político experiente baseado apenas em observação qualitativa.

O horizonte de cinco a dez anos para o microtargeting de precisão psicológica é a personalização em tempo real de mensagens eleitorais que se adaptam continuamente ao estado emocional e cognitivo detectado do receptor, calibrando conteúdo, tom, momento e canal para maximizar o impacto sobre a decisão de voto de cada eleitor individualmente. Esse sistema não precisa de deepfakes para ser politicamente disruptivo: ele produz manipulação psicológica em escala industrial usando apenas conteúdo que passaria por qualquer verificação de autenticidade, porque o que é manipulado não é a autenticidade do conteúdo mas a vulnerabilidade específica do receptor no momento específico de sua exposição.

Como argumenta Zuboff (2019), a diferença entre predição comportamental e modificação comportamental é técnica, não ética: um sistema suficientemente bom na predição inevitavelmente descobre que modificar o ambiente informacional do sujeito é o caminho mais eficiente para aumentar a precisão de suas predições, o que é funcionalmente equivalente à modificação do comportamento (ZUBOFF, 2019, p. 293-296). Para o direito eleitoral, essa equivalência não tem resposta adequada nos instrumentos disponíveis em 2026: a regulação de microtargeting eleitoral desenvolvida até aqui foca nas bases legais para o uso de dados, não nos limites éticos e jurídicos da capacidade de personalização psicológica que esses dados viabilizam. Preencher essa lacuna é a tarefa

regulatória mais urgente e mais difícil do próximo ciclo de desenvolvimento do direito eleitoral.

O Problema da Governança: Quem Regula os Reguladores?

A análise das trajetórias tecnológicas precedentes converge para uma questão de governança que é anterior e mais fundamental do que qualquer escolha regulatória específica: quem tem a capacidade, a legitimidade e os incentivos para regular tecnologias de IA eleitoral que evoluem mais rapidamente do que qualquer sistema institucional democrático foi projetado para acompanhar?

O TSE demonstrou, com a Resolução nº 23.732/2024, que autoridades eleitorais com poder normativo próprio têm vantagem estrutural sobre legislaturas em termos de velocidade de resposta. Mas a vantagem de velocidade tem um preço democrático: regulação produzida por autoridade não eleita sem aprovação parlamentar tem déficit de legitimidade que se torna mais visível à medida que a regulação se torna mais intrusiva sobre a liberdade de expressão e de comunicação política. A regulação de IA generativa em 2024 estava suficientemente distante dos limites da liberdade de expressão para que esse déficit não fosse politicamente saliente. A regulação de agentes autônomos de IA, de microtargeting de precisão psicológica e de deepfakes em tempo real em 2028 e 2030 estará substancialmente mais próxima desses limites, e a legitimidade democrática do regulador será uma questão que não poderá ser evitada.

A segunda dimensão do problema de governança é a territorial: as tecnologias de IA eleitoral são desenvolvidas predominantemente em empresas sediadas nos Estados Unidos e na China, distribuídas globalmente por plataformas que operam sob múltiplas jurisdições e utilizadas em processos eleitorais que são, por definição, soberanos. Nenhuma autoridade eleitoral nacional, por mais sofisticada que seja sua regulação doméstica, tem jurisdição sobre os sistemas de IA que afetam suas eleições.

A Resolução TSE nº 23.732/2024 regula o uso de IA eleitoral no Brasil mas não tem qualquer mecanismo de influência sobre o desenvolvimento dos sistemas que essa regulação busca governar. A cooperação regulatória internacional na área de IA eleitoral está anos atrás da cooperação necessária, e o ritmo atual de desenvolvimento institucional não é compatível com a velocidade de evolução tecnológica.

A terceira dimensão é a mais perturbadora e a que a literatura de governança de IA menos sabe como endereçar: o problema do alinhamento regulatório em contexto de IA geral. Se os sistemas de IA continuarem a progredir na direção de capacidades cada vez

mais gerais, o problema de governança eleitoral da IA deixará de ser um problema de governança de ferramentas específicas para se tornar um problema de governança de sistemas com capacidade de agência que supera a de qualquer regulador. Esse horizonte, que a literatura técnica denomina de IA de nível humano geral ou superinteligência, está fora do escopo do direito eleitoral vigente de forma absoluta: nenhum dos instrumentos analisados nesta obra foi projetado para governar sistemas de IA que possam superar a capacidade cognitiva dos reguladores que os governam. A antecipação desse horizonte não é paranoia: é a exigência mínima de prudência institucional que qualquer regulador responsável deve exercer quando o objeto de sua regulação tem trajetória de desenvolvimento conhecida e acelerada.

Democracia Aumentada: O Potencial Subutilizado da IA para Fortalecer a Participação

A análise dos riscos da IA para a democracia eleitoral, que necessariamente domina a maior parte desta obra, não deve obscurecer o fato de que as mesmas tecnologias que criam esses riscos têm potencial para fortalecer dimensões da participação democrática que os sistemas políticos contemporâneos negligenciam estruturalmente. A diferença entre IA como ameaça à democracia e IA como fortalecedora da democracia não está na tecnologia em si: está nas escolhas de design, nos incentivos que as governam e nas estruturas de poder que determinam quem controla o desenvolvimento e a implementação.

Sistemas de IA podem tornar o processo legislativo genuinamente participativo em escala que as assembleias representativas não conseguem replicar manualmente: ao sintetizar, agregar e priorizar contribuições de milhares ou milhões de cidadãos sobre políticas específicas, mantendo a diversidade de perspectivas em vez de reduzi-la a denominadores mínimos comuns que o processo político tradicional produz. Taiwan implementou sistemas desta natureza com resultados documentados de aumento de qualidade do debate público e de legitimidade percebida das decisões resultantes, precisamente o tipo de inovação democrática que os sistemas de IA mal utilizados para desinformação estão simultaneamente tornando mais necessária e mais difícil de confiar.

Sistemas de IA podem tornar a verificação de informações eleitorais instantânea e universal: integrados diretamente aos sistemas de distribuição de conteúdo, eles podem fornecer contexto, verificação e perspectiva alternativa em tempo real para qualquer conteúdo que o eleitor receba, transformando o fact-checking de uma prática especializada

e posterior em uma camada de infraestrutura informacional que opera preventivamente. A diferença entre esse sistema e a censura algorítmica que alguns temem está no design: sistemas que adicionam contexto e perspectivas alternativas sem remover conteúdo preservam a liberdade de expressão enquanto reduzem o dano epistêmico da desinformação.

Sistemas de IA podem tornar a prestação de contas de mandatários radicalmente mais transparente: ao processar automaticamente todas as declarações, votos e posicionamentos de um político ao longo de sua carreira, identificando inconsistências, evoluções e compromissos cumpridos e descumpridos, eles podem fornecer ao eleitor um retrato de seus representantes com profundidade e precisão que nenhum jornalismo de investigação manual conseguiria replicar em escala. O mesmo sistema que em mãos de operadores maliciosos produz deepfakes que fabricam declarações nunca proferidas em mãos de instituições com governança adequada pode verificar em segundos se uma declaração atribuída a um político é autêntica e contextualizada.

O potencial democrático da IA é real e precisa ser articulado com a mesma seriedade com que seus riscos são analisados, precisamente porque a narrativa exclusivamente de risco produz um viés regulatório que pode bloquear usos benéficos junto com os prejudiciais. A distinção que toda a análise desta obra procurou construir, entre IA usada para enganar e IA usada com transparência, entre IA que explora vulnerabilidades e IA que fortalece capacidades, entre IA que concentra poder e IA que o distribui, é a distinção que os reguladores, os cidadãos e os desenvolvedores precisam internalizar como critério orientador das escolhas que determinarão que tipo de democracia as tecnologias do próximo decênio tornarão possível.

A democracia que o Brasil terá em 2036 não está determinada pelas tecnologias que existirão em 2036: está determinada pelas escolhas que serão feitas nos próximos anos sobre como governar, desenvolver e usar as tecnologias que já existem. Essas escolhas estão sendo feitas agora, por atores que frequentemente não têm consciência plena de que estão fazendo escolhas com consequências democráticas de longo prazo. Este livro é uma contribuição para que mais desses atores façam essas escolhas com os olhos abertos.

O quadro a seguir traduz as trajetórias tecnológicas analisadas nas seções anteriores em perspectiva operacional, organizando cinco capacidades emergentes por estágio atual de desenvolvimento, trajetória esperada, implicação eleitoral e resposta regulatória necessária. A leitura da coluna de trajetória esperada revela o dado mais

perturbador do quadro: todas as cinco capacidades atingirão disponibilidade em dispositivos comuns entre 2027 e 2030, o que significa que o próximo ciclo eleitoral brasileiro de 2030 ocorrerá num ambiente tecnológico substancialmente diferente do de 2026, e que a preparação regulatória para esse ciclo precisa começar agora.

A coluna de resposta regulatória necessária é o coração analítico do quadro: ela revela que as respostas adequadas às cinco capacidades emergentes não são variações dos instrumentos já disponíveis na Resolução TSE nº 23.732/2024, mas instrumentos de natureza diferente, protocolos de autenticação de identidade em vez de controle de conteúdo, obrigação de supervisão humana em vez de responsabilização por resultado, certificação de transmissões ao vivo em vez de rotulagem de peças produzidas.

Essa mudança de natureza dos instrumentos necessários é o argumento mais concreto disponível para a tese de que o direito eleitoral da IA de próxima geração não será construído por aperfeiçoamento incremental do marco regulatório vigente, mas por uma revisão de segunda geração que ainda não foi iniciada por nenhuma democracia do mundo, incluindo o Brasil. A distância entre o estágio atual de 2026 e a trajetória esperada de 2028 a 2030 na coluna central é, em última análise, a medida exata do tempo disponível para que essa revisão ocorra antes que a necessidade se torne urgência.

Tabela 56. Capacidades Emergentes da IA Generativa e Implicações Eleitorais

Capacidade	Estágio Atual (2026)	Trajetória Esperada	Implicação Eleitoral	Resposta Regulatória Necessária
Síntese de vídeo em tempo real	Tecnicamente possível com hardware especializado; custo alto.	Disponível em dispositivos comuns até 2028-2030.	Deepfakes ao vivo em videoconferências e transmissões; impossibilidade de autenticar aparições remotas de candidatos.	Protocolos de autenticação biométrica para aparições remotas de candidatos; padrões C2PA.
Clonagem de voz com poucos segundos de áudio	Disponível comercialmente com 3-5 segundos de referência.	Qualidade crescente; disponível em apps gratuitos até 2027.	Áudios deepfake acessíveis a qualquer ator com smartphone; saturação de desinformação de áudio.	Marcadores digitais obrigatórios em conteúdo de áudio eleitoral; detecção em plataformas de áudio.

Agentes de IA autônomos	Sistemas capazes de executar sequências de tarefas sem supervisão humana contínua.	Capacidades crescentes; uso comercial amplo até 2027-2028.	Campanhas de desinformação autônomas que se adaptam em tempo real sem intervenção humana.	Responsabilização legal por sistemas autônomos; obrigação de supervisão humana em contexto eleitoral.
Personalização generativa em escala	Produção de variantes de mensagens para cada receptor individualmente.	Integração com sistemas de ad tech até 2027.	Cada eleitor recebe mensagem única gerada por IA, impossível de monitorar ou auditar em escala.	Obrigação de registro de conteúdo gerado por IA mesmo quando personalizado; auditoria por amostragem.
Deepfakes de imagem em tempo real	Apps de câmera com filtros de substituição de rosto já disponíveis.	Qualidade de deepfake real-time até 2027.	Candidatos podem aparecer em vídeo ao vivo com aparência alterada; eleitores não conseguem verificar autenticidade.	Certificação de transmissões ao vivo de candidatos; protocolo de autenticidade para debates e entrevistas.

Fonte: Elaborado pelo autor com base em Russell e Norvig (2022), Brynjolfsson e McAfee (2014) e relatórios técnicos de OpenAI, Google e Anthropic (2024-2025).

O Metaverso e as Eleições Virtuais: Um Horizonte Plausível

O desenvolvimento de ambientes de realidade virtual e aumentada, genericamente denominados metaverso, abre perspectivas para a comunicação política que a maioria dos sistemas regulatórios eleitorais ainda não começou a endereçar seriamente, em parte porque a penetração de mercado desses ambientes permanece restrita a segmentos específicos de usuários em 2026, e em parte porque a complexidade jurídica que eles introduzem desencoraja a antecipação regulatória de problemas que ainda não se materializaram em casos concretos. Essa postura de esperar para ver é compreensível como gestão de atenção institucional, mas é estrategicamente equivocada: as tecnologias de realidade aumentada seguem trajetória de redução de custo e expansão de funcionalidades que tornará a penetração em escala de consumo de massa plausível antes do final desta década, e a janela para construir marcos regulatórios adequados antes que

os primeiros conflitos eleitorais se materializem está se fechando mais rapidamente do que os ciclos legislativos conseguem perceber.

A relevância dos ambientes virtuais para a comunicação política não está apenas na possibilidade de comícios virtuais ou debates em ambientes imersivos, embora esses sejam os casos de uso mais óbvios. Está na combinação de três características que os distinguem qualitativamente de todos os outros canais de comunicação política analisados nesta obra.

A primeira é a presença simulada: ambientes de realidade virtual e aumentada produzem no usuário uma experiência de presença física que nenhuma transmissão de vídeo bidimensional consegue replicar, com consequências para o impacto persuasivo das mensagens políticas que a pesquisa em comunicação ainda está começando a documentar. A segunda é a personalização radical: em ambientes virtuais, cada participante pode ter uma experiência simultaneamente coletiva e individualmente customizada, com o avatar do candidato adaptando comportamentos, gestos e até aspectos de sua aparência para corresponder às preferências detectadas de cada interlocutor individual. E a terceira é a rastreabilidade comportamental: ambientes virtuais coletam dados comportamentais de uma granularidade sem precedente, incluindo movimentos oculares, padrões de atenção, respostas fisiológicas mediadas por sensores de dispositivo e comportamentos de navegação no espaço virtual que revelam preferências e vulnerabilidades que o próprio usuário não reconheceria como politicamente relevantes.

Do ponto de vista regulatório, o metaverso apresenta desafios qualitativamente novos que a extrapolação dos instrumentos vigentes não consegue endereçar adequadamente. O primeiro é a identificação dos participantes em ambientes onde a representação por avatares é estrutural: quando um eleitor interage com o avatar de um candidato num comício virtual, está interagindo com o candidato real, com um agente de IA que simula o candidato, com um operador humano controlando o avatar em tempo real ou com alguma combinação dessas possibilidades? Nenhum dos instrumentos de autenticidade disponíveis em 2026 foi projetado para operar em ambientes onde a dissociação entre identidade física e representação digital é uma característica do sistema, não uma anomalia.

O segundo desafio é a natureza jurídica das interações eleitorais em ambientes virtuais descentralizados que operam em infraestrutura tecnológica global sob jurisdição fragmentada. A regulação eleitoral brasileira, incluindo a Resolução TSE nº 23.732/2024, foi construída sobre a premissa de que as plataformas de comunicação eleitoral são entidades jurídicas identificáveis com representação legal no Brasil ou com obrigação de cooperação com a Justiça Eleitoral por força de acordos de cooperação. Plataformas de

metaverso baseadas em tecnologia blockchain, com governança descentralizada e sem jurisdição nacional definida, podem operar fora de qualquer mecanismo de enforcement que o TSE tem disponível atualmente.

O terceiro desafio é a regulação de avatares de candidatos produzidos por IA generativa em ambientes virtuais. Se hoje a vedação de deepfakes de candidatos estabelecida pelo artigo 9º da Resolução TSE nº 23.732/2024 alcança com relativa clareza vídeos sintéticos distribuídos em plataformas convencionais, a regulação de avatares interativos em tempo real que representam candidatos em ambientes virtuais exige uma interpretação expansiva do conceito de conteúdo sintético que a Resolução não antecipou. Um avatar que simula o candidato em conversas individuais com eleitores num ambiente de realidade aumentada, respondendo em tempo real com a voz e a aparência do candidato usando IA generativa, é tecnicamente um agente virtual para fins do artigo 2º, inciso III, da Resolução, mas produz uma experiência de presença e autenticidade que nenhum chatbot de texto consegue replicar e cujo potencial de engano é qualitativamente distinto.

A criação de marcos regulatórios para a comunicação política em ambientes virtuais é uma das fronteiras mais abertas do direito eleitoral digital e uma agenda que precisa ser iniciada antes que a urgência de casos concretos force respostas apressadas. As perguntas que esses marcos precisarão responder incluem: qual é o equivalente funcional da rotulagem obrigatória em ambientes imersivos onde interrupções textuais comprometem a experiência do usuário? Como o princípio da autenticidade eleitoral se aplica a ambientes onde a representação digital é estruturalmente distinta da identidade física? E como a jurisdição eleitoral brasileira se estende a eventos que ocorrem em infraestrutura tecnológica global fora do controle de qualquer regulador nacional? Nenhuma dessas perguntas tem resposta adequada no ordenamento vigente, e o ritmo atual de desenvolvimento tecnológico sugere que elas precisarão de resposta antes que estejamos prontos para fornecê-la.

Inteligência Artificial e Desinformação: A Escalada Assimétrica

Uma das tendências mais preocupantes no horizonte da IA eleitoral é o que os pesquisadores denominam escalada assimétrica: a velocidade com que as capacidades ofensivas de produção de desinformação se desenvolvem supera consistentemente a velocidade com que as capacidades defensivas de detecção e resposta são desenvolvidas.

Essa assimetria não é acidental nem superável por investimento adicional em pesquisa defensiva: ela reflete uma diferença estrutural nos incentivos, nos recursos e nas

restrições que governam os dois lados da equação, e compreendê-la com precisão é condição para não desperdiçar recursos regulatórios em estratégias que a assimetria estrutural condena ao fracasso.

Do lado ofensivo, a desinformação eleitoral é um mercado competitivo em que atores com diferentes recursos e motivações investem continuamente no aprimoramento de suas capacidades, impulsionados pelos retornos políticos e econômicos do sucesso. Esses atores não precisam obedecer a princípios éticos, não precisam obter aprovação institucional para experimentar novas abordagens, não precisam publicar seus métodos para revisão por pares e não precisam demonstrar que suas inovações não causam danos antes de implementá-las. A velocidade de inovação ofensiva é limitada apenas por recursos financeiros e capacidade técnica, dois fatores que estão progressivamente se tornando mais acessíveis com a democratização das ferramentas de IA.

Do lado defensivo, a detecção e o combate à desinformação dependem de investimentos institucionais que são frequentemente subfinanciados em relação à escala do problema, mais lentos na adoção de novas tecnologias por razões de processo institucional e sujeitos a restrições legais e éticas que os atores ofensivos não enfrentam. Um sistema de detecção automática de deepfakes desenvolvido por uma instituição pública precisa passar por processos de aquisição, validação, aprovação orçamentária e implantação que levam meses ou anos. Um sistema de geração de deepfakes atualizado por um ator privado pode ser implantado em dias. Essa assimetria de velocidade não é superável por determinação política ou por investimento adicional dentro do paradigma regulatório vigente: ela é estrutural.

Chesney e Citron (2019) formularam o argumento mais preciso disponível sobre a dinâmica dessa assimetria no contexto específico dos deepfakes: os geradores precisam apenas produzir conteúdo que engane humanos; os detectores precisam identificar conteúdo que já enganou humanos e foi amplamente distribuído. O gerador opera no tempo de produção; o detector opera no tempo de resposta. Em ambiente de circulação acelerada, o dano é causado no tempo de produção e o detector chega no tempo de resposta, quando o conteúdo já alcançou seu público-alvo e formou percepções que a correção posterior raramente consegue reverter completamente (CHESNEY; CITRON, 2019, p. 154-160).

A assimetria tecnológica tem uma dimensão adicional que a análise de Chesney e Citron não esgota: a assimetria de dados. Sistemas de geração de desinformação são treinados em dados que incluem todas as desinformações bem-sucedidas do passado, o que significa que cada operação bem-sucedida de desinformação contribui para melhorar a próxima. Sistemas de detecção são treinados em dados de desinformações identificadas,

o que significa que desinformações que não foram detectadas não contribuem para melhorar os sistemas de defesa. Essa dinâmica de aprendizado assimétrico produz uma vantagem cumulativa dos sistemas ofensivos sobre os defensivos que cresce com o tempo, não diminui.

A resposta regulatória mais eficaz para a escalada assimétrica não está na competição direta entre sistemas de geração e detecção, que a estrutura da assimetria condena ao fracasso de longo prazo. Está em três abordagens complementares que operam em pontos distintos da cadeia de produção e distribuição de desinformação. A primeira é a proveniência criptográfica, que desloca o problema de detecção de conteúdo para autenticação de origem: em vez de perguntar se um vídeo foi gerado por IA, perguntar se tem cadeia de proveniência verificável, o que é tecnicamente muito mais acessível porque não depende de detectar artefatos mas de verificar registros criptográficos. A segunda é a responsabilização de intermediários por amplificação em vez de apenas por hospedagem, que cria incentivos para que plataformas desenvolvam sistemas de detecção internos com recursos superiores aos que qualquer regulador externo consegue mobilizar.

A terceira é o investimento em literacia midiática que desenvolve resiliência no receptor, a única estratégia que não depende de vencer a corrida armamentista tecnológica porque opera num plano diferente: não na detecção do conteúdo sintético, mas na redução da vulnerabilidade do eleitor ao seu impacto.

O Projeto de Lei nº 2.338/2023: O Marco Legal da IA no Brasil

O Projeto de Lei nº 2.338/2023, que institui o Marco Legal da Inteligência Artificial no Brasil, representa o esforço legislativo mais abrangente disponível para complementar e em alguns aspectos modificar o arcabouço regulatório estabelecido pela Resolução TSE nº 23.732/2024. Sua análise neste capítulo é necessariamente provisória, dado que o projeto estava em tramitação avançada ao tempo da redação desta obra, com possibilidade de emendas substantivas antes de sua aprovação final, mas os elementos centrais de sua arquitetura estão suficientemente consolidados para análise das implicações eleitorais que produzirão quando em vigor.

A arquitetura do PL 2.338/2023 adota abordagem baseada em risco com inspiração declarada no AI Act europeu, classificando sistemas de IA por nível de risco e estabelecendo obrigações proporcionais a cada categoria. Para o contexto eleitoral, essa arquitetura tem implicações que vão além da conformidade regulatória específica: ela representa uma mudança de paradigma do modelo setorial que o TSE adotou com a

Resolução nº 23.732/2024 para um modelo horizontal que enquadra a IA eleitoral como subconjunto de uma categoria mais ampla de sistemas de alto risco, com obrigações que decorrem da classificação de risco e não da legislação eleitoral específica.

As disposições mais relevantes do PL 2.338/2023 para o contexto eleitoral concentram-se em três eixos. O primeiro é a classificação dos sistemas de IA utilizados para influenciar eleitores ou para interferir em processos eleitorais como sistemas de alto risco, sujeitos às obrigações mais exigentes do projeto: avaliação de conformidade antes da implantação, registro em base de dados nacional de sistemas de IA, documentação técnica detalhada, medidas de supervisão humana e mecanismos de explicabilidade que permitam aos afetados compreender as decisões tomadas pelo sistema. Essa classificação tem implicação direta para ferramentas de microtargeting eleitoral, para chatbots eleitorais e para sistemas de análise de dados eleitorais: eles precisarão cumprir as obrigações de sistemas de alto risco independentemente de estarem sujeitos à Resolução TSE nº 23.732/2024.

O segundo eixo é a exigência de avaliação de impacto de direitos fundamentais para sistemas de IA em processos eleitorais, mecanismo inspirado no Relatório de Impacto à Proteção de Dados da LGPD mas com escopo mais amplo que inclui o impacto sobre direitos políticos, liberdade de expressão e integridade do processo democrático. Essa avaliação prévia, se implementada com a profundidade que o texto do projeto sugere, seria o instrumento mais eficaz disponível para identificar riscos eleitorais de sistemas de IA antes de sua implantação, ao invés de aguardar que esses riscos se materializem em casos concretos. A limitação estrutural da avaliação de impacto, que o histórico de instrumentos similares no campo da proteção de dados revela, é que avaliações produzidas pelo próprio operador do sistema têm valor limitado sem mecanismo de verificação independente que as valide ou conteste.

O terceiro eixo é a proibição de sistemas de IA que utilizem técnicas subliminares para influenciar comportamentos políticos, dispositivo que dialoga diretamente com a vedação do artigo 9º, inciso V, da Resolução TSE nº 23.732/2024, mas com alcance potencialmente mais amplo por operar horizontalmente em todo o ecossistema de IA e não apenas em contexto de propaganda eleitoral. A articulação entre a vedação da Resolução TSE, que se aplica ao uso em campanha, e a proibição do PL 2.338/2023, que se aplicaria ao desenvolvimento e à comercialização do sistema, poderia criar um regime de dupla proibição que alcança tanto o produtor da ferramenta quanto o operador que a utiliza em contexto eleitoral.

A interação entre o PL 2.338/2023 e a Resolução TSE nº 23.732/2024, quando o projeto for aprovado, produzirá um sistema regulatório de camadas que o operador de tecnologia eleitoral precisará navegar com precisão. Obrigações que decorrem do marco geral de IA e obrigações que decorrem da regulação eleitoral específica não são necessariamente redundantes: elas frequentemente se complementam, com o marco geral estabelecendo os requisitos de desenvolvimento e documentação do sistema e a regulação eleitoral estabelecendo os requisitos de uso no contexto específico da campanha. Mas podem também ser conflitantes, quando as obrigações de transparência do marco geral de IA entram em tensão com estratégias de confidencialidade de campanhas que não querem revelar suas ferramentas e metodologias a adversários. Essa tensão entre transparência regulatória e confidencialidade estratégica é uma das que o sistema regulatório de IA eleitoral brasileiro precisará resolver antes que o ciclo de 2030 torne o problema urgente.

📌 PL 2.338/2023 — Disposições Relevantes para o Contexto Eleitoral

Art. 14, IV (sistemas de alto risco): Sistemas de IA utilizados para influenciar o comportamento de pessoas em contexto de processos eleitorais ou democráticos, incluindo sistemas de microtargeting eleitoral e geração de conteúdo político personalizado. Art. 15 (sistemas de IA proibidos): É vedado o desenvolvimento, a comercialização e o uso de sistemas de IA que empreguem técnicas subliminares, manipulativas ou enganosas para influenciar o comportamento de pessoas, especialmente em contextos eleitorais. Art. 27 (obrigações de transparência): Os responsáveis por sistemas de IA de alto risco devem manter registro auditável das decisões tomadas pelo sistema e disponibilizá-lo às autoridades competentes mediante requisição fundamentada.

Brasil. PL 2.338/2023 (Marco Legal da IA). Versão aprovada pela Comissão Especial do Senado Federal, 2024.

Governança Democrática da IA: Quem Decide as Regras?

A discussão sobre regulação da IA eleitoral não pode ser reduzida a uma questão técnica ou jurídica sem que essa redução produza exatamente o tipo de déficit democrático que a regulação pretende prevenir. É, fundamentalmente, uma questão de governança democrática de primeira ordem: quem decide como a IA pode ser usada nas eleições que determinam quem governa? Com base em quais valores, produzidos por quais processos, com participação de quem e com que mecanismos de accountability para as consequências das decisões tomadas? Essas perguntas não têm resposta técnica: têm resposta política,

e a qualidade dessa resposta determina se a regulação da IA eleitoral fortalece ou enfraquece a democracia que pretende proteger.

O modelo atual de governança da IA eleitoral no Brasil concentra o poder normativo no TSE, que produz resoluções com base em processo consultivo que inclui audiências públicas, análise comparada de regulações internacionais e participação de especialistas técnicos e jurídicos. Esse modelo tem virtudes reais que este livro documentou extensivamente: velocidade de resposta, especificidade normativa e capacidade de adaptação que sistemas legislativos tradicionais não conseguem replicar. Mas tem também um déficit de legitimidade democrática que se torna mais saliente à medida que as decisões regulatórias se tornam mais intrusive sobre o exercício da liberdade de expressão e de comunicação política.

Uma resolução que define o que constitui conteúdo sintético vedado, que estabelece quando um chatbot está simulando interação humana de forma irregular e que determina quais técnicas de persuasão algorítmica são incompatíveis com a democracia está tomando decisões que afetam profundamente o espaço público democrático, e é legítima perguntar se o processo pelo qual essas decisões são tomadas tem a representatividade que seu impacto exige.

Essa observação não é uma crítica à competência ou à boa-fé do TSE, que demonstrou capacidade técnica e responsividade notáveis no campo da regulação de IA eleitoral. É uma observação sobre os limites estruturais de qualquer modelo de regulação técnica para questões com implicações profundas sobre o exercício da soberania popular. O princípio democrático que fundamenta a legitimidade das eleições, que as regras que governam a competição pelo poder devem ser produzidas com participação dos competidores e dos cidadãos que os elegem, aplica-se com força especial à regulação da IA eleitoral: quando as regras que governam como candidatos podem se comunicar com eleitores são produzidas por um tribunal cujos membros não são eleitos, há uma tensão constitucional que os processos participativos do TSE atenuam mas não eliminam.

Habermas (1997), cuja teoria da democracia deliberativa fornece um dos fundamentos normativos mais influentes para pensar a legitimidade de decisões em sociedades pluralistas, argumenta que a validade de normas que afetam todos os cidadãos depende da possibilidade de que todos os afetados participem, em condições de igualdade, do processo discursivo que as produz. Aplicado ao contexto da regulação de IA eleitoral, esse critério revela que processos consultivos que incluem audiências públicas são necessários mas insuficientes: audiências públicas tendem a ser dominadas por atores organizados com recursos para participar, produzindo uma seleção não representativa das

perspectivas que deveriam informar a regulação. Os cidadãos com menor literacia midiática, que são os mais vulneráveis às condutas que a regulação visa coibir, são sistematicamente os menos capazes de participar dos processos que definem como essa proteção será estruturada (HABERMAS, 1997, p. 187-193).

Bobbio (1986), num argumento que permanece analiticamente poderoso quatro décadas depois de sua formulação, identificou que a qualidade de uma democracia não se mede apenas pelas regras que governam as eleições, mas pela extensão em que as regras que governam os processos de decisão que afetam os cidadãos são elas próprias produzidas democraticamente. A meta-democracia, a democracia sobre como as regras democráticas são feitas, é o critério pelo qual sistemas que se autodeclaram democráticos revelam a profundidade real de seu compromisso com o princípio da soberania popular (BOBBIO, 1986, p. 9-19). A governança da IA eleitoral é, nesse sentido, um teste de meta-democracia: a democracia que permite que as regras do jogo eleitoral sejam determinadas por processos opacos ou por atores não representativos está cedendo um ponto de controle sobre sua própria reprodução que é muito difícil de recuperar depois.

A inclusão de representantes da sociedade civil, de movimentos sociais, de grupos historicamente sub-representados no processo político, como populações indígenas, quilombolas e periferias urbanas, e do público em geral no processo de construção da regulação da IA eleitoral é uma condição de legitimidade que os processos regulatórios atuais ainda não satisfazem plenamente. Não porque os reguladores atuais tenham interesses ilegítimos ou perspectivas inadequadas: mas porque regulação produzida sem a participação dos mais vulneráveis à desinformação eleitoral tende a proteger melhor os atores que têm recursos para participar do processo do que os que têm mais a perder com a ausência de proteção adequada.

O PL 2.338/2023, com seu Conselho Nacional de Inteligência Artificial, representa um passo relevante na direção de uma governança da IA com maior representatividade. Mas seu potencial de contribuição para a governança democrática da IA eleitoral depende de duas condições que o texto do projeto não garante automaticamente: a composição efetiva do conselho, que precisará incluir não apenas especialistas técnicos e representantes de indústria mas vozes de grupos com menor poder político e menor capacidade de articulação institucional; e a qualidade dos processos participativos que o conselho adotar, que precisarão ir além de consultas formais para incluir mecanismos de deliberação que permitam a formação de perspectivas informadas por parte de cidadãos sem expertise técnica prévia.

A experiência internacional fornece modelos de governança participativa da IA que podem informar o desenvolvimento brasileiro. O modelo taiwanês de governo aberto, com plataformas digitais que permitem deliberação coletiva sobre políticas complexas com participação de cidadãos sem expertise técnica, demonstrou que é possível produzir consenso regulatório sobre questões tecnológicas com representatividade genuinamente ampla sem sacrificar a qualidade técnica das decisões resultantes.

O modelo do Comitê Europeu de Proteção de Dados, que coordena as autoridades nacionais de proteção de dados dos países membros e produz orientações por processo deliberativo entre reguladores com mandatos democráticos distintos, oferece um modelo de governança regulatória especializada com maior legitimidade democrática do que o modelo de autoridade única. E o modelo de citizen assemblies, assembleias de cidadãos selecionados por amostragem representativa para deliberar sobre questões complexas, tem sido usado com sucesso no Irlanda e na França para produzir recomendações de política sobre questões em que o processo legislativo convencional estava bloqueado por polarização ou por captura por interesses organizados.

Nenhum desses modelos é transplantável diretamente para o contexto institucional brasileiro sem adaptação. Mas todos eles apontam para uma direção comum: a governança democrática de tecnologias que afetam profundamente o exercício da soberania popular requer processos que vão além da consulta técnica especializada para incluir deliberação genuinamente representativa. A pergunta que o Brasil precisa responder antes de 2030 não é apenas como regular a IA eleitoral, mas quem tem o direito e a responsabilidade de fazer essas escolhas e por meio de quais processos. A resposta a essa pergunta é ela própria uma escolha democrática de primeira importância, e a urgência de endereçá-la cresce a cada ciclo eleitoral em que a tecnologia avança mais rapidamente do que a governança.

■ RESUMO DO CAPÍTULO:

Este capítulo empreendeu a análise prospectiva mais ambiciosa desta obra, tratando o futuro da IA eleitoral não como futurologia especulativa mas como análise de trajetórias em curso com grau identificável de confiança, distinguindo o que é altamente provável no horizonte de dois a cinco anos, o que é plausível no horizonte de cinco a dez anos e o que é especulativo além desse horizonte. A abertura estabeleceu o fundamento metodológico a partir de três autores centrais. Russell e Norvig identificaram a não-linearidade como o princípio que torna a prospectiva baseada em extrapolação linear sistematicamente inadequada para o campo da IA. Brynjolfsson e McAfee articularam o mecanismo econômico das tecnologias exponenciais e o ponto de inflexão da curva que a IA generativa atingiu entre 2020 e 2024. Dafoe sistematizou a tensão estrutural entre velocidade de inovação tecnológica e velocidade de adaptação institucional como o problema de governança mais fundamental que democracias enfrentam no campo da IA. A análise da quarta geração de modelos de linguagem identificou três avanços com implicações

eleitorais diretas: a expansão radical de janelas de contexto que viabiliza análises estratégicas de escala e complexidade antes impossíveis, produzindo vantagem competitiva da mesma ordem que sistemas de alta frequência conferiram a traders financeiros; a integração genuinamente multimodal que colapsa o pipeline de produção de deepfakes de múltiplas ferramentas especializadas para interface de prompt único, transformando deepfakes de arma de atores sofisticados em commodity disponível a qualquer campanha; e a autonomia operacional crescente de agentes de IA que executam objetivos estratégicos sem instrução humana específica, criando cadeias de responsabilização suficientemente opacas para que negação plausível seja industrialmente escalável. Os agentes autônomos de IA foram analisados como ruptura na natureza das operações de influência, não apenas evolução de capacidades, com três questões jurídicas para as quais o ordenamento vigente não tem resposta: atribuição de autoria quando o agente decide autonomamente produzir conteúdo irregular, detecção quando redes de personas sintéticas aprendem a imitar comportamento orgânico, e o problema filosófico mais profundo da ininteligibilidade operacional do conceito de vontade popular quando campanhas de influência totalmente automatizadas simulam fidelmente a formação espontânea de opinião pública. Os deepfakes em tempo real foram identificados como a trajetória mais disruptiva no horizonte de três a cinco anos, por eliminar a janela técnica de intervenção que existe entre produção e distribuição e por exigir uma migração regulatória do modelo de controle de conteúdo para o modelo de autenticação de identidade que a Resolução TSE nº 23.732/2024 não foi projetada para fornecer. O microtargeting de precisão psicológica foi analisado como ameaça que não requer deepfakes para ser politicamente disruptiva, porque opera na vulnerabilidade do receptor e não na autenticidade do conteúdo, com a equivalência funcional entre predição e modificação comportamental identificada por Zuboff como o problema regulatório mais urgente e menos endereçado do próximo ciclo. O problema de governança foi desenvolvido em três dimensões: a legitimidade democrática do regulador, cujo déficit crescerá à medida que a regulação se aproximar dos limites da liberdade de expressão; a territorialidade, porque nenhuma autoridade eleitoral nacional tem jurisdição sobre os sistemas de IA que afetam suas eleições; e o alinhamento regulatório em contexto de IA geral, que está fora do escopo do direito eleitoral vigente de forma absoluta. O metaverso foi analisado como horizonte plausível com três desafios regulatórios qualitativamente novos: identificação em ambientes onde avatares são estruturais, jurisdição em plataformas descentralizadas e regulação de avatares interativos que produzem experiência de presença e autenticidade qualitativamente distinta de conteúdo sintético convencional. A escalada assimétrica foi desenvolvida com a adição da dimensão de assimetria de dados, ausente na análise de Chesney e Citron: desinformações bem-sucedidas melhoram os sistemas ofensivos enquanto desinformações não detectadas não contribuem para os sistemas defensivos, produzindo vantagem cumulativa que cresce com o tempo. As três abordagens complementares que operam fora da corrida armamentista, proveniência criptográfica, responsabilização por amplificação e literacia midiática, foram identificadas como as estratégias regulatórias com maior potencial de eficácia estrutural. O PL 2.338/2023 foi analisado em seus três eixos eleitoralmente relevantes: classificação de sistemas de influência eleitoral como alto risco, avaliação de impacto de direitos fundamentais e proibição de técnicas subliminares, com a tensão entre transparência regulatória e confidencialidade estratégica identificada como problema não resolvido. A governança democrática da IA foi desenvolvida com Habermas sobre as condições deliberativas de validade normativa e Bobbio sobre a meta-democracia como critério de qualidade democrática, culminando na apresentação de três modelos internacionais de governança participativa com seus princípios transferíveis. O capítulo encerrou com o argumento que é também a tese final do livro: a democracia que o Brasil terá em 2036 não está determinada pelas tecnologias que existirão em 2036, mas pelas escolhas que serão feitas nos próximos anos sobre como governar, desenvolver e usar as tecnologias que já existem.

QUESTÕES PARA REFLEXÃO E DEBATE

1. O capítulo argumenta que a Resolução TSE nº 23.732/2024 foi calibrada para um mundo em que a produção de conteúdo sintético de alta qualidade ainda exigia recursos não triviais, e que esse mundo

está desaparecendo. A democratização radical das ferramentas de deepfake transforma a lógica regulatória de cima para baixo, que alcança atores sofisticados com recursos identificáveis, em instrumento insuficiente para um ecossistema em que qualquer eleitor com smartphone pode produzir e distribuir conteúdo sintético convincente. Que modelo regulatório alternativo seria mais adequado para esse cenário: responsabilização de plataformas pela amplificação, proveniência criptográfica obrigatória, ou uma combinação de instrumentos que ainda não existe em nenhuma democracia?

2. A questão da atribuição de autoria em sistemas de IA autônomos é identificada como a vulnerabilidade estrutural mais significativa do direito eleitoral no horizonte de 2028. Que doutrina de responsabilidade, baseada em quais princípios do direito civil, do direito penal ou do direito eleitoral, seria mais adequada para responsabilizar candidatos por violações cometidas por agentes autônomos sem instrução humana específica para o ato irregular?

3. O dividendo do mentiroso em escala sistêmica, o benefício que atores desonestos extraem da desconfiança generalizada em toda evidência, foi identificado como consequência que persiste mesmo quando os sistemas de detecção melhoram. Esse fenômeno sugere que a estratégia de investimento em sistemas de detecção cada vez mais sofisticados pode paradoxalmente aumentar a vulnerabilidade democrática ao produzir mais ceticismo sobre a autenticidade de qualquer conteúdo. Como os reguladores deveriam calibrar o investimento em detecção para maximizar a proteção democrática sem ampliar o dividendo do mentiroso?

4. O microtargeting de precisão psicológica que se adapta em tempo real ao estado emocional e cognitivo detectado de cada eleitor foi identificado como ameaça que produz manipulação psicológica em escala industrial usando conteúdo que passaria por qualquer verificação de autenticidade. A regulação disponível foca nas bases legais para uso de dados mas não nos limites da capacidade de personalização psicológica que esses dados viabilizam. Que princípios éticos e jurídicos deveriam orientar a construção de limites sobre a personalização psicológica de mensagens políticas, e como esses limites seriam operacionalizáveis na prática?

5. A governança democrática da IA eleitoral enfrenta uma tensão estrutural entre a velocidade de resposta que o modelo TSE oferece e a legitimidade democrática que apenas processos com participação mais ampla podem fornecer. O modelo de citizen assemblies, com cidadãos selecionados por amostragem representativa deliberando sobre questões técnicas complexas com suporte especializado, foi apresentado como alternativa promissora. Que adaptações específicas ao contexto institucional brasileiro seriam necessárias para tornar esse modelo viável no ciclo regulatório que precede as eleições de 2030?

6. O metaverso apresenta o problema da dissociação estrutural entre identidade física e representação digital como característica do sistema, não como anomalia. A regulação eleitoral baseada no conceito de candidato como pessoa física identificável enfrenta um desafio conceitual fundamental nesse ambiente: quando um avatar de candidato interage com um eleitor num ambiente virtual, qual é a relação jurídica estabelecida, e como o direito eleitoral deveria caracterizá-la para fins de responsabilização, propaganda eleitoral e proteção do eleitor?

7. A assimetria de dados identificada no capítulo, em que desinformações bem-sucedidas melhoram sistemas ofensivos enquanto desinformações não detectadas não contribuem para sistemas defensivos, sugere que a estratégia mais eficaz de longo prazo pode ser a transparência obrigatória sobre operações de IA eleitoral, que tornaria detectáveis as desinformações atualmente invisíveis. Que nível de transparência sobre ferramentas, metodologias e resultados de operações de IA eleitoral seria compatível com a proteção da confidencialidade estratégica legítima de campanhas?

8. O argumento final do capítulo, que a democracia que o Brasil terá em 2036 está sendo determinada pelas escolhas feitas agora, implica que as decisões regulatórias sobre IA eleitoral têm consequências constitucionais de longo prazo que superam em importância as consequências eleitorais imediatas que motivam a maioria dessas decisões. Que mecanismos institucionais poderiam garantir que as escolhas regulatórias sobre IA eleitoral sejam avaliadas pela sua adequação aos fundamentos constitucionais democráticos de longo prazo, e não apenas pela sua efetividade nas eleições do próximo ciclo?

NA PRÁTICA | O que este capítulo significa para quem trabalha com política e campanhas

1. A Resolução TSE 23.732/2024 está envelhecendo em tempo real. Prepare-se para a segunda geração.

A Resolução foi calibrada para um mundo em que deepfakes exigiam recursos significativos, ferramentas especializadas e conhecimento técnico. Esse mundo está desaparecendo. Modelos multimodais de quarta geração colapsam todo o pipeline de produção de conteúdo sintético para uma interface de prompt único acessível a qualquer usuário. Antes de 2028, qualquer ator com acesso à internet terá capacidade de produzir deepfakes que em 2024 exigiam equipes especializadas. O profissional de campanha e o advogado eleitoral que planejam para 2030 precisam começar a acompanhar os trabalhos do Congresso sobre o PL 2.338/2023 e as discussões do TSE sobre revisão da Resolução. A segunda geração regulatória está sendo gestada agora.

2. Agentes de IA autônomos criam um problema de responsabilidade que você precisa antecipar nos contratos de hoje.

Se você contrata uma agência que utiliza agentes de IA autônomos para executar estratégias de campanha, e esse agente decide autonomamente produzir conteúdo que viola a Resolução TSE, a cadeia de responsabilização entre sua instrução original e a violação específica pode ser suficientemente longa para que defesas de desconhecimento sejam levantadas por todos os elos. Mas o TSE responsabilizará o candidato objetivamente. A proteção disponível hoje é contratual: cláusula expressa que proíbe uso de agentes autônomos sem supervisão humana contínua em decisões de conteúdo eleitoral, com responsabilidade solidária da agência por qualquer violação praticada por sistemas automatizados. Inclua essa cláusula em qualquer contrato firmado a partir de agora.

3. Clonagem de voz com três segundos de referência já é realidade. Sua campanha está preparada?

Ferramentas comerciais disponíveis em 2026 clonam a voz de qualquer pessoa com três a cinco segundos de referência de áudio. Isso significa que qualquer declaração pública de seu candidato, qualquer entrevista gravada, qualquer discurso postado nas redes sociais, é material de treinamento suficiente para produzir áudios sintéticos convincentes. A questão não é se isso vai acontecer com seu candidato: é quando. O protocolo de resposta a deepfakes de áudio precisa estar estabelecido antes do período eleitoral, com perito forense especializado em análise de áudio pré-contratado e protocolo de notificação às plataformas de distribuição de áudio como Spotify, podcasts e grupos de WhatsApp. A clonagem de voz já não é ameaça futura: é risco presente.

4. O dividendo do mentiroso é real e cresce com cada deepfake descoberto. Sua estratégia de resposta precisa considerar isso.

Cada vez que um deepfake eleitoral é exposto, uma parte do eleitorado não conclui que o deepfake era falso: conclui que toda evidência pode ser falsa. Essa desconfiança generalizada beneficia estruturalmente os atores que produzem mais desinformação, porque contamina retroativamente a credibilidade de registros autênticos. Para o candidato que responde a deepfakes, isso tem uma implicação estratégica específica: a resposta deve incluir não apenas a desmentida do conteúdo específico, mas elementos que reforcem a credibilidade geral do candidato como fonte confiável. Responder apenas ao fato específico sem endereçar a erosão de confiança que o ambiente de deepfakes produz é responder de forma incompleta ao dano real sofrido.

5. Microtargeting que explora vulnerabilidades cognitivas já é tecnicamente possível e regulatoriamente subendereçado.

A regulação de microtargeting eleitoral disponível em 2026 foca nas bases legais para uso de dados: consentimento, legítimo interesse, dados sensíveis. Ela não regula os limites da capacidade de personalização psicológica que esses dados viabilizam. Sistemas que identificam estados emocionais em tempo quase real e calibram mensagens eleitorais para explorar vulnerabilidades cognitivas específicas estão operando numa zona regulatória sem cobertura clara. Para o advogado eleitoral, isso significa dois riscos simultâneos: o risco de que seu candidato seja vítima de microtargeting psicológico por adversários, e o risco de que ferramentas de marketing contratadas por sua campanha estejam operando em violação de normas que ainda não foram formalmente estabelecidas mas que serão construídas retroativamente pela jurisprudência.

6. A democracia que o Brasil terá em 2036 está sendo determinada pelas escolhas que você faz agora.

Este não é um argumento abstrato. É um argumento sobre responsabilidade profissional concreta. Os profissionais que atuam em campanhas eleitorais, os advogados que as assessoram, os dirigentes que tomam decisões sobre o uso de tecnologia, os legisladores que votam o PL 2.338/2023 e os ministros do TSE que editam resoluções estão todos fazendo escolhas com consequências democráticas de longo prazo que excedem em importância as consequências eleitorais imediatas que motivam essas decisões. Usar a IA para enganar ou manipular pode funcionar num ciclo eleitoral. Erode a confiança institucional que torna qualquer vitória eleitoral futura sustentável. A pergunta que todo profissional precisa fazer não é apenas 'isso é legal?' mas 'que tipo de democracia estou ajudando a construir com essa escolha?'

7. O metaverso parece distante. Não está. Comece a acompanhar agora para não ser surpreendido em 2028.

Em 2016, o WhatsApp parecia irrelevante para eleições presidenciais. Em 2018, foi o principal vetor de desinformação eleitoral do Brasil. A curva de adoção de tecnologias de realidade aumentada segue a mesma trajetória: penetração restrita hoje, potencial de massa amanhã. Plataformas de metaverso que permitem comícios virtuais, interações personalizadas com avatares de candidatos e coleta de dados comportamentais de granularidade sem precedente ainda não têm cobertura regulatória eleitoral adequada em nenhuma democracia do mundo. O profissional que começa a acompanhar essa agenda agora, entendendo os desafios de autenticação de identidade, jurisdição de plataformas descentralizadas e regulação de avatares interativos, estará anos à frente de adversários que descobrirão o problema quando ele já for urgente.

CAPÍTULO 18

É possível salvar a democracia?

Propostas, Caminhos e o Imperativo da Ação Coletiva na Era da Inteligência Artificial

A democracia não se salva sozinha. Ela se salva com cidadãos que a valorizam o suficiente para defendê-la com as armas de cada época. Hoje, essas armas incluem a regulação da IA, a literacia midiática e a coragem de dizer a verdade quando mentir é mais fácil e mais lucrativo

A Pergunta que Não Pode Ser Evitada

É possível salvar a democracia? A pergunta que intitula este capítulo final pode soar hiperbólica para quem chegou a este livro sem ter acompanhado de perto a acumulação de evidências descritas nos dezessete capítulos anteriores. Para quem chegou até aqui, ela é a pergunta mais honesta, mais urgente e mais difícil que se pode formular no campo em que esta obra se situa, e sua hipérbole aparente é o primeiro sintoma do problema que ela descreve: quando a deterioração democrática é gradual o suficiente, a pergunta sobre se ela está ocorrendo parece exagerada precisamente para quem está no meio do processo.

A democracia, como argumentaram Levitsky e Ziblatt (2018) em *Como as Democracias Morrem*, não morre subitamente por golpes abruptos e dramáticos na maioria dos casos contemporâneos. Ela se deteriora gradualmente, por acumulação de erosões pequenas que, isoladamente, parecem toleráveis, mas cujo conjunto acaba por comprometer irreversivelmente as condições de possibilidade do regime democrático genuíno. O mecanismo é preciso: cada erosão individual é racionalizada como exceção necessária, resposta proporcional a uma ameaça específica ou ajuste técnico sem implicações sistêmicas. A erosão sistêmica só se torna visível quando já é difícil de reverter, porque o conjunto de ajustes individuais produziu uma realidade na qual as normas e as instituições que protegiam a democracia perderam a robustez que tornava sua proteção eficaz (LEVITSKY; ZIBLATT, 2018, p. 76-82).

A manipulação algorítmica do ambiente informacional eleitoral é precisamente esse tipo de erosão: invisível para a maioria dos eleitores que a experimenta, difusa na sua

responsabilidade entre plataformas, campanhas, anunciantes e desenvolvedores de ferramentas, e lenta o suficiente para que cada passo individual possa ser racionalizado como tecnicamente necessário, comercialmente legítimo ou politicamente justificado.

Nenhum ator específico acorda de manhã com a intenção de destruir a democracia. A consequência democrática emerge da agregação de escolhas individuais que nenhum de seus autores reconheceria como destruidora se analisada isoladamente. Esse é o mecanismo que torna o problema ao mesmo tempo mais difícil de denunciar, porque exige olhar para o conjunto, e mais urgente de endereçar, porque quanto mais gradual for a deterioração mais difícil será revertê-la.

A resposta à pergunta que intitula este capítulo é, como qualquer resposta honesta a uma questão desta magnitude, ao mesmo tempo sim e depende. Sim, porque os instrumentos necessários para proteger a integridade democrática na era da IA existem, são conhecidos e são tecnicamente implementáveis. Cada um dos capítulos anteriores desta obra identificou instrumentos concretos: regulação específica de IA eleitoral como a Resolução TSE nº 23.732/2024, sistemas de detecção e resposta a deepfakes, protocolos de proteção de dados, obrigações de transparência das plataformas, programas de literacia midiática, mecanismos de governança participativa. Nenhum desses instrumentos é perfeito, e cada um tem limitações analisadas com precisão ao longo desta obra. Mas o conjunto deles, implementado com consistência e coordenação, é suficiente para limitar o dano a um nível compatível com a manutenção de processos eleitorais com legitimidade democrática razoável.

Depende, porque a implementação efetiva desses instrumentos exige um conjunto de condições que não são garantidas pela simples existência dos instrumentos: vontade política dos atores com poder de implementação, investimento institucional sustentado ao longo de ciclos orçamentários e eleitorais que favorecem o curto prazo, cooperação internacional numa arena em que os interesses nacionais tendem à competição e não à colaboração, e, acima de tudo, cidadãos que compreendam o problema em profundidade suficiente para exigir soluções com a mesma urgência com que problemas mais visíveis são tratados. A última condição é a mais difícil de satisfazer e a mais fundamental: sem uma cidadania informada sobre os mecanismos pelos quais a IA eleitoral pode comprometer a autonomia política, as outras condições carecem do suporte democrático que as torna sustentáveis no longo prazo.

Este capítulo final procura sistematizar esses instrumentos e identificar as condições de sua efetividade, não como inventário técnico mas como argumento normativo: a democracia não é um estado que se alcança e se mantém automaticamente, mas uma

conquista que exige trabalho ativo de cada geração que a herda. A geração que governa o Brasil em 2026 herdou uma democracia que sobreviveu a desafios sérios nos anos recentes. Sua responsabilidade é entregá-la à próxima geração em condição não apenas de sobrevivência mas de fortalecimento. As escolhas que determinarão se essa responsabilidade será cumprida estão sendo feitas agora, por atores que incluem os leitores desta obra.

O Que Esta Obra Construiu: Uma Síntese Integrada

Ao longo de dezessete capítulos, esta obra construiu um argumento que pode ser enunciado em uma tese central e em quatro proposições subsidiárias que a sustentam. A tese central é que a inteligência artificial generativa introduz uma descontinuidade qualitativa, não apenas quantitativa, nos riscos para a integridade eleitoral democrática. Não porque a manipulação política seja nova, a propaganda política existe desde que a política existe, mas porque a combinação de escala industrial, personalização psicológica e imperceptibilidade que as ferramentas de IA tornam possíveis cria condições para a distorção do processo de formação de preferências políticas em magnitude que nenhuma tecnologia anterior havia alcançado. Essa descontinuidade qualitativa exige resposta regulatória, institucional e educacional que vai além do aperfeiçoamento incremental dos instrumentos existentes.

A primeira proposição subsidiária é que o Brasil chegou às eleições de 2026 com o marco regulatório mais específico do mundo sobre IA eleitoral, mas com lacunas operacionais que a análise crítica honesta não pode minimizar. A Resolução TSE nº 23.732/2024 é um avanço genuíno e pioneiro. Ela não é suficiente, por si mesma, para proteger a integridade eleitoral de 2026 contra todos os riscos documentados nesta obra, e será ainda menos suficiente para 2030 sem uma revisão substantiva que a maioria dos atores institucionais ainda não começou a planejar. O reconhecimento simultâneo do avanço e da insuficiência é a única postura analítica que produz diagnóstico útil.

A segunda proposição subsidiária é que a responsabilização jurídica efetiva pelo uso irregular de IA eleitoral depende de três condições que a análise dos capítulos de direito eleitoral revelou como parcialmente não satisfeitas: tipificação adequada das condutas novas habilitadas pela IA, atribuição de responsabilidade em cadeias de produção deliberadamente opacas, e capacidade de detecção e prova em escala industrial. A criação de tipos infracionais administrativos pela Resolução TSE foi a resposta mais eficaz disponível para o problema da tipificação. O modelo de responsabilidade solidária em

cadeia é a resposta mais promissora para a atribuição. E o desenvolvimento de sistemas automatizados de detecção que operem na mesma escala que as violações é a única resposta estruturalmente adequada para o problema da prova em escala.

A terceira proposição subsidiária é que a proteção de dados pessoais dos eleitores não é uma obrigação paralela às obrigações eleitorais: é uma condição necessária para que o processo eleitoral cumpra sua função democrática de permitir que cidadãos formem preferências políticas autênticas. Essa articulação entre LGPD e direito eleitoral, desenvolvida no Capítulo 13, é a que tem maior potencial de transformação prática para campanhas e para a Justiça Eleitoral: ela transforma a conformidade com a LGPD de exercício de compliance em instrumento de proteção democrática, com a força normativa correspondente.

A quarta proposição subsidiária é que nenhum dos instrumentos jurídicos, regulatórios ou tecnológicos analisados nesta obra é suficiente isoladamente, e que sua efetividade depende da combinação com investimento em literacia midiática que desenvolve nos eleitores a resiliência que nenhuma regulação consegue produzir por substituição. A camada de amplificação orgânica da desinformação, composta por cidadãos que redistribuem conteúdo de boa-fé, só pode ser reduzida por capacidade crítica desenvolvida em escala populacional. E essa capacidade é o único instrumento que protege a democracia aonde a regulação não chega e onde a detecção tecnológica falha.

As Condições da Democracia Resiliente

A pergunta sobre se é possível salvar a democracia tem resposta condicional que depende de cinco condições que esta obra identificou como necessárias, nenhuma das quais é suficiente isoladamente e todas as quais precisam ser desenvolvidas simultaneamente para produzir o nível de resiliência que o ambiente tecnológico de 2026 exige.

A primeira condição é a regulação adaptativa: marcos normativos que não apenas respondem a tecnologias já implantadas mas antecipam trajetórias de desenvolvimento com grau suficiente de precisão para criar obrigações preventivas. A Resolução TSE nº 23.732/2024 foi um passo relevante nessa direção, mas o mecanismo de revisão periódica obrigatória que tornaria a adaptação regulatória sistemática em vez de reativa ainda não foi institucionalizado. Sem esse mecanismo, a regulação chegará sempre um passo atrás da tecnologia que busca governar, e a janela entre a inovação tecnológica e a resposta regulatória será a janela explorada por atores com motivação e recursos para fazê-lo.

A segunda condição é a capacidade institucional de fiscalização em escala: sistemas automatizados de detecção de violações que operem na mesma escala em que as violações ocorrem. A regulação que depende exclusivamente de denúncias formais para identificar violações é uma regulação que protege melhor os candidatos com recursos para monitorar e denunciar do que os sem esses recursos. O desenvolvimento de capacidade técnica de fiscalização pelo TSE e pelo CIEDDE é uma prioridade que precisa ser financiada antes de 2028, não improvisada em resposta às primeiras violações em escala das eleições de 2030.

A terceira condição é a cooperação internacional efetiva: acordos entre democracias para compartilhar inteligência sobre operações de desinformação, coordenar respostas regulatórias e exercer pressão conjunta sobre desenvolvedores de tecnologia localizados fora de qualquer jurisdição nacional individual. O Brasil tem posição privilegiada para liderar essa agenda na América Latina, exatamente como liderou a agenda de regulação específica de IA eleitoral com a Resolução TSE nº 23.732/2024. A construção de uma rede de democracias com marcos regulatórios de IA eleitoral mutuamente compatíveis seria o contributo mais significativo que o Brasil poderia fazer para a integridade democrática global no próximo ciclo.

A quarta condição é o investimento em literacia midiática com escala e continuidade comparáveis ao que a Finlândia desenvolveu ao longo de três décadas. Os retornos desse investimento são de horizonte geracional, não eleitoral, o que os torna sistematicamente subinvestidos em processos orçamentários de curto prazo. A criação de mecanismos que protejam esse investimento dos ciclos políticos, como fundos específicos com mandato de longo prazo ou metas constitucionais de educação midiática, é a condição estrutural que permite que os outros instrumentos operem sobre um eleitorado com capacidade crítica crescente em vez de decrescente.

A quinta condição é a mais fundamental e a que esta obra pode apenas nomear mas não garantir: a vontade coletiva de preservar as condições de possibilidade da democracia mesmo quando isso limita vantagens táticas de curto prazo. Como argumentou Sandel (2012), a democracia pressupõe cidadãos dispostos a praticar as virtudes cívicas que ela requer, incluindo a disposição de submeter interesses particulares a normas comuns que protegem o processo pelo qual todos competem. Um sistema eleitoral em que os principais beneficiários potenciais da desinformação controlam os instrumentos de fiscalização ou têm incentivos para subvertê-los não pode ser protegido apenas por instrumentos jurídicos: depende de um conjunto mínimo de compromissos normativos compartilhados que as

instituições sustentam mas que os cidadãos precisam reproduzir continuamente (SANDEL, 2012, p. 221-226).

Uma Palavra Final aos Leitores Desta Obra

Este livro foi escrito com a convicção de que informação de qualidade sobre a interseção entre inteligência artificial e democracia eleitoral é, ela própria, um instrumento de proteção democrática. Não porque informação seja suficiente, os capítulos anteriores demonstraram com abundância que informação sem regulação, sem instituições e sem vontade política é insuficiente. Mas porque sem informação adequada, regulação, instituições e vontade política carecem da base de legitimidade que as torna eficazes no longo prazo.

Os dezessete capítulos anteriores construíram esse fundamento de informação para três públicos que a obra buscou alcançar simultaneamente. Para profissionais de campanha e consultores de marketing político, construiu o mapa de riscos e oportunidades que permite operar com mais eficácia dentro dos limites que protegem o processo eleitoral. Para advogados eleitorais, construiu o arcabouço doutrinário e jurisprudencial que permite orientar clientes e litigar casos com fundamento adequado à complexidade tecnológica que o campo agora exige. E para cidadãos interessados na qualidade da democracia brasileira, construiu o enquadramento conceitual que permite compreender o que está em jogo nas escolhas tecnológicas e regulatórias que determinarão o tipo de democracia que o Brasil terá em 2036.

A democracia que o Brasil herdou de gerações que lutaram por ela sob condições incomparavelmente mais difíceis do que as que este livro descreve merece o mesmo nível de comprometimento ativo. As ameaças que a IA eleitoral apresenta são reais, documentadas e crescentes. Os instrumentos para enfrentá-las existem, são conhecidos e são implementáveis. O que faz a diferença, em última instância, é se os atores com poder de implementação, incluindo os leitores desta obra, escolhem agir com o conhecimento disponível ou aguardar a urgência que chegará mais cedo do que o calendário eleitoral sugere.

A pergunta que intitula este capítulo tem resposta. Mas a resposta não está neste livro: está nas escolhas que serão feitas por quem o leu.

O que Aprendemos

Ao longo dos dezoito capítulos desta obra, construímos progressivamente uma compreensão integrada do impacto da inteligência artificial sobre o processo democrático eleitoral que nenhum capítulo individualmente poderia fornecer. O valor dessa construção cumulativa não está apenas no conhecimento específico de cada capítulo, mas na teia de conexões entre eles: a regulação ganha sentido quando articulada com a análise técnica; a análise técnica ganha urgência quando articulada com a jurisprudência; a jurisprudência ganha contexto quando articulada com os casos concretos; e os casos concretos ganham profundidade quando situados nos fundamentos democráticos que justificam toda a empresa regulatória.

É útil, neste capítulo de fechamento, sintetizar os principais aprendizados de cada parte antes de avançar para as propostas, não como catálogo de conteúdos mas como mapeamento das linhas argumentativas que esta obra procurou desenvolver com consistência do primeiro ao último capítulo.

Sobre a natureza do problema.

Os capítulos iniciais estabeleceram que a inteligência artificial não é uma ameaça à democracia eleitoral no sentido de uma força externa que ataca um sistema sadio: ela é uma tecnologia que amplifica as vulnerabilidades que os sistemas democráticos já continham e que os instrumentos regulatórios disponíveis foram construídos para gerir em escalas muito menores. A propaganda política existe desde que a política existe. A manipulação de informação existe desde que a informação é produzida em escala. O que a IA generativa introduz é uma descontinuidade quantitativa que se converte em descontinuidade qualitativa: quando qualquer ator com acesso à internet consegue produzir conteúdo sintético convincente em escala industrial, com personalização psicológica e distribuição automatizada, o problema muda de natureza, não apenas de magnitude.

Sobre os mecanismos técnicos.

A análise das tecnologias subjacentes à desinformação eleitoral, GANs, modelos de difusão, síntese de voz, modelos de linguagem de grande escala e sistemas de microtargeting, revelou um padrão recorrente que o profissional do direito eleitoral precisa internalizar: cada avanço tecnológico que torna o conteúdo sintético mais convincente é acompanhado por um avanço que torna sua detecção mais difícil. A corrida armamentista entre geração e detecção não é simétrica: os sistemas de geração têm incentivos, recursos

e velocidade de inovação sistematicamente superiores aos dos sistemas de detecção. Essa assimetria não é superável por investimento adicional em detecção: exige a mudança de paradigma de controle de conteúdo para autenticação de proveniência que o padrão C2PA representa.

Sobre os casos brasileiros.

A análise dos três ciclos eleitorais brasileiros recentes revelou um padrão que se repete com regularidade suficiente para ser tratado como lei empírica do campo: a natureza da lacuna regulatória muda a cada ciclo, mas sua existência se mantém constante. Em 2018, a lacuna era de marco normativo. Em 2022, era de definição jurídica precisa. Em 2024, era de detecção sistemática em escala. Em 2026, será de protocolo ANPD-TSE para dados inferidos e de fiscalização de agentes autônomos. A conclusão prática dessa regularidade é que a pergunta relevante para cada ciclo não é se haverá lacunas, mas de qual natureza serão e se o sistema estará mais bem preparado para identificá-las antes que causem dano irreversível.

Sobre o marco regulatório.

A Resolução TSE nº 23.732/2024 foi analisada com o rigor que seu caráter pioneiro exige: como avanço genuíno e como instrumento com limitações estruturais identificáveis. Seu principal mérito é a especificidade: ela é o instrumento normativo mais detalhado sobre IA eleitoral disponível em qualquer democracia ao tempo de sua edição, com definições operacionais, vedações absolutas e regime sancionatório próprio. Suas principais lacunas são a indefinição do limiar técnico de "significativamente modificado", a ausência de mecanismo de certificação de ferramentas, a dependência de detecção por denúncia e a ausência de revisão periódica obrigatória. A Resolução TSE nº 23.760/2026, ao integrar a regulação de IA ao calendário eleitoral com a vedação dos três dias pré-eleitorais, representou o amadurecimento do sistema. Mas a revisão de segunda geração que as trajetórias tecnológicas do Capítulo 17 exigem ainda não foi iniciada.

Sobre a responsabilização.

A análise do sistema de responsabilização revelou que sua efetividade depende da articulação das três camadas, administrativa, cível e penal, de forma que cada uma

contribua com o que oferece de melhor: a camada administrativa com sua agilidade e seu padrão probatório mais acessível, a camada cível com seu potencial de consequências definitivas sobre o resultado eleitoral, e a camada penal com sua força simbólica e deterrente de longo prazo. O maior obstáculo identificado não é a ausência de instrumentos, mas a dissociação deliberada entre financiadores, produtores e distribuidores que atores sofisticados constroem para criar negação plausível em cada elo da cadeia. A resposta do direito eleitoral para esse obstáculo é o critério do benefício eleitoral como substituto da prova de autoria, consolidado pela jurisprudência do TSE mas ainda em construção para o contexto específico da IA.

Sobre a proteção de dados.

A articulação entre LGPD e direito eleitoral, um campo em que a produção doutrinária permanece desproporcionalmente escassa em relação à importância do tema, revelou que a proteção de dados não é obrigação paralela às obrigações eleitorais: é condição necessária para que o processo eleitoral cumpra sua função democrática. A questão mais urgente identificada, o regime jurídico dos dados inferidos que revelam orientação política a partir de dados comportamentais aparentemente neutros, ainda não tem resposta consolidada na jurisprudência da ANPD ou do TSE, e a zona de risco que essa ausência cria é precisamente onde sistemas de microtargeting psicográfico operam com maior frequência.

Sobre as ferramentas.

A análise das ferramentas digitais para proteção da integridade eleitoral estabeleceu o princípio que orienta todo o Capítulo 14: a pergunta mais importante não é qual ferramenta usar, mas como interpretar seus resultados e quais são seus limites de confiabilidade. Ferramentas de detecção são instrumentos de triagem, não de prova. A prova requer análise forense com cadeia de custódia. O protocolo de resposta a incidentes de deepfake eleitoral, com suas cinco fases sobrepostas e seus responsáveis específicos pré-definidos, é o instrumento mais operacionalmente relevante que esta obra oferece para profissionais de campanha, precisamente porque converte conhecimento em procedimento executável sob pressão máxima.

Sobre a prática.

Os instrumentos operacionais do Capítulo 15, o checklist de conformidade, o guia de usos lícitos, a matriz de riscos, as cláusulas contratuais e o papel do DPO, convergiram para um argumento que precisa ser repetido na conclusão: a arquitetura de conformidade só funciona quando implementada antes do período eleitoral e de forma integrada. Cada instrumento isolado é necessário mas insuficiente. Sua força está na combinação, e a combinação tem de estar em funcionamento antes que o primeiro incidente a torne urgente.

Sobre a literacia midiática.

A análise do déficit estrutural de literacia midiática no Brasil e das experiências internacionais de referência revelou que o investimento em capacidade crítica do eleitorado é o instrumento com maior retorno de longo prazo disponível para a democracia brasileira, e simultaneamente o mais sistematicamente subinvestidos por ser de horizonte geracional. Os cinco princípios transferíveis das experiências finlandesa, estoniana e taiwanesa, integração curricular transversal, formação de professores como condição necessária, progressão de complexidade, parceria com plataformas e enquadramento institucional adequado, fornecem o mapa de implementação que políticas públicas de literacia midiática no Brasil precisam seguir para produzir impacto antes de 2030.

Sobre o futuro

O Capítulo 17 estabeleceu o argumento mais perturbador desta obra: o direito eleitoral da IA de próxima geração não será construído por aperfeiçoamento incremental do marco regulatório vigente, mas por uma revisão de segunda geração que ainda não foi iniciada por nenhuma democracia do mundo. As cinco capacidades emergentes analisadas, síntese de vídeo em tempo real, clonagem de voz com poucos segundos de referência, agentes autônomos, personalização generativa em escala e deepfakes de imagem em tempo real, atingirão disponibilidade em dispositivos comuns entre 2027 e 2030. O próximo ciclo eleitoral brasileiro de 2030 ocorrerá num ambiente tecnológico radicalmente diferente do de 2026. A preparação para esse ciclo precisa começar agora.

Propostas de aperfeiçoamento normativo

Com base na análise desenvolvida ao longo desta obra, apresentamos um conjunto de propostas concretas de aperfeiçoamento do marco regulatório da IA eleitoral no Brasil,

organizadas por prazo de implementação e por ator responsável, com realismo sobre as possibilidades institucionais de cada esfera e sobre os obstáculos que cada proposta precisará superar para sair do papel.

A organização por prazo não é apenas editorial: ela reflete uma análise sobre o que é institucional e politicamente viável em cada horizonte temporal. Propostas de curto prazo, até as eleições de 2026, precisam ser implementáveis sem aprovação legislativa, porque o calendário não permite percorrer o processo legislativo completo antes do pleito. Propostas de médio prazo, no ciclo 2026-2030, dependem da janela legislativa que se abre após as eleições gerais e que é o momento mais produtivo do ciclo para reformas regulatórias estruturais. E propostas de longo prazo envolvem construção institucional que não tem prazo eleitoral mas que precisa ser iniciada agora para produzir resultados no horizonte de uma década.

Propostas de Curto Prazo — Até as Eleições de 2026

A primeira proposta é a criação de um protocolo formal de cooperação entre o TSE e a ANPD, que estabeleça fluxos de comunicação, divisão de competências e procedimentos de investigação conjunta para casos em que o uso irregular de dados pessoais em campanhas eleitorais constitua simultaneamente infração à LGPD e infração eleitoral. Como analisado no Capítulo 13, a ausência desse protocolo é a principal lacuna operacional do sistema de proteção de dados em contexto eleitoral, e suas consequências práticas incluem investigações paralelas com resultados inconsistentes, duplo ônus processual para investigados e fragmentação probatória que beneficia operações deliberadamente estruturadas para explorar a fronteira entre as duas jurisdições. A implementação dessa proposta não exige aprovação legislativa: pode ser feita por portaria conjunta TSE-ANPD com criação de grupo de trabalho interinstitucional permanente. O obstáculo não é jurídico: é a construção da cultura de cooperação entre instituições com tradições regulatórias distintas e que historicamente operaram em campos sem sobreposição. Essa cultura não se constrói por portaria: exige investimento deliberado em relações institucionais que precede a cooperação operacional em casos concretos.

A segunda proposta é a adoção pelo TSE de padrões técnicos de autenticidade de conteúdo baseados no protocolo C2PA, que insere assinaturas criptográficas nos arquivos de mídia no momento de sua captura, permitindo a verificação posterior de autenticidade sem dependência de sistemas de detecção que a corrida armamentista tecnológica progressivamente torna insuficientes. Como argumentado no Capítulo 14, a resposta mais

promissora ao problema da pós-autenticidade não está na competição entre sistemas de geração e detecção, que os sistemas de geração tendem a vencer, mas na mudança de paradigma de controle de conteúdo para autenticação de proveniência. A integração dessa tecnologia ao ecossistema eleitoral brasileiro, incluindo a exigência de uso de câmeras e sistemas de transmissão certificados C2PA para transmissões oficiais de candidatos em debates e entrevistas, é a resposta técnica mais robusta disponível ao problema da pós-autenticidade. A implementação requer resolução do TSE e acordos de cooperação com fabricantes de equipamentos e com plataformas de distribuição de conteúdo, o que é factível no horizonte de 2026 se iniciado imediatamente.

A terceira proposta é a extensão obrigatória do Programa Fato ou Boato do TSE ao período de pré-campanha, com funcionamento permanente ao longo de todo o ano eleitoral e não apenas nos períodos de campanha oficial, e a ampliação de sua capacidade de processamento para dar conta do volume crescente de conteúdo suspeito que circula no ambiente digital brasileiro. Como demonstrado no Capítulo 11, a extensão das obrigações regulatórias ao período de pré-campanha por critério funcional e não temporal torna incoerente manter as ferramentas de verificação restritas ao período oficial de campanha: se a desinformação eleitoral opera o ano inteiro, as ferramentas de defesa precisam operar o ano inteiro. A ampliação do programa exige aumento de orçamento do CIEDDE e das agências de fact-checking parceiras, além de integração com plataformas de mensageria que hoje não cooperam adequadamente com os sistemas de verificação.

Propostas de Médio Prazo — Ciclo Legislativo 2026-2030

A quarta proposta é a aprovação do PL 2.338/2023 com inclusão de dispositivos específicos para o contexto eleitoral que vão além do texto atual, endereçando três lacunas identificadas ao longo desta obra: a obrigação de auditoria independente de sistemas de recomendação de plataformas digitais em períodos eleitorais, que é o instrumento mais eficaz para responsabilizar algoritmos de amplificação que operam como vetores de desinformação sem produzir conteúdo irregular eles próprios; a proibição expressa de microtargeting baseado em dados psicográficos inferidos sem consentimento específico em contexto político, que endereça o problema da zona de indeterminação sobre dados sensíveis inferidos identificado no Capítulo 13; e a criação de fundo público para financiamento de ferramentas de detecção e verificação de conteúdo sintético eleitoral, que reduz a dependência de ferramentas estrangeiras para proteção de processo eleitoral que é por definição soberano. A aprovação do PL com essas emendas exige articulação entre

TSE e Congresso que precisa ser construída no primeiro ano do novo mandato presidencial.

A quinta proposta é a criação de um regime de responsabilidade civil objetiva para plataformas digitais por danos eleitorais decorrentes de conteúdo irregular que permaneça disponível após notificação formal da Justiça Eleitoral, elevando a força normativa do prazo de vinte e quatro horas e da responsabilidade solidária estabelecidos pela Resolução TSE nº 23.732/2024 ao nível de lei ordinária. Como analisado no Capítulo 8, as plataformas têm incentivos estruturais que não são neutralizados pela existência de uma resolução administrativa cujo regime de enforcement depende de acordos de cooperação voluntários: a transformação desse regime em lei ordinária com responsabilidade objetiva cria os incentivos financeiros que a conformidade voluntária não produz por si mesma. A proposta exige lei complementando o Marco Civil da Internet e enfrentará resistência das plataformas, que têm histórico de intenso lobbying contra a expansão de sua responsabilidade por conteúdo de terceiros.

A sexta proposta é a inclusão da educação para a democracia digital e literacia midiática como componente curricular obrigatório na educação básica, com progressão de complexidade ao longo dos anos escolares, seguindo os cinco princípios transferíveis das experiências internacionais de referência identificados no Capítulo 16: integração transversal nas disciplinas existentes, formação específica de professores como condição necessária, progressão de complexidade garantida, parceria com plataformas de fact-checking e enquadramento como política de Estado e não de governo. O retorno dessa proposta é de horizonte geracional, o que a torna sistematicamente desfavorecida em disputas orçamentárias de curto prazo. A criação de mecanismos de proteção do investimento, como fundo específico com mandato de longo prazo ou meta constitucional de educação midiática, é a condição estrutural que distingue políticas que produzem impacto das que são descontinuadas no primeiro ciclo de corte orçamentário.

Propostas de Longo Prazo — Agenda para a Próxima Década

A sétima proposta é a construção de um regime internacional de governança da IA eleitoral no âmbito da Organização dos Estados Americanos e das Nações Unidas, que estabeleça padrões mínimos de transparência, responsabilização e proteção da integridade eleitoral aplicáveis globalmente. Como argumentado no Capítulo 17, a natureza transnacional das plataformas digitais e dos sistemas de IA eleitoral torna ineficaz qualquer regulação que opere exclusivamente no âmbito nacional: a Resolução TSE nº 23.732/2024

regula o uso de IA eleitoral no Brasil mas não tem mecanismo de influência sobre o desenvolvimento dos sistemas que essa regulação busca governar. O Brasil tem posição privilegiada para liderar essa agenda na América Latina, exatamente como liderou a agenda de regulação específica de IA eleitoral no plano doméstico: um país que pioneirou o marco regulatório mais específico do mundo tem credibilidade para propor que esse marco se torne referência para acordos multilaterais. A construção desse regime exige engajamento do Executivo federal, do Ministério das Relações Exteriores e do TSE em foros multilaterais, com horizonte de negociação de pelo menos uma década.

A oitava proposta é a criação de um laboratório nacional de pesquisa em integridade eleitoral digital, vinculado ao TSE com autonomia científica garantida por modelo institucional adequado, dedicado ao desenvolvimento e à manutenção de ferramentas de detecção de conteúdo sintético, monitoramento de operações de desinformação, análise forense eleitoral e formação de pesquisadores e profissionais no campo. O laboratório funcionaria como centro de excelência que reduz progressivamente a dependência de ferramentas estrangeiras para a proteção de um processo eleitoral que é, por definição, expressão da soberania nacional. Sua autonomia científica é condição de credibilidade: um laboratório percebido como instrumento de qualquer governo terá seus resultados questionados precisamente nos casos em que sua independência seria mais necessária. A criação exige lei específica, dotação orçamentária permanente e modelo de governança que inclua universidades parceiras e representação da sociedade civil.

O Papel de Cada Ator: Uma Agenda Distribuída

A proteção da integridade democrática na era da IA não é responsabilidade de nenhum ator isolado. É, por natureza, uma responsabilidade distribuída que exige a contribuição coordenada de atores com papéis, recursos e capacidades distintos e que não pode ser delegada inteiramente a nenhum deles. Quando candidatos terceirizam a responsabilidade para plataformas, plataformas a devolvem para a regulação, a regulação a projeta sobre as campanhas e as campanhas a atribuem aos adversários que praticam desinformação, o resultado é a paralisia coletiva que beneficia precisamente os atores que preferem ausência de responsabilização a qualquer distribuição que os inclua.

O quadro a seguir organiza o que cada categoria de ator pode fazer agora, o que deve fazer no médio prazo e o que não deve fazer nunca, não como lista de obrigações formais mas como diagnóstico de onde cada ator tem maior impacto potencial e maiores riscos de contribuir para a deterioração que todos dizem querer evitar. A coluna do que não deve fazer nunca é intencionalmente a mais direta: ela enumera as condutas que,

independentemente de qualquer discussão sobre gradações de culpabilidade e estratégias de minimização de responsabilidade, são incompatíveis com o compromisso mínimo com a integridade democrática que qualquer ator que pretende participar do processo eleitoral precisa sustentar.

Tabela 58. Agenda de Ação por Categoria de Ator

Ator	O que pode fazer agora	O que deve fazer no médio prazo	O que não deve fazer nunca
Candidatos e campanhas	Implementar checklist de conformidade com Res. TSE 23.732/2024; inserir cláusulas de conformidade em contratos com fornecedores de tecnologia; nomear DPO.	Investir em ferramentas de monitoramento de desinformação; treinar equipe em literacia midiática; adotar padrões C2PA para conteúdo oficial.	Contratar agências de disparos em massa não declarados; utilizar deepfakes de adversários; microtargeting com dados sensíveis sem consentimento.
Advogados eleitorais	Incluir análise de conformidade com IA na due diligence eleitoral; orientar clientes sobre riscos de responsabilidade solidária; acompanhar jurisprudência do TSE.	Desenvolver expertise em intersecção entre LGPD e direito eleitoral; construir conhecimento técnico básico sobre ferramentas de IA.	Assessorar clientes no uso de tecnologias claramente vedadas pela regulação; omitir riscos jurídicos relevantes.
Plataformas digitais	Implementar a Resolução TSE 23.732/2024 com rigor; responder em 24 horas às notificações; disponibilizar dados para pesquisadores credenciados.	Desenvolver ferramentas de detecção de deepfakes integradas às plataformas; implementar padrões C2PA; transparência algorítmica.	Monetizar conteúdo eleitoral irregular; resistir a ordens judiciais de remoção; tratar compliance eleitoral como custo a minimizar.
Pesquisadores e academia	Desenvolver e publicar ferramentas open source de detecção; produzir análises empíricas do ecossistema de desinformação eleitoral; formar pesquisadores no campo.	Construir bases de dados de referência para detecção de deepfakes em português; colaborar com o TSE em desenvolvimento de capacidades.	Produzir conhecimento que facilite a desinformação; omitir conflitos de interesse em pesquisas financiadas por plataformas.

Cidadãos e eleitores	Verificar conteúdo antes de compartilhar; usar ferramentas de fact-checking; reportar conteúdo suspeito ao TSE pelo canal Fato ou Boato.	Engajar em programas de literacia midiática; exigir transparência de candidatos sobre o uso de IA em campanhas.	Compartilhar conteúdo desinformativo mesmo que corresponda às suas preferências políticas; tratar deepfakes como evidência sem verificação.
Instituições eleitorais (TSE/TREs)	Aplicar a Resolução TSE 23.732/2024 com consistência; operacionalizar o CIEDDE; firmar protocolo com ANPD.	Adotar padrões C2PA; criar laboratório de forense digital eleitoral; ampliar programas de literacia midiática.	Usar tecnologias de vigilância digital de eleitores sem base legal; limitar o acesso de pesquisadores a dados eleitorais necessários à fiscalização.

Fonte: Elaborado pelo autor, 2026.

A Democracia como Prática Cotidiana: Uma Nota Final

Este livro começou com uma pergunta: quem realmente decide as eleições na era da inteligência artificial? Percorremos dezoito capítulos para construir uma resposta rigorosa, multidimensional e, nos momentos em que a análise honesta exigiu, desconfortável. É justo, neste último momento, oferecer também uma resposta simples. Quem decide, ainda, é o cidadão.

O voto, no momento em que é depositado na urna eletrônica, é o ato soberano de um ser humano que, apesar de todas as tentativas de manipulação, de todas as bolhas de filtro e câmaras de eco, de todos os deepfakes e disparos em massa, chegou àquela cabine eleitoral com uma escolha. Essa escolha pode ter sido influenciada de formas que o próprio eleitor não percebeu. Mas ela foi feita.

A questão não é se a IA decide as eleições: ela não decide, pelo menos não diretamente. A questão é se as condições em que os eleitores formam suas preferências são suficientemente livres, informadas e autônomas para que o resultado eleitoral possa ser chamado, com honestidade, de expressão da vontade popular. É essa pergunta que cada ferramenta regulatória, cada protocolo de conformidade e cada aula de literacia midiática descritos nesta obra procura responder positivamente. A democracia não sobrevive pela força de seus instrumentos jurídicos, por mais bem desenhados que sejam. Ela sobrevive pela qualidade dos cidadãos que a habitam.

Tocqueville (2005), ao analisar a democracia americana no século XIX com a perspicácia de quem observava de fora um experimento político sem precedentes históricos, identificou que a condição mais fundamental da democracia não é nem a Constituição nem as leis nem as instituições: é o estado de espírito dos cidadãos que as sustentam. Uma democracia habitada por cidadãos que tratam a política como espetáculo em que são espectadores passivos, que acreditam que as decisões relevantes são tomadas por outros e que a participação política individual não tem consequências mensuráveis, é uma democracia que já perdeu o principal recurso de que depende para sobreviver: a energia cívica dos próprios cidadãos que ela é suposta a servir (TOCQUEVILLE, 2005, vol. 2, p. 389-395).

A desinformação eleitoral, com toda sua sofisticação tecnológica, opera sobre essa vulnerabilidade existencial da democracia: ela não apenas distribui informação falsa, ela distribui desorientação e ceticismo que convencem cidadãos de que a participação genuína é impossível porque tudo é manipulado, todos os atores são igualmente desonestos e a distinção entre verdade e mentira é operacionalmente irrelevante. Esse cinismo não é apenas consequência da desinformação: é seu produto mais valoroso para quem a instrumentaliza, porque cidadãos cínicos se abstêm, e abstenção favorece estruturalmente os atores com maior capacidade de mobilizar bases menores de apoio intenso.

O voto programado, manipulado, condicionado por algoritmos que conhecem nossas vulnerabilidades melhor do que nós mesmos, é a antítese da democracia. O voto livre, consciente, informado, resultado de uma deliberação genuína por um cidadão que conhece seus direitos e sabe reconhecer quando está sendo manipulado, é o único voto que merece o nome de democrático. E a distância entre esses dois tipos de voto é precisamente o que esta obra procurou, em cada um de seus dezoito capítulos, contribuir para reduzir.

A democracia não é um sistema que funciona automaticamente. É uma conquista que precisa ser renovada a cada geração, com as ferramentas e contra os inimigos de cada época. Os instrumentos descritos nesta obra, regulatórios, jurídicos, tecnológicos e educacionais, são as ferramentas desta época. Os inimigos desta época têm nomes: desinformação, manipulação algorítmica, deepfakes e, mais difícil de combater do que todos eles, a indiferença de cidadãos que acreditam que a democracia vai se defender sozinha.

Não vai. E protegê-la é, no final, a tarefa de todos nós!

REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, Érika; OLIVEIRA, Marcos. Desinformação e eleições no Brasil: ecossistema, padrões e desafios regulatórios. Rio de Janeiro: NetLab/UFRJ, 2021.

BALKIN, Jack M. Free speech is a triangle. Columbia Law Review, v. 118, n. 7, p. 2011–2056, 2018.

BENKLER, Yochai; FARIS, Robert; ROBERTS, Hal. Network Propaganda: manipulation, disinformation, and radicalization in American politics. Oxford: Oxford University Press, 2018.

BENKLER, Yochai. The Wealth of Networks: how social production transforms markets and freedom. New Haven: Yale University Press, 2006.

BESSI, Alessandro; FERRARA, Emilio. Social bots distort the 2016 U.S. Presidential election online discussion. First Monday, v. 21, n. 11, 2016.

BIONI, Bruno Ricardo. Proteção de dados pessoais: a função e os limites do consentimento. 2. ed. Rio de Janeiro: Forense, 2021.

BOBBIO, Norberto. O futuro da democracia: uma defesa das regras do jogo. Rio de Janeiro: Paz e Terra, 1986.

BOND, Robert M. et al. A 61-million-person experiment in social influence and political mobilization. Nature, v. 489, p. 295–298, 2012.

BRASIL. Código Eleitoral. Lei nº 4.737, de 15 de julho de 1965. Institui o Código Eleitoral. Diário Oficial da União, Brasília, DF, 19 jul. 1965.

BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Senado Federal, 1988.

BRASIL. Lei Complementar nº 64, de 18 de maio de 1990. Estabelece casos de inelegibilidade. Diário Oficial da União, Brasília, DF, 21 maio 1990.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil (Marco Civil da Internet). Diário Oficial da União, Brasília, DF, 24 abr. 2014.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União, Brasília, DF, 15 ago. 2018.

BRASIL. Lei nº 14.197, de 1º de setembro de 2021. Acrescenta ao Código Penal os crimes contra o Estado Democrático de Direito. Diário Oficial da União, Brasília, DF, 2 set. 2021.

BRASIL. Lei nº 9.096, de 19 de setembro de 1995. Dispõe sobre partidos políticos. Diário Oficial da União, Brasília, DF, 20 set. 1995.

BRASIL. Lei nº 9.504, de 30 de setembro de 1997. Estabelece normas para as eleições. Diário Oficial da União, Brasília, DF, 1 out. 1997.

BRASIL. Projeto de Lei nº 2.338, de 2023. Dispõe sobre o uso da Inteligência Artificial. Brasília, DF: Senado Federal, 2023.

BRASIL. SUPREMO TRIBUNAL FEDERAL. Recurso Extraordinário nº 1.037.396/SP. Relator: Min. Dias Toffoli. Tema 987: Responsabilidade civil de provedor de aplicações de internet por conteúdo gerado por terceiros. Brasília, DF: STF, 2023.

BRASIL. TRIBUNAL SUPERIOR ELEITORAL. Apostila Curso IA e Eleições: módulo 1 — fundamentos jurídicos e regulatórios. Brasília, DF: TSE/CIEDDE, 2026.

BRASIL. TRIBUNAL SUPERIOR ELEITORAL. Apostila Curso IA e Eleições: módulo 2 — deepfakes, desinformação e conteúdo sintético. Brasília, DF: TSE/CIEDDE, 2026.

BRASIL. TRIBUNAL SUPERIOR ELEITORAL. Portaria TSE nº 544, de 2022. Institui o Centro Integrado de Enfrentamento à Desinformação e Defesa da Democracia Eleitoral (CIEDDE). Brasília, DF: TSE, 2022.

BRASIL. TRIBUNAL SUPERIOR ELEITORAL. Resolução nº 23.610, de 18 de dezembro de 2019. Dispõe sobre propaganda eleitoral. Diário da Justiça Eleitoral, Brasília, DF, 2019.

BRASIL. TRIBUNAL SUPERIOR ELEITORAL. Resolução nº 23.732, de 28 de março de 2024. Dispõe sobre o uso de Inteligência Artificial nas eleições de 2024. Diário da Justiça Eleitoral, Brasília, DF, 2024.

BRASIL. TRIBUNAL SUPERIOR ELEITORAL. Resolução nº 23.760, de 2 de março de 2026. Calendário Eleitoral para as Eleições Gerais de 2026. Diário da Justiça Eleitoral, Brasília, DF, 2026.

BRASIL. TRIBUNAL SUPERIOR ELEITORAL. RQ-0060125/2022. Relator: Min. Alexandre de Moraes. Brasília, DF: TSE, 2022.

BRIGNULL, Harry. Dark patterns: deception vs. honesty in UI design. A List Apart, n. 338, 2011.

BRYNJOLFSSON, Erik; McAFEE, Andrew. The Second Machine Age: work, progress, and prosperity in a time of brilliant technologies. New York: W.W. Norton, 2014.

C2PA. C2PA Technical Specification 2.0. Coalition for Content Provenance and Authenticity, 2024. Disponível em: https://c2pa.org/specifications/specifications/2.0/. Acesso em: 20 jan. 2025.

CADWALLADR, Carole; GRAHAM-HARRISON, Emma. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian; The Observer, 17 mar. 2018. Disponível em: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. Acesso em: 10 jan. 2025.

CÂNDIDO, Joel José. Direito eleitoral brasileiro. 17. ed. Bauru: EDIPRO, 2022.

CASTELLS, Manuel. Comunicación y poder. Madrid: Alianza Editorial, 2009.

CHESNEY, Robert; CITRON, Danielle Keats. Deepfakes and the new disinformation war: the coming age of post-truth geopolitics. Foreign Affairs, v. 98, n. 1, p. 147–155, jan./fev. 2019.

CIALDINI, Robert B. Influence: science and practice. 5. ed. Boston: Pearson, 2009.

CIALDINI, Robert B. Influence: the psychology of persuasion. New York: William Morrow, 1984.

CLAYTON, Katherine et al. Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. Political Behavior, v. 42, n. 4, p. 2019–2036, 2020.

CRUICKSHANK, Iain; CARLEY, Kathleen. A comparison of misinformation and disinformation on Twitter/X after its acquisition by Elon Musk. Harvard Kennedy School Misinformation Review, 2023.

DAFOE, Allan. AI governance: a research agenda. Future of Humanity Institute, University of Oxford, 2020.

ESTADOS UNIDOS. Section 230, Communications Decency Act. 47 U.S.C. § 230. Washington, D.C., 1996.

FEDERAL TRADE COMMISSION. FTC imposes $5 billion penalty and sweeping new privacy restrictions on Facebook. Washington, D.C.: FTC, 2019. Disponível em: https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook. Acesso em: 10 jan. 2025.

FERRARA, Emilio et al. The rise of social bots. Communications of the ACM, v. 59, n. 7, p. 96–104, 2016.

FGV DAPP. Robôs, redes sociais e política no Brasil: estudo sobre interferências ilegítimas no debate público na web, riscos para a democracia e processo eleitoral de 2018. Rio de Janeiro: FGV DAPP, 2018.

FGV DAPP. Robôs, redes sociais e política no Brasil. Rio de Janeiro: FGV DAPP, 2018.

FRICKER, Miranda. Epistemic Injustice: power and the ethics of knowing. Oxford: Oxford University Press, 2007.

GOMES, José Jairo. Direito eleitoral. 19. ed. São Paulo: Atlas, 2023.

GOODFELLOW, Ian et al. Generative Adversarial Networks. In: Advances in Neural Information Processing Systems. Montreal: NeurIPS, 2014.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep Learning. Cambridge: MIT Press, 2016.

GPOPAI – GRUPO DE PESQUISA EM POLÍTICAS PÚBLICAS DE ACESSO À INFORMAÇÃO DA USP. Monitoramento de boatos e desinformação nas eleições 2018. São Paulo: USP, 2018.

HABERMAS, Jürgen. Direito e democracia: entre facticidade e validade. Rio de Janeiro: Tempo Brasileiro, 1997. 2 v.

HOBBS, Renee. Digital and Media Literacy: a plan of action. Washington, D.C.: The Aspen Institute, 2010.

HOWARD, Philip N. Lie Machines: how to save democracy from troll armies, deceitful robots, junk news operations, and political operatives. New Haven: Yale University Press, 2020.

HOWARD, Philip N.; KOLLANYI, Bence. Bots, #StrongerIn, and #Brexit: computational propaganda during the UK-EU referendum. Oxford Internet Institute, COMPROP Research Note, jun. 2016.

INTERNATIONAL FEDERATION OF ELECTION SYSTEMS (IFES). AI and elections: 2024 global review. Washington, D.C.: IFES, 2024.

KAHNEMAN, Daniel. Rápido e devagar: duas formas de pensar. Rio de Janeiro: Objetiva, 2011.

KOSINSKI, Michal; STILLWELL, David; GRAEPEL, Thore. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, v. 110, n. 15, p. 5802–5805, 2013.

LEONARDI, Marcel. Tutela e privacidade na internet. São Paulo: Saraiva, 2012.

LEVITSKY, Steven; ZIBLATT, Daniel. Como as democracias morrem. Rio de Janeiro: Zahar, 2018.

MATEJIC, Nicolas. Social Media for the Political Campaigner. Saarbrücken: Lambert Academic Publishing, 2018.

McCRAE, Robert R.; COSTA, Paul T. The five-factor theory of personality. In: JOHN, Oliver P.; ROBINS, Richard W.; PERVIN, Lawrence A. (org.). Handbook of Personality: theory and research. 3. ed. New York: Guilford Press, 2008. p. 159–181.

MINISTÉRIO PÚBLICO ELEITORAL. Relatório de investigações sobre disparos em massa pelo WhatsApp nas eleições de 2018. Brasília, DF: MPE, 2019.

NEWMAN, Nic et al. Reuters Institute Digital News Report 2023. Oxford: Reuters Institute for the Study of Journalism, 2023.

NIGHTINGALE, Sophie J.; FARID, Hany. AI-synthesized faces are indistinguishable from real faces and more trustworthy. Proceedings of the National Academy of Sciences, v. 119, n. 8, 2022.

OPEN SOCIETY INSTITUTE. Media Literacy Index 2023. Sofia: European Policies Initiative/Open Society Institute, 2023.

OVERSIGHT BOARD. Annual Report 2023. Meta Oversight Board, 2023.

OXFORD INTERNET INSTITUTE. Computational Propaganda Research Project. Oxford: OII, 2016.

PARIS, Britt; DONOVAN, Joan. Deepfakes and Cheap Fakes: the manipulation of audio and visual evidence. Data & Society Research Institute, 2019.

PARISER, Eli. O filtro invisível: o que a internet está escondendo de você. Rio de Janeiro: Zahar, 2011.

PENNYCOOK, Gordon; RAND, David G. Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, v. 116, n. 7, p. 2521–2526, 2019.

PINHEIRO, Patrícia Peck. Proteção de dados pessoais: comentários à Lei n. 13.709/2018 – LGPD. 3. ed. São Paulo: Saraiva, 2021.

POTTER, W. James. Media Literacy. 8. ed. Thousand Oaks: SAGE, 2016.

REINO UNIDO. Online Safety Act 2023. London: His Majesty's Stationery Office, 2023.

REUTERS INSTITUTE FOR THE STUDY OF JOURNALISM. Digital News Report 2023. Oxford: Reuters Institute/University of Oxford, 2023.

RUSSELL, Stuart; NORVIG, Peter. Artificial Intelligence: a modern approach. 4. ed. Hoboken: Pearson, 2022.

SALGADO, Eneida Desiree. Princípios constitucionais eleitorais. Belo Horizonte: Fórum, 2015.

SANDEL, Michael J. A justiça: o que é fazer a coisa certa. 15. ed. Rio de Janeiro: Civilização Brasileira, 2012.

SOLOVE, Daniel J. Nothing to Hide: the false tradeoff between privacy and security. New Haven: Yale University Press, 2013.

SOUZA, Fábio; NOGUEIRA, Rodrigo; LOTUFO, Roberto. BERTimbau: pretrained BERT models for Brazilian Portuguese. In: INTELLIGENT SYSTEMS (BRACIS), 9., 2020, Rio Grande. Anais [...]. Cham: Springer, 2020. p. 403–417.

SUNSTEIN, Cass R. #Republic: divided democracy in the age of social media. Princeton: Princeton University Press, 2017.

SUNSTEIN, Cass R. Republic.com 2.0. Princeton: Princeton University Press, 2007.

TANG, Audrey. Digital democracy with Taiwanese characteristics. New York: Columbia Global Policy Initiative, 2020.

THALER, Richard H.; SUNSTEIN, Cass R. Nudge: improving decisions about health, wealth, and happiness. New Haven: Yale University Press, 2008.

TOCQUEVILLE, Alexis de. A democracia na América. 2. ed. São Paulo: Martins Fontes, 2005. 2 v.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016. Regulamento Geral sobre a Proteção de Dados (RGPD). Jornal Oficial da União Europeia, Bruxelas, 4 maio 2016.

UNIÃO EUROPEIA. Regulamento (UE) 2022/2065 do Parlamento Europeu e do Conselho, de 19 de outubro de 2022. Digital Services Act (DSA). Jornal Oficial da União Europeia, Bruxelas, 27 out. 2022.

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024. Lei da Inteligência Artificial (AI Act). Jornal Oficial da União Europeia, Bruxelas, 12 jul. 2024.

VAROL, Onur et al. Online human-bot interactions: detection, estimation, and characterization. In: AAAI INTERNATIONAL CONFERENCE ON WEB AND SOCIAL MEDIA, 11., 2017, Montreal. Proceedings [...]. Palo Alto: AAAI Press, 2017.

VASWANI, Ashish et al. Attention is all you need. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 30., 2017, Long Beach. Proceedings [...]. New York: Curran Associates, 2017.

VOSOUGHI, Soroush; ROY, Deb; ARAL, Sinan. The spread of true and false news online. Science, v. 359, n. 6380, p. 1146–1151, mar. 2018.

WARDLE, Claire; DERAKHSHAN, Hossein. Information Disorder: toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe, 2017.

WU, Tim. The Attention Merchants: the epic scramble to get inside our heads. New York: Knopf, 2016.

ZUBOFF, Shoshana. A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2019.

Nas campanhas eleitorais de hoje, a disputa mais decisiva não acontece nos palanques. Acontece nos algoritmos.

Democracia Hackeada é a obra que o profissional de marketing político precisava e ainda não tinha: uma análise rigorosa, tecnicamente fundamentada e diretamente aplicável sobre como a Inteligência Artificial transformou, de forma silenciosa e radical, o campo onde trabalhamos.

Deepfakes, microtargeting psicográfico, bots que simulam presença, plataformas que lucram com a polarização, financiadores invisíveis da desinformação: este livro não apenas descreve esses fenômenos. Ele os diseca, mapeia seus riscos jurídicos e entrega ao leitor as ferramentas para navegar nesse ambiente com estratégia, conformidade e vantagem competitiva.

Com base na Resolução TSE nº 23.732/2024, o marco regulatório mais avançado do mundo sobre IA eleitoral, Rodolpho Raphael traça a fronteira entre o que é permitido, o que é arriscado e o que pode custar um mandato.

Quem entende as regras do novo jogo chega na frente. Quem ignora, paga o preço nas urnas, ou na Justiça Eleitoral.

"A eleição de 2026 será vencida ou perdida no domínio que poucos ainda dominam. Este livro é o mapa."

Rodolpho Raphael é uma das vozes mais completas na interseção entre comunicação, política e tecnologia no Brasil contemporâneo. Jornalista, historiador, filósofo, teólogo, pedagogo e administrador, sua formação plural não é acidental. É o alicerce de uma visão que poucos profissionais do campo conseguem oferecer: ao mesmo tempo analítica, estratégica e humanamente enraizada. Com mais de 15 anos de atuação direta em campanhas eleitorais e na comunicação institucional de mandatos nos âmbitos legislativo e executivo, Raphael alia a experiência de quem já esteve nas trincheiras das disputas políticas à solidez de quem as estuda com método e rigor acadêmico. Professor de ensino superior há 12 anos, é o criador e fundador da primeira Pós-Graduação em Comunicação e Marketing Político da Paraíba, lançada em 2016. Uma iniciativa pioneira que formou e qualificou profissionais que hoje atuam em campanhas por todo o Brasil. Sua atuação na docência reflete uma convicção que atravessa toda a sua obra: conhecimento sem aplicação é inerte, e prática sem reflexão é cega. Democracia Hackeada é a síntese natural dessa trajetória. O livro de quem viveu as campanhas por dentro, ensinou as regras do jogo por fora e agora entrega ao leitor o que realmente está em disputa quando a Inteligência Artificial entra no campo eleitoral.

ISBN: 978-65-02-02203-0
9 786502 022030