REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

INTELIGÊNCIA ARTIFICIAL E PROPAGANDA ELEITORAL: NOVOS DESAFIOS À DEMOCRACIA INDIRETA

ARTIFICIAL INTELLIGENCE AND ELECTORAL PROPAGANDA: NEW CHALLENGES TO INDIRECT DEMOCRACY

INTELIGENCIA ARTIFICIAL Y PROPAGANDA ELECTORAL: NUEVOS DESAFÍOS PARA LA DEMOCRACIA INDIRECTA

Claudio José Langroiva Pereira
Livre Docente em Direito Processual Penal – PUC/SP,
Pós-doutor em Direito Público pelo Instituto de Altos Estudos de Pós-doutorado da Università di Bologna
Doutor em Direito das Relações Sociais– PUC/SP,
Professor Doutor de Direito Processual Penal – PUC/SP,
Líder do Grupo de Pesquisa "Políticas Públicas de Segurança e Direitos Humanos".
Advogado criminal
E-mail: claudiopereira@pucsp.br

Marcelo Carita Correra
Especialista em Direito Penal e Econômico – FGV Law
Doutor em Direito Processual Penal – PUC/SP
Mestre em Direito Penal – PUC/SP
Pós-doutor em Direito Público pelo Instituto de Altos Estudos de Pós-doutorado da Università di Bologna,
Email: mcorrera@gmail.com

Resumo

O presente artigo, elaborado pelo método lógico-dedutivo e construído a partir da literatura pátria e internacional, tem como escopo a investigação da utilização da inteligência artificial em campanhas eleitorais e como esse fenômeno opera na formação de vontade do eleitor. A pergunta de pesquisa envolve a viabilidade jurídica do uso dessa tecnologia, bem como eventuais efeitos positivos e negativos sobre o pleito eleitoral. Trata-se de estudo qualitativo, mas que se baseia em estudos empíricos realizados por outros pesquisadores sobre vieses cognitivos no processo de escolha, bem como sobre a influência da inteligência artificial, especialmente a generativa, os *chatbots* e os algoritmos em redes sociais. O primeiro capítulo foi dedicado ao estudo da premissa de que o ser humano possui livre-arbítrio para realizar suas escolhas. Após esse estudo, no segundo capítulo, o trabalho verificou os dados extraídos dos estudos de outros profissionais sobre o viés cognitivo e a influência de informações. Após a depuração dos dados, foi realizada uma pesquisa crítica dos resultados obtidos. Embora o trabalho tenha foco na legislação do Brasil, estudos relevantes utilizados neste artigo como base teórica foram realizados em instituições de ensino dos Estados Unidos da América, tendo a população e o processo eleitoral do referido país como paradigma. A conclusão foi dedicada à análise conjunta dos dados obtidos e a resposta foi que a utilização da tecnologia em questão pode implicar a manipulação da vontade do eleitor. A solução proposta foi uma ampliação da regulamentação, inclusive fora do período eleitoral, de forma a que o usuário seja informado dos riscos envolvidos na utilização da tecnologia e da falta de neutralidade plena.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

Palavras-chave: Inteligência Artificial; Livre-arbítrio; Inconsciência; Regulamentação, Deepfake.

Abstract

This article, developed through a logical-deductive approach and based on domestic and international literature, aims to analyze the use of artificial intelligence in election campaigns and how this phenomenon affects the formation of the voter's will. The research question explores the legal viability of using this technology, as well as its potential positive and negative effects on the electoral process. This is a qualitative study, yet it draws on empirical research conducted by other scholars regarding cognitive biases in the decision-making process, alongside the influence of artificial intelligence—specifically generative AI, chatbots, and social media algorithms. The first chapter is dedicated to examining the premise that human beings possess the free will to make their own choices. Following this analysis, the second chapter examines data drawn from studies by other experts on cognitive biases and the influence of information. After refining the data, a critical analysis of the results was conducted. Although this paper focuses on Brazilian legislation, relevant studies used as a theoretical foundation for this article were conducted at educational institutions in the United States of America, using the U.S. population and electoral process as a paradigm. The conclusion provides a comprehensive analysis of the collected data, establishing that the use of this technology can result in the manipulation of the voter's will. The proposed solution involves expanding regulations—even outside the election period—to ensure that users are informed about the risks associated with this technology and its lack of complete neutrality.

Keywords: Artificial Intelligence; Free Will; Unconsciousness; Regulation; Deepfake.

Resumen

El presente artículo, elaborado mediante el método lógico-deductivo y construido a partir de la literatura nacional e internacional, tiene como propósito el análisis de la utilización de la inteligencia artificial en las campañas electorales y cómo este fenómeno incide en la formación de la voluntad del elector. La pregunta de investigación plantea la viabilidad jurídica del uso de esta tecnología, así como los posibles efectos positivos y negativos sobre el proceso electoral. Se trata de una investigación cualitativa, pero que se fundamenta en estudios empíricos realizados por otros investigadores sobre los sesgos cognitivos en el proceso de elección, así como en la influencia de la inteligencia artificial, especialmente la generativa, los chatbots y los algoritmos en las redes sociales. El primer capítulo se dedicó al estudio de la premisa de que el ser humano posee libre albedrío para tomar sus decisiones. Tras este análisis, en el segundo capítulo, el trabajo examinó los datos extraídos de estudios de otros profesionales sobre el sesgo cognitivo y la influencia de la información. Luego de la depuración de los datos, se llevó a cabo un análisis crítico de los resultados obtenidos. Aunque el trabajo se centra en la legislación de Brasil, estudios relevantes utilizados en este artículo como base teórica fueron llevados a cabo en instituciones educativas de los Estados Unidos de América, tomando a la población y el proceso electoral de dicho país como paradigma. La conclusión se dedicó al análisis conjunto de los datos obtenidos y la respuesta fue que el uso de la tecnología en cuestión puede implicar la manipulación de la voluntad del elector. La solución propuesta consistió en una ampliación de la regulación, incluso fuera del período electoral, de manera que el usuario esté informado de los riesgos que conlleva el uso de la tecnología y de la falta de neutralidad plena.

Palabras clave: Inteligencia Artificial; Libre albedrío; Inconsciencia; Regulación; Deepfake.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

1. Introdução

A incorporação crescente da inteligência artificial nas dinâmicas de comunicação política constitui um dos fenômenos mais relevantes do constitucionalismo e do direito eleitoral contemporâneos. Se, por um lado, essas tecnologias ampliam a capacidade de difusão de informações, de segmentação de mensagens e de interação entre candidaturas e eleitorado, por outro, também potencializam riscos associados à manipulação informacional, à opacidade algorítmica, à circulação de conteúdos sintéticos enganosos e à interferência sobre a formação da vontade do eleitor. Em um cenário no qual campanhas eleitorais passam a operar com apoio de sistemas de inteligência artificial generativa, *chatbots*, algoritmos de recomendação e ferramentas de impulsionamento digital, impõe-se investigar em que medida o uso dessas tecnologias é juridicamente admissível e quais limites normativos devem ser estabelecidos para preservação da integridade do processo democrático. Nesse universo tecnológico, destacam-se os Modelos de Linguagem de Grande Escala (*Large Language Models* — LLMs), sistemas de inteligência artificial treinados a partir de enormes volumes de textos, capazes de processar, gerar e manipular linguagem natural. Tais modelos baseiam-se na arquitetura *Transformer*, apresentada em 2017 (Vaswani et al., 2017), que revolucionou o campo do processamento de linguagem natural e viabilizou o atual estágio da inteligência artificial generativa.

A relevância do problema decorre da circunstância de que o debate sobre liberdade de escolha política não pode mais ser travado com base em concepção simplificada do eleitor como agente integralmente racional e imune a condicionamentos externos. A literatura filosófica, psicológica e neurocientífica tem demonstrado, de modo cada vez mais consistente, que a tomada de decisão humana resulta da interação entre elementos conscientes e inconscientes, fatores contextuais, vieses cognitivos e estruturas de processamento informacional que escapam, ao menos em parte, ao controle reflexivo do sujeito. Em consequência, o ambiente digital contemporâneo não pode ser tratado como

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

mero espaço neutro de circulação de ideias. Ao contrário, trata-se de ecossistema tecnicamente estruturado para selecionar, priorizar, recomendar e amplificar conteúdos, influenciando percepções, preferências e comportamentos de forma muitas vezes invisível ao próprio usuário.

No contexto eleitoral, essa constatação assume especial relevância. A utilização da inteligência artificial em campanhas pode favorecer práticas legítimas de comunicação política, como a ampliação do acesso à informação, a facilitação do diálogo com o eleitor e a otimização de estratégias de mobilização. Contudo, as mesmas ferramentas podem ser empregadas para fins incompatíveis com a normalidade e a legitimidade das eleições, a exemplo da produção de *deepfakes*, da personalização persuasiva em escala, da manipulação conversacional automatizada, da amplificação de vieses políticos por algoritmos de recomendação e da exploração de vulnerabilidades cognitivas do eleitorado. O problema jurídico, portanto, não reside apenas no conteúdo manifestamente ilícito, mas também na arquitetura tecnológica que medeia a formação da opinião política e condiciona a autonomia informacional dos cidadãos.

É nesse quadro que se insere o presente artigo, cujo objetivo consiste em analisar a utilização da inteligência artificial em campanhas eleitorais e seus efeitos sobre a formação da vontade do eleitor. A pesquisa busca responder à seguinte indagação central: é juridicamente viável o uso da inteligência artificial no processo eleitoral e, em caso afirmativo, em que condições esse uso pode ser compatibilizado com a liberdade de expressão, a igualdade de oportunidades entre candidaturas e a autonomia decisória do eleitor? Para enfrentar essa questão, o trabalho examina, de um lado, os fundamentos teóricos relacionados ao livre-arbítrio e à formação da vontade; de outro, analisa estudos empíricos sobre vieses cognitivos, influência informacional e impacto da inteligência artificial, especialmente da inteligência artificial generativa, dos *chatbots* e dos algoritmos de recomendação em redes sociais, sobre a opinião e o comportamento político.

O estudo foi desenvolvido por meio do método lógico-dedutivo, com apoio em bibliografia nacional e estrangeira. Trata-se de pesquisa qualitativa, mas

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

fundada também na análise de evidências empíricas produzidas por terceiros, extraídas sobretudo de investigações realizadas no campo da neurociência, da psicologia cognitiva, da ciência política e dos estudos sobre plataformas digitais. Embora o foco normativo do artigo recaia sobre o ordenamento jurídico brasileiro, parte relevante da base empírica utilizada decorre de estudos conduzidos nos Estados Unidos da América, país cujo processo eleitoral tem servido de campo para observação dos efeitos da inteligência artificial, da segmentação algorítmica e da comunicação política digital.

Quanto ao percurso metodológico, o estudo adota abordagem teórico-normativa, articulando revisão bibliográfica interdisciplinar — filosofia da mente, neurociência cognitiva e ciência da computação aplicada à persuasão — com análise documental da produção empírica recente sobre persuasão e viés em modelos de linguagem e algoritmos de recomendação, com prioridade a estudos com metodologia e resultados validados por pares, distinguindo-se expressamente, nas referências, os trabalhos revisados por pares dos disponibilizados como preprint. Sobre essa base, procede-se à análise dogmática do microssistema regulatório eleitoral brasileiro (Resoluções TSE nº 23.610/2019, 23.671/2021, 23.732/2024 e 23.755/2026 e Código Eleitoral), com o objetivo de identificar, à luz da evidência empírica, as insuficiências da disciplina vigente e formular proposições de aprimoramento normativo. Não se realiza, portanto, teste empírico próprio das soluções propostas, limite expressamente reconhecido nas considerações finais.

A hipótese desenvolvida ao longo do trabalho é a de que a inteligência artificial, quando empregada em campanhas eleitorais e em ecossistemas digitais de circulação política, possui potencial concreto para interferir, de forma inconsciente, na formação da vontade do eleitor, não apenas por meio da produção de conteúdos falsos ou manipulados, mas também pela operação de mecanismos menos visíveis de persuasão, recomendação e direcionamento informacional. Por essa razão, sustenta-se que a resposta jurídica adequada não consiste em proibição absoluta da tecnologia, mas em seu enquadramento regulatório por meio de deveres reforçados de transparência, informação

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

qualificada, mitigação de vieses, responsabilização dos agentes envolvidos e proteção da autonomia informacional dos usuários.

A estrutura do artigo desenvolve-se em três eixos principais. Em primeiro lugar, examina-se a premissa clássica do livre-arbítrio e sua releitura à luz de contribuições contemporâneas da neurociência e da teoria cognitiva, a fim de demonstrar que a vontade humana se forma em ambiente permeado por condicionamentos múltiplos e nem sempre conscientes. Em seguida, são analisadas evidências empíricas acerca da influência de informações, conteúdos midiáticos, algoritmos de recomendação e modelos de linguagem sobre cognição, memória, empatia e decisão política. Por fim, procede-se à análise crítica da disciplina normativa brasileira aplicável ao tema, com especial atenção à regulamentação expedida pelo Tribunal Superior Eleitoral, identificando-se eventuais avanços, insuficiências e lacunas, especialmente no que se refere aos modelos de linguagem de grande escala.

Ao final, pretende-se demonstrar que a proteção da democracia, no ambiente digital, exige mais do que a repressão posterior a abusos evidentes. Exige, sobretudo, a construção de um modelo regulatório capaz de enfrentar a opacidade tecnológica, reduzir assimetrias informacionais e preservar as condições mínimas para uma deliberação política livre, informada e não manipulada. Nesse sentido, a discussão proposta insere-se no esforço mais amplo de adaptação do direito eleitoral e do direito digital aos desafios impostos por tecnologias capazes de remodelar, de forma profunda, a própria infraestrutura da esfera pública contemporânea.

2. Metodologia

O presente estudo adota abordagem teórico-normativa, de natureza qualitativa e método lógico-dedutivo, articulando três planos: (i) revisão bibliográfica interdisciplinar (filosofia da mente, neurociência cognitiva e ciência da computação aplicada à persuasão); (ii) análise documental da produção empírica recente sobre persuasão e viés em modelos de linguagem e algoritmos

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

de recomendação; e (iii) análise dogmática do microssistema regulatório eleitoral brasileiro.

A seleção da literatura empírica é de natureza narrativa e exploratória, e não sistemática, limitação que se reconhece expressamente. Adotaram-se como critérios de inclusão: (a) recorte temporal concentrado no quinquênio 2021–2026, justificado pela emergência dos LLMs conversacionais de grande escala; (b) pertinência direta ao objeto (persuasão algorítmica, viés político de IA, efeitos de feeds e recomendação sobre atitudes políticas); e (c) prioridade a estudos com metodologia e resultados validados por pares, distinguindo-se expressamente, nas referências, os trabalhos revisados por pares dos disponibilizados como preprint. A inclusão pontual de preprints justifica-se pela velocidade da produção científica sobre LLMs, cujos achados mais recentes frequentemente antecedem a publicação revisada; tais fontes são empregadas de modo subsidiário e sempre identificadas como não revisadas. Reconhece-se que a seleção narrativa comporta risco de viés de confirmação, mitigado pela inclusão de achados de vetores ideológicos divergentes, o que reforça — e não enfraquece — a tese da não neutralidade.

A mobilização de estudos majoritariamente anglófonos justifica-se pela escassez de pesquisa empírica equivalente em língua portuguesa, e seus achados são utilizados como base para compreender vulnerabilidades cognitivas exploráveis, e não como prova direta de manipulação eleitoral no Brasil. Os estudos neurocientíficos sobre livre-arbítrio, empatia e decisão de júri são empregados em chave analógica limitada: não demonstram automaticamente manipulação eleitoral — contextos de júri e pleito diferem em escala, temporalidade e mediação social —, mas fornecem fundamento para a premissa de vulnerabilidade decisória. As especificidades culturais, linguísticas e institucionais do eleitorado brasileiro recomendam cautela na transposição, conforme registrado nas considerações finais. A análise normativa procede pela técnica dogmático-jurídica, identificando objeto, conduta regulada, sujeito responsável e lacunas de cada dispositivo.

3. Do livre arbítrio

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

Benjamin Libet (1985, p. 529), ao realizar estudo sobre o tema da consciência e liberdade, concluiu que “processos associados com responsabilidade individual e livre-arbítrio ‘operariam’ não para iniciar um ato voluntário, mas para selecionar e controlar os resultados volitivos” (tradução nossa).

Joana Gonçalves Coelho (2014, p. 157) está correta quando, ao realizar análise do estudo acima referenciado, afirma que: “apenas deveríamos deslocar o foco, ou seja, o livre-arbítrio não está na possibilidade de interferir nos acontecimentos cerebrais inconscientes que precedem a ação, portanto, não está na capacidade de iniciar um ato voluntário, mas sim no poder de controlar a realização ou não do ato”.

Registrem-se críticas profundas aos estudos de Benjamin Libet. Schurger et al. (2021) afirmam que o experimento de Libet consagrou-se como evidência neurocientífica contra o livre-arbítrio ao mostrar que o *readiness potential* (RP) precede em centenas de milissegundos o relato consciente da intenção de mover (*W-time*), sugerindo que o cérebro "decide antes" do agente. Os autores contestam essa leitura ao demonstrar, com base em modelos estocásticos de decisão (SDMs), que o RP pode ser plenamente explicado como artefato da média de flutuações neurais aleatórias alinhadas retrospectivamente ao movimento. Segundo o *Accumulation-to-Bound Model*, a ação é desencadeada quando o ruído cortical, somado a um viés tônico, cruza um limiar, o início do RP. Os autores reconhecem que os SDMs trazem novos desafios filosóficos, mas concluem que, mesmo que o modelo estocástico esteja equivocado, as críticas metodológicas e filosóficas acumuladas já bastam para invalidar a tese de Libet.

Logo, qualquer que seja a posição adotada no debate sobre o livre-arbítrio, seja a que enfatiza a capacidade de iniciar ações, seja a que destaca o controle posterior do processo decisório, seja, ainda, a que compreende a decisão como resultado de dinâmicas neurocognitivas complexas, não decorre daí a negação da agência humana. Ao contrário, permanece possível sustentar que o eleitor, ao exercer o direito ao voto, possui capacidade, ao menos em alguma medida, de buscar informações, avaliar alternativas, responder a razões e formar convicções políticas. No processo eleitoral, portanto, parte-se da premissa de que a decisão do

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

eleitor pode envolver deliberação e análise de opções disponíveis no espaço público. É justamente essa premissa mínima de autonomia decisória, e não a suposição de racionalidade perfeita, que torna juridicamente relevante investigar em que medida fatores externos, como a arquitetura informacional das redes sociais e o uso de inteligência artificial, podem interferir na formação da vontade política. Sebástian Borges de Albuquerque Mello (2015, p. 88) pondera que: “a maior parte daquilo que cada pessoa faz ou sente não está sob o seu controle consciente. [...] O cérebro é visto como um órgão que toma decisões sem que nossa consciência perceba”.

Thomas Hills (2019) propõe uma reformulação do problema clássico do livre-arbítrio a partir de uma perspectiva neurocognitiva, afastando-se tanto de concepções metafísicas tradicionais quanto de reducionismos que o tratam como mera ilusão. Em vez de concebê-lo como uma propriedade binária, existente ou inexistente, o autor o apresenta como um conjunto integrado de capacidades cognitivas graduais, ancoradas em processos biológicos específicos. Nesse enquadramento, o livre-arbítrio emerge da articulação entre quatro dimensões fundamentais: a capacidade de gerar variabilidade comportamental (isto é, produzir alternativas), o ajuste dessa variabilidade em função de objetivos hierarquicamente organizados, a deliberação orientada por representações internas e a construção dinâmica de um “eu” consciente que participa da seleção entre possibilidades.

Segundo o autor, tal arquitetura permite compreender o agente como um sistema que explora o espaço de ações possíveis por meio de mecanismos parcialmente imprevisíveis, mas não arbitrários, uma vez que essa imprevisibilidade é direcionada por metas. Nesse sentido, a deliberação não consiste apenas na escolha entre opções previamente dadas, mas na própria produção dessas opções por meio de simulações internas de futuros possíveis, frequentemente baseadas na recombinação de experiências passadas.

A conclusão central do artigo reside na tese de que o livre-arbítrio deve ser entendido como um processo de autoconstrução generativa, no qual o sujeito simultaneamente produz e avalia versões alternativas de si mesmo, de modo que identidade e decisão emergem em conjunto.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

Entendemos que o trabalho de Mudrik *et al.* (2022) fornece um resumo sobre o "estado da arte" dos estudos sobre livre-arbítrio. A partir de uma revisão de estudos empíricos, os autores argumentam que fatores inconscientes influenciam nosso comportamento, mas não eliminam o livre-arbítrio, embora revelem limites importantes sobre o grau de liberdade que possuímos em situações específicas. Diversas pesquisas indicam que nossas escolhas podem ser moldadas por elementos dos quais não temos consciência. Por exemplo, mudanças sutis na forma como opções são apresentadas, como a ordem em uma lista ou a escolha padrão, podem direcionar decisões sem que percebamos. Da mesma forma, características aparentemente irrelevantes, como a aparência de uma pessoa ou estereótipos sociais, podem influenciar julgamentos, inclusive em contextos sérios como decisões judiciais. Além disso, estados internos como cansaço ou fome também afetam nossas escolhas, assim como estímulos subliminares, que operam abaixo do nível da consciência.

Diante dessas evidências, os autores colocam a seguinte pergunta de pesquisa: se nossas decisões são influenciadas por fatores inconscientes, ainda podemos considerá-las livres? O artigo responde a essa questão analisando o papel da consciência no processo decisório. Demonstram que não é necessário estarmos conscientes de todos os fatores envolvidos para que uma decisão seja livre. No entanto, a falta de consciência sobre como esses fatores influenciam nossas escolhas pode comprometer nossa autonomia.

Em outras palavras, o problema não é simplesmente não perceber certos estímulos, mas não entender sua influência sobre nossas decisões. Para avaliar o impacto dessas influências, o artigo adota dois critérios fundamentais para o livre-arbítrio: o controle sobre as próprias ações e a capacidade de responder a razões. Uma decisão é considerada mais livre quando o indivíduo consegue agir de acordo com seus objetivos e valores, reconhecendo e avaliando motivos relevantes para agir. As influências inconscientes podem enfraquecer esses aspectos, por exemplo, ao nos levar a tomar decisões baseadas em fatores irrelevantes ou a agir de maneira automática. Contudo, isso não significa que perdemos completamente o

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

controle ou a capacidade de raciocinar; significa apenas que essas capacidades podem ser reduzidas em determinadas circunstâncias.

Dessa forma, a principal conclusão do artigo é que o livre-arbítrio não é algo absoluto, mas gradual. Em vez de sermos totalmente livres ou completamente determinados, nossa liberdade varia conforme o contexto e as condições em que tomamos decisões. Em situações em que estamos atentos, informados e menos sujeitos a vieses, tendemos a agir com maior liberdade. Já em contextos marcados por cansaço, pressões externas ou influências inconscientes, nossa liberdade pode ser limitada. Por fim, os autores destacam um ponto importante: compreender os fatores que influenciam nossas decisões pode, na verdade, aumentar nossa liberdade. Ao nos tornarmos conscientes de vieses e influências inconscientes, ganhamos a possibilidade de corrigi-los ou evitá-los, fortalecendo nosso controle e nossa capacidade de agir de acordo com nossas razões.

Outro aspecto importante é o *nudge.* Thaler e Sunstein (2021) criam o conceito de paternalismo libertário. Diferentemente do idealizado *homo economicus*, o ser humano é falível e movido por dois sistemas mentais (o automático, intuitivo e rápido, e o reflexivo, deliberativo e lento) e sujeito a vieses sistemáticos como a inércia, o viés do *status quo* e a falácia do planejamento.

Conforme estabelece Claudio José Langroiva Pereira et al. (2020), esses vieses sistemáticos decorrem, em larga medida, do funcionamento das heurísticas, estratégias cognitivas simplificadas que o cérebro utiliza para tomar decisões rápidas em situações de complexidade ou incerteza. Embora úteis e eficazes na maioria das situações cotidianas, as heurísticas tornam-se fontes de erro sistemático em contextos que exigem julgamento cuidadoso, dando origem a desvios previsíveis na formação da vontade. É justamente sobre esse substrato cognitivo, operado pelo sistema automático e rápido, que atuam tanto os nudges quanto, como se verá adiante, as ferramentas de inteligência artificial e os algoritmos de recomendação.

Portanto, partindo da premissa de que não existe arquitetura de escolha neutra, pois toda forma de apresentar opções influencia decisões, os autores defendem que governos e instituições privadas devem atuar como arquitetos de

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

escolha conscientes, projetando ambientes decisórios que tornem as opções benéficas mais fáceis e prováveis por meio de *nudges*, pequenos empurrões previsíveis que preservam integralmente a liberdade de escolha sem proibir alternativas ou alterar incentivos econômicos significativos, aplicáveis a domínios como poupança previdenciária, doação de órgãos, saúde, educação e meio ambiente.

O ponto principal é que os *nudges*, que também têm potencial para incidir sobre temas relacionados às eleições e política, podem, se livremente aplicados, influenciar a vontade popular e afetar o processo democrático. Apesar da inexistência de um design neutro, esse é um tema que deve estar no foco do legislador, de forma a evitar que esses "pequenos empurrões" desequilibrem a disputa eleitoral.

Assim, é possível fixar a premissa de que o livre-arbítrio é caracterizado por uma profunda interação entre elementos puramente conscientes e inconscientes. O inconsciente, por sua vez, é uma formação psíquica abstrata, complexa e multifatorial que envolve desde experiências passadas (incluindo educação) até mesmo fatores ambientais imediatos, tais como privação de sono, fome e temperatura ambiente.

Cabe, portanto, responder se a combinação de redes sociais, viés algorítmico e técnicas de chatbot em LLMs pode atuar sobre o substrato inconsciente da decisão, a ponto de induzir, e não apenas informar, a vontade política do eleitor.

Fixadas tais premissas, delimita-se o que se entende por manipulação, distinguindo-a da persuasão legítima. Adotam-se os critérios de Mudrik et al. (2022) — controle sobre as próprias ações e capacidade de responder a razões. A persuasão lícita opera sobre o segundo critério, oferecendo razões que o eleitor pode aceitar ou rejeitar conscientemente. A manipulação caracteriza-se pela degradação opaca desses critérios, atuando sobre o substrato inconsciente sem que o agente compreenda a influência sofrida. Três elementos a qualificam: (i) interferência por vias que escapam à deliberação consciente; (ii) opacidade quanto à existência e à dimensão da influência; e (iii) degradação da

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

responsividade a razões, mediante vieses, distorções factuais ou estados mentais artificialmente induzidos — em diálogo direto com o art. 242 do Código Eleitoral. Disso decorre uma tipologia gradativa, com tratamentos jurídicos diferenciados: (a) informação legítima e (b) persuasão política ordinária — lícitas, sem restrição; (c) persuasão segmentada transparente — lícita, sujeita a dever de identificação; (d) persuasão opaca baseada em dados pessoais inferidos — sujeita a auditoria e limites de tratamento de dados; (e) desinformação sintética e (f) deepfake — vedadas (art. 9º-C); (g) indução subliminar e (h) manipulação cognitiva ilícita — vedadas e sujeitas a responsabilização. O critério distintivo entre persuasão humana e algorítmica que justifica tratamento especial não é a finalidade persuasiva — comum a ambas —, mas o conjunto escala, opacidade, adaptabilidade hiperpersonalizada, simulação conversacional e dificuldade de auditoria e *accountability*.

Cumpre ainda distinguir três planos de interferência, relevantes para o *enforcement:* o favorecimento indireto ainda possui objeto identificável (candidatura beneficiada); o viés estrutural molda percepções sem recomendação eleitoral explícita; e a influência subliminar pode não ser atribuível a candidatura determinada. Quanto mais a influência se desloca do favorecimento identificável para o viés estrutural difuso, maior a dificuldade de imputação e fiscalização — o que fundamenta as propostas de auditoria adiante.

4. Da influência da inteligência artificial

Na verdade, a preocupação com a influência da comunicação sobre o inconsciente e o processo de tomada de decisão antecede as mudanças decorrentes da era digital e das redes sociais. Sobre o viés do devido processo legal, na década de 1970, a Suprema Corte dos Estados Unidos se viu diante de um conflito com o mesmo pano de fundo aqui discutido. No caso *Nebraska Press Association v. Stuart* (EUA, 1976), um magistrado de primeira instância, responsável por conduzir um caso de homicídio que vinha recebendo ampla cobertura da mídia, expediu uma decisão que vedava aos jornalistas a divulgação

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

ou veiculação das confissões prestadas pelo réu às autoridades policiais. Na avaliação do juiz, tal restrição era indispensável para assegurar que o acusado tivesse acesso a um julgamento imparcial. A premissa da ordem judicial era o reconhecimento de que a ampla cobertura midiática poderia atingir pessoas que iriam compor o júri, afastando a imparcialidade. Ou seja, a exposição à notícia afastaria o *fair trial*.

A Suprema Corte analisou a decisão à luz da liberdade de expressão e estabeleceu que, embora o caso de homicídio realmente tenha provocado forte repercussão na mídia antes do julgamento, impor uma censura prévia à imprensa não seria uma solução eficaz para preservar os direitos do réu. Nas palavras do tribunal: "não é possível silenciar toda uma comunidade sobre um tema que afeta diretamente sua própria vida". *Gag orders* contra a imprensa são, portanto, remédios inadmissíveis. Anote-se que, diante da leitura do precedente, é possível concluir que a restrição à publicidade somente seria viável, em caráter excepcional, se houvesse provas concretas dos riscos, ou seja, em caráter excepcional e concretamente demonstrado.

Posteriormente, em *Richmond Newspapers, Inc. v. Virginia* (EUA, 1980), a Suprema Corte se viu diante de um caso envolvendo um homicídio que passou por diversos julgamentos que acabaram sendo anulados. Diante disso, o juiz responsável determinou que as audiências seguintes ocorressem sem acesso do público nem da imprensa. O julgamento se deu à luz da Primeira Emenda (liberdade de expressão) e da Sexta Emenda (júri imparcial). Decidiu-se que o direito de o público acompanhar julgamentos criminais, embora não esteja escrito de forma expressa, está subentendido nas proteções oferecidas pela Primeira Emenda. Destacou-se que a referida emenda também ampara o direito de as pessoas se reunirem em espaços públicos, entre os quais se incluem os tribunais. Novamente adotou a premissa de que a ampla exposição não afeta a capacidade das pessoas estabelecerem decisões racionais.

No Brasil, no HC 225180 (Brasil, 2023), restou estabelecido que a rotineira veiculação de notícias sobre fatos criminosos por intermédio da imprensa, sobretudo com as facilidades atuais de propagação da notícia, não é capaz de,

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

somente pela notoriedade assumida pelo caso, tornar o corpo de jurados tendencioso. Ou seja, estabeleceu a necessidade de prova concreta da perda da imparcialidade e, novamente, afastou a premissa da influência, *per se*, de meios de comunicação sobre a formação da vontade e imparcialidade.

As decisões revelam uma tensão entre o padrão jurídico de prova concreta da perda de imparcialidade e os achados empíricos sobre influência informacional, tensão que, em nossa leitura, recomenda reavaliação à luz das especificidades dos ambientes algorítmicos. Christine L. Ruva e Christina C. Guenther (2015) realizaram pesquisa de como a exposição à publicidade negativa antes do julgamento (PTP) e a deliberação em grupo influenciam as decisões, a memória e as percepções de júris simulados. Os resultados mostraram que jurados expostos a informações negativas prévias apresentaram maior tendência a considerar o réu culpado, avaliar sua credibilidade de forma mais negativa e cometer erros de memória, frequentemente confundindo informações externas com evidências do julgamento. A deliberação em grupo teve efeitos ambíguos, podendo reduzir a tendência à condenação entre jurados não expostos à PTP, mas também prejudicar a precisão da memória. Além disso, nas discussões em grupo, jurados expostos à PTP tenderam a interpretar evidências ambíguas de maneira favorável à acusação e a desconsiderar a ausência de provas sólidas. Mesmo com instruções judiciais para ignorar informações externas e com a deliberação coletiva, o viés gerado pela exposição prévia à mídia não é eliminado, pois está associado a distorções de memória, formação de impressões tendenciosas e interpretações enviesadas das evidências, que continuam a influenciar significativamente as decisões finais dos jurados[1].

Sobre liberdade de expressão e controle do mérito do discurso, há estudo sobre possíveis influências na formação da opinião e comportamento de pessoas expostas a discurso de ódio. Soral et al. (2023), após realização de pesquisa empírica, afirmam que a onipresença generalizada do discurso de ódio afeta as atitudes e os comportamentos. A exposição pode levar ao preconceito, à

[1] O texto trata de discussão em grupo no âmbito do júri, uma vez que, nos Estados Unidos da América, há debate entre os jurados.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

desumanização e à falta de empatia em relação a membros de grupos externos. Utilizando análises de todo o cérebro e de regiões de interesse, os autores afirmam demonstrar que a exposição a linguagem depreciativa sobre migrantes reduz a resposta cerebral à dor de outras pessoas na junção temporoparietal direita (rTPJ), independentemente do pertencimento ao grupo. Como a rTPJ está associada a processos relevantes para a tomada de perspectiva, sua atividade reduzida pode estar relacionada a uma menor propensão de adotar a perspectiva psicológica dos outros.

Essa situação se torna ainda mais grave quando constatamos a "disfunção narcotizante", nos termos de Merton e Lazarsfeld (2000), em que o ser humano perde o protagonismo na construção de conclusões e passa a, passivamente, receber, como se verdades fossem, as informações fornecidas pela *mass media* e quando a informação vem juntamente com o rótulo de opinião de um especialista, o que Pierre Bourdieu (1997) chama de "*fast thinkers".*

Considerando as premissas dos estudos como corretas, isto é, a possibilidade de influência sobre o processo decisório (inclusive em processos judiciais) mediante a prévia exposição a determinados conteúdos, é preciso questionar como e em que medida esse efeito se dá sobre o processo eleitoral, com foco em redes sociais e *chatbots*. Para tanto, vamos utilizar dados obtidos por outros estudos científicos de neurociência com metodologia e resultados validados por pares.

Hackenburg et al. (2025) estabeleceram que o avanço dos modelos de linguagem de grande escala (*Large Language Models* — LLMs) tem suscitado preocupações quanto ao seu potencial de influenciar crenças e moldar a opinião pública em contextos políticos. O estudo teve como objetivo identificar e mensurar sistematicamente as principais "alavancas" que determinam a persuasividade da inteligência artificial (IA) conversacional em discussões políticas. Para tanto, os autores conduziram três experimentos em larga escala, totalizando 76.977 respostas obtidas de 42.357 participantes, nos quais foram empregados 19 LLMs distintos, incluindo modelos especificamente pós-treinados para fins persuasivos, aplicados a 707 questões políticas do contexto britânico. Adicionalmente, procedeu-

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

se à verificação da precisão factual de 466.769 afirmações geradas pelos modelos. Quatro fatores foram manipulados de forma independente: escala computacional, métodos de pós-treinamento, personalização com base em dados do usuário e estratégias retóricas induzidas por *prompting*. Os resultados evidenciaram que as técnicas de pós-treinamento e de *prompting* retórico constituíram as alavancas mais robustas, elevando a persuasividade em até 51% e 27%, respectivamente, ao passo que o aumento da escala do modelo e a personalização exerceram efeitos comparativamente menores. Verificou-se, ainda, que o principal mecanismo subjacente à eficácia persuasiva foi a densidade informacional, isto é, a concentração de alegações factuais nos argumentos gerados.

Segundo os autores, as mesmas técnicas que ampliaram a capacidade persuasiva dos modelos reduziram, de forma sistemática, sua precisão factual. Conclui-se, portanto, que o potencial persuasivo da IA conversacional contemporânea decorre menos do poder computacional bruto ou da personalização e mais das estratégias de refinamento e instrução que mobilizam a habilidade dos LLMs de produzir informação de modo rápido e estratégico. Tal constatação aponta para um dilema estrutural relevante, na medida em que a otimização desses sistemas para fins persuasivos tende a intensificar a disseminação de conteúdos imprecisos ou enganosos.

Fisher et al. (2026) afirmam existir uma incorporação dos Modelos de Linguagem de Grande Escala (*Large Language Models* — LLMs) às práticas cotidianas de busca de informação, deliberação pública e formulação de políticas. Ponderam que, embora a literatura especializada já tenha documentado de forma robusta a presença de vieses inerentes a esses sistemas, sejam eles de natureza social, demográfica ou partidária, permanecia em aberto a questão de como tais vieses, quando mobilizados em interações dinâmicas e dialógicas, repercutem sobre a cognição e a conduta política dos usuários. Houve dois experimentos interativos com 299 participantes norte-americanos autoidentificados como Democratas ou Republicanos. Foram testados os efeitos da exposição a LLMs partidariamente enviesados, liberais, conservadores ou neutros, sobre a formação de opiniões em tópicos políticos pouco conhecidos e sobre decisões alocativas em

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

um cenário de distribuição orçamentária municipal. Empregando o modelo GPT-3.5-turbo manipulado mediante prefixos de *prompt* e validado pelo *Political Compass Test*, os autores estruturaram um delineamento experimental 3×2, no qual os participantes interagiam livremente com o modelo via *chatbot*, sem prévio conhecimento de seu eventual enviesamento. Os resultados obtidos revelam, de modo convergente e estatisticamente significativo, que a interação com LLMs enviesados conduz os usuários a deslocarem suas opiniões e decisões na direção do viés do modelo, independentemente de sua filiação partidária prévia e, notadamente, mesmo quando tal viés contraria suas convicções ideológicas declaradas, fenômeno que desafia diretamente a literatura clássica sobre resistência à mudança de atitude em interações de curta duração. Verificou-se, ademais, que apenas 54% dos participantes foram capazes de identificar corretamente o viés do modelo com o qual interagiam e que, mais preocupante ainda, o reconhecimento explícito do viés não o atenua de forma suficiente para neutralizá-lo, contrariando hipóteses tradicionalmente sustentadas pela pesquisa sobre vieses na mídia convencional. Em contrapartida, observou-se que o letramento prévio em inteligência artificial se correlacionou modestamente com uma menor susceptibilidade ao efeito do viés, sugerindo que a familiaridade crítica com as limitações desses sistemas pode operar como fator de proteção cognitiva.

À luz desses achados, os autores sustentam que os LLMs introduzem dinâmicas inéditas no ecossistema informacional contemporâneo, na medida em que são percebidos simultaneamente como autoritativos e objetivos, propriedade que amplifica seu potencial de influência em comparação às fontes midiáticas convencionais. Conclui-se, portanto, que a influência exercida por LLMs enviesados sobre a opinião e a decisão política configura um risco democrático tangível e não negligenciável, dotado de potencial para reconfigurar processos eleitorais, debates públicos e a própria arquitetura das políticas públicas, sobretudo em contextos de polarização acentuada.

Gauthier et al. (2026) investigam, por meio de um experimento de campo com 4.965 usuários ativos do X nos Estados Unidos, conduzido ao longo de sete semanas em 2023, se o algoritmo de feed da plataforma exerce influência causal

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

sobre atitudes políticas. O estudo dialoga diretamente com o achado contraintuitivo do experimento da Meta de 2020, segundo o qual substituir o feed algorítmico pelo cronológico não produzira efeitos detectáveis sobre polarização ou opinião política. Hipotetizando que tal resultado mascara efeitos persistentes da exposição inicial à curadoria algorítmica, os autores examinam separadamente os efeitos de ligar e desligar o algoritmo, e analisam o mecanismo causal por meio das contas que os usuários passam a seguir. O estudo foi conduzido de forma independente, sem cooperação da plataforma, e pré-registrado.

Os resultados destacam sua assimetria. Ligar o algoritmo, para quem usava o *feed* cronológico, deslocou as opiniões políticas em direção a posições mais conservadoras: aumentou em 0,11 desvios-padrão a priorização de pautas conservadoras, em 0,08 desvios-padrão a crença de que as investigações criminais contra Trump são inaceitáveis, e em 0,12 desvios-padrão as atitudes pró-Kremlin sobre a guerra na Ucrânia. Em termos absolutos, os participantes expostos ao algoritmo se tornaram 7,4 pontos percentuais menos propensos a ter visão positiva do presidente ucraniano Zelensky e 5,5 pontos percentuais mais propensos a considerar inaceitáveis as investigações contra Trump. Em contraste radical, desligar o algoritmo não produziu nenhum dos efeitos simétricos esperados: as atitudes políticas permaneceram inalteradas. Notavelmente, em ambas as direções, partidarismo autodeclarado, polarização afetiva e bem-estar subjetivo não foram afetados, sugerindo que essas dimensões mais rígidas da identidade política são resistentes a sete semanas de exposição.

A análise do conteúdo capturado revelou que o algoritmo do X promove conteúdo com 408% a 508% mais engajamento, aumenta em 19,9% a probabilidade de aparecer conteúdo conservador (contra apenas 3,1% para conteúdo progressista), reduz em 58,1% a presença de mídia tradicional e aumenta em 27,4% a presença de ativistas políticos. Crucialmente, o algoritmo promove conteúdo conservador tanto para Democratas quanto para Republicanos e Independentes, embora a persuasão seja mais forte entre estes últimos.

A grande contribuição teórica do estudo, contudo, está na elucidação do mecanismo causal que explica a assimetria observada e que reconcilia os achados

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

aparentemente contraditórios da literatura anterior. Os autores demonstram que a exposição ao algoritmo leva os usuários a seguir contas de ativistas conservadores, efeito de 0,18 desvios-padrão, e que esses usuários continuam seguindo essas contas mesmo após o algoritmo ser desligado. Como resultado, mesmo no *feed* cronológico subsequente, os usuários previamente expostos ao algoritmo veem 60% mais *posts* conservadores do que aqueles que nunca passaram pelo *feed* algorítmico. Em outras palavras, o algoritmo opera não apenas como um curador instantâneo de conteúdo, mas como um arquiteto persistente da rede social do usuário.

Damião et al. (2026) investigam a presença de viés político em motores de busca (Google, Bing, DuckDuckGo e Yahoo) e em modelos de linguagem de grande escala (ChatGPT-4o e Microsoft Copilot) durante o período crítico que antecedeu dois pleitos relevantes de 2024: as eleições para o Parlamento Europeu e as eleições presidenciais dos Estados Unidos. O objeto da pesquisa consistiu em verificar se tais sistemas algorítmicos, ao responderem às consultas politicamente neutras, super-representam determinadas entidades políticas ou agendas para além do que seria razoavelmente explicado por fatores externos como sondagens eleitorais, resultados de pleitos anteriores ou saliência mediática, problematizando, assim, sua suposta neutralidade informacional. Para tanto, os autores desenvolveram uma metodologia robusta, baseada na implementação de bots automatizados que simulavam comportamento humano de navegação, distribuídos geograficamente em cinco países europeus (Áustria, Alemanha, Irlanda, Polônia e Portugal) e em quinze condados estadunidenses com diferentes orientações políticas, totalizando aproximadamente 4.360 consultas que resultaram em 4.358 páginas de resultados de busca e cerca de 200 respostas de LLMs. As menções a políticos, partidos, espectros ideológicos e, no caso norte-americano, temas politicamente polarizados, foram extraídas mediante combinação de técnicas automatizadas com validação manual, sendo posteriormente mapeadas em famílias ideológicas e comparadas com três indicadores externos de referência. Os resultados demonstraram, no contexto europeu, uma super-representação sistemática e estatisticamente significativa de entidades associadas à Direita

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

Radical em todos os motores de busca analisados e no ChatGPT, fenômeno que persistiu inclusive em países onde tais agremiações possuem baixa expressão eleitoral, como Irlanda e Portugal, com as menções nos SEs excedendo em cerca de 80% a saliência mediática observada. No contexto estadunidense, embora as diferenças tenham sido mais sutis em razão do sistema bipartidário, identificou-se que o Google favoreceu fortemente temas prioritários para o eleitorado republicano, ao passo que os demais motores apresentaram leve inclinação em direção a pautas democratas. De forma geral, os modelos de linguagem revelaram comportamento mais equilibrado, ainda que marcado por taxas variáveis de recusa em responder. Notou-se também que, no caso europeu, os LLMs apresentaram uma tendência centralizadora com sobrerrepresentação concomitante da Direita Radical e dos Verdes. Conclui-se, portanto, que existe um viés algorítmico sistemático e mensurável que não pode ser plenamente explicado por fatores contextuais externos, sugerindo que mesmo consultas neutras produzem representações enviesadas do cenário político, com potencial impacto sobre eleitores indecisos e, consequentemente, sobre processos democráticos. Diante da opacidade inerente aos algoritmos proprietários, que impede a identificação precisa das causas desses padrões, sejam elas relacionadas a engajamento de usuários, otimização para mecanismos de busca, demanda de pesquisa ou ajustes deliberados, os autores defendem a urgência de auditorias independentes e maior transparência algorítmica.

Em síntese, a evidência empírica acumulada nos últimos cinco anos converge para um conjunto de conclusões robustas que reconfiguram o debate sobre algoritmos e democracia. Primeiro, algoritmos de redes sociais não são neutros: amplificam sistematicamente certos tipos de conteúdo, particularmente conteúdo conservador, agitador e gerado por ativistas, em detrimento de mídia tradicional e conteúdo moderado. Segundo, seus efeitos políticos são reais, mas operam de forma sutil e persistente: não alteram identidade partidária ou polarização afetiva no curto prazo, mas modificam prioridades de políticas, percepções sobre eventos atuais e, sobretudo, a estrutura das redes sociais dos usuários. Terceiro, o paradigma de "ligar/desligar" o algoritmo é metodologicamente

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

insuficiente: estudos que apenas desligam o algoritmo subestimam dramaticamente seus efeitos reais, pois ignoram o efeito persistente sobre as escolhas de *followings*. Quarto, a influência algorítmica é assimétrica: opera com mais intensidade ao expor usuários a novos conteúdos do que ao retirá-los desse fluxo.

O estudo de Ye, Luceri e Ferrara (2025), apresenta uma auditoria dos vieses algorítmicos da plataforma X (antigo Twitter) ao longo de seis semanas, abrangendo o mês anterior e as duas semanas posteriores à eleição presidencial norte-americana de 2024. Diante do fato de que aproximadamente 50% dos *tweets* exibidos nas *timelines* dos usuários provêm de contas que eles não seguem, os autores investigam especificamente as recomendações fora da rede de contatos, uma dimensão pouco explorada na literatura anterior, que tendia a tratar a *timeline* algorítmica como entidade unificada.

Metodologicamente, os pesquisadores empregaram uma auditoria do tipo sock-puppet com 120 contas artificiais distribuídas em quatro grupos — neutras, esquerdistas, direitistas e balanceadas (30 em cada) —, coletando 9,79 milhões de tweets ao longo do período eleitoral. As contas partidárias seguiam dez veículos de mídia selecionados a partir do *AllSides Media Bias Chart*, sendo sete de viés moderado (*center-left* ou *center-right*) e três de viés acentuado (*left* ou *right*), além de figuras políticas-chave de cada espectro: Kamala Harris, Tim Walz e as bancadas democratas no caso das contas de esquerda; Donald Trump, JD Vance e as bancadas republicanas no caso das contas de direita. As contas balanceadas seguiam cinco veículos *center-left* e *cinco center-right*, juntamente com ambos os candidatos presidenciais, configurando uma dieta informacional centrista. As contas neutras não seguiam ninguém, simulando usuários recém-cadastrados e funcionando como sonda do comportamento padrão do algoritmo. Para mensurar exposição, desigualdade e amplificação, os autores combinaram três instrumentos: (i) a ocorrência ponderada por mil tweets, métrica que pondera a visibilidade de cada tweet por sua posição na *timeline* mediante uma função de decaimento exponencial, calibrada de modo que os 20% iniciais concentrem cerca de 70% da atenção do usuário; (ii) o coeficiente de Gini, complementado pela curva de Lorenz, para quantificar a concentração de exposição entre os usuários

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

recomendados; e (iii) a razão média de amplificação (*mean amplification ratio*), que compara a exposição recebida por contas partidárias àquela observada nas contas balanceadas, tomadas como baseline.

Os resultados revelaram três padrões centrais: (1) o algoritmo concentra a exposição em poucos usuários de alta popularidade, com contas direitistas experimentando o maior nível de desigualdade; (2) tanto contas de esquerda quanto de direita recebem amplificação de mais de 50% acima da baseline para vozes ideologicamente alinhadas, com redução simétrica da exposição a perspectivas opostas; e (3) contas neutras enfrentam um viés padrão em direção ao conteúdo conservador, com figuras de direita representando 30,16% da exposição entre os 20 usuários mais recomendados, contra apenas 12,92% para figuras de esquerda.

No desfecho, os autores articulam três contribuições centrais. Em primeiro lugar, demonstram que os achados divergem de estudos anteriores que sugeriam uma tendência centrista do algoritmo de X, indicando uma possível mudança no comportamento da plataforma após a aquisição por Elon Musk, que teria deslocado o sistema de recomendação da promoção de conteúdo moderado para o reforço explícito de preferências ideológicas existentes. Em segundo lugar, identificam uma transformação qualitativa no perfil dos atores amplificados: o algoritmo prioriza não apenas veículos de mídia tradicional e legisladores eleitos, mas também comentaristas políticos e influenciadores não institucionais, possivelmente em função da priorização de contas verificadas e assinantes pagos, fenômeno que pode introduzir sensacionalismo e desinformação no debate público. Em terceiro lugar, os autores enfatizam que a amplificação de conteúdo ideologicamente alinhado em recomendações fora da rede sugere que as câmaras de eco operam não apenas na composição das redes sociais, mas também no enquadramento ideológico do conteúdo circulante, e que o viés padrão direitista para contas neutras implica que novos usuários são expostos precocemente a conteúdo partidário, com potenciais efeitos sobre a formação de suas perspectivas políticas.

O estudo de Motoki, Pinho Neto e Rodrigues (2023) propõe uma abordagem empírico-quantitativa inovadora para mensurar o viés político em modelos de

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

linguagem de larga escala, com foco específico no ChatGPT. A metodologia articula a aplicação do questionário *Political Compass* com técnicas de impersonificação, nas quais o modelo é instado a responder sob diferentes personas, abrangendo posições médias e radicais de esquerda e direita, além de perfis profissionais e contextos nacionais distintos, como Brasil, Estados Unidos e Reino Unido. Para mitigar a natureza probabilística das respostas geradas pelo modelo, os autores realizam cem repetições de cada questionário, randomizam a ordem das perguntas e empregam a técnica de *bootstrap* com mil reamostragens, assegurando robustez estatística aos achados. Complementarmente, foram executados testes de robustez como o dose-resposta, o teste de perguntas neutras (placebo) e o de alinhamento profissional, os quais reforçam a validade dos resultados ao descartar a hipótese de que sejam meros artefatos ou flutuações aleatórias.

Os resultados revelam um viés sistemático do ChatGPT em direção ao espectro político de esquerda, evidenciado por uma correlação quase perfeita, da ordem de 0,96, com as posições dos democratas nos Estados Unidos, do então candidato Lula no Brasil e do Partido Trabalhista no Reino Unido, ao passo que as correlações com posições republicanas ou conservadoras se mostraram insignificantes ou negativas. Tais achados foram validados por análises adicionais em outros países e por testes de neutralidade, o que confere consistência aos resultados. As possíveis origens desse viés residem tanto nos dados de treinamento, marcados por uma sobrerrepresentação de fontes alinhadas à esquerda, quanto no próprio algoritmo, que pode amplificar vieses preexistentes. A principal contribuição da pesquisa, portanto, reside, embora tenha utilizado somente o ChatGPT, na proposição de um método acessível, replicável e estatisticamente robusto para a auditoria de vieses em sistemas de inteligência artificial. Registre-se que o achado de viés à esquerda diz respeito especificamente ao ChatGPT, não autorizando generalização para os LLMs como classe — o que, longe de fragilizar a conclusão, confirma que o vetor do viés é idiossincrático de cada modelo, função de seu corpus e de seu *fine-tuning*.

O relatório *Freedom of Expression and Elections in the Digital Age*

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

(A/HRC/59/50) (ONU, 2025), apresentado por Irene Khan, Relatora Especial das Nações Unidas para a promoção e proteção do direito à liberdade de opinião e expressão, ao Conselho de Direitos Humanos em junho de 2025, dedica análise substantiva ao papel das mídias sociais e da inteligência artificial na configuração contemporânea dos processos eleitorais democráticos. Embora o relatório aborde dimensões amplas, incluindo o marco jurídico internacional, a desinformação patrocinada por Estados, o declínio da mídia tradicional e violências de gênero em contextos eleitorais, esta análise concentra-se especificamente nas informações da Relatora sobre o papel das mídias sociais e da inteligência artificial.

No que concerne às mídias sociais, o relatório parte de uma constatação empírica de relevância histórica: pesquisa conduzida pela UNESCO e Ipsos, no ano de 2024, revelou que essas plataformas ultrapassaram a televisão como principal fonte de notícias, segundo pesquisa abrangendo 16 países. Tal deslocamento do epicentro informacional, segundo Khan, traz implicações profundas para os processos eleitorais, na medida em que as redes sociais foram identificadas como o principal vetor contemporâneo de desinformação e discurso de ódio. A Relatora documenta, ainda, uma migração significativa da comunicação política para plataformas de mensagens fechadas, notadamente WhatsApp, Telegram, Viber e WeChat, bem como para ambientes inicialmente concebidos para outras finalidades, como Discord e Twitch. Essa migração agrava os desafios de moderação e transparência, conforme ilustra o caso indiano, em que o partido governista Bharatiya Janata operou pelo menos cinco milhões de grupos de WhatsApp para disseminação de conteúdo político.

Articula-se a essa transformação a ascensão de uma nova figura no ecossistema midiático: o influenciador digital. O relatório dedica atenção particular ao fenômeno, destacando que tais agentes, diferentemente dos jornalistas profissionais, não estão vinculados a padrões éticos, obrigações de verificação factual ou exigências de divulgação de afiliações políticas. Pesquisa citada pela Relatora indica que aproximadamente trinta por cento dos influenciadores entrevistados em Bangladesh e na Índia relataram ter sido objeto de intimidação política para apresentar candidatos em seus conteúdos. A eleição indiana de 2024,

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

denominada por analistas como "YouTube election", exemplifica a centralidade adquirida por essa figura no engajamento eleitoral contemporâneo, fenômeno também documentado em campanhas na Bélgica, no México, na Nigéria e nos Estados Unidos.

Outra dimensão crítica abordada pelo relatório diz respeito à amplificação algorítmica enviesada. Khan documenta que, durante as eleições alemãs de 2025, setenta e oito por cento do conteúdo político recomendado pelo TikTok e sessenta e quatro por cento no X favoreciam o partido de extrema-direita *Alternative für Deutschland*, apesar de tal agremiação contar com menos de um quarto da preferência nas pesquisas eleitorais. Esse descompasso entre a representação algorítmica e a representação política empírica revela, segundo a Relatora, que o modelo de negócios das plataformas, fundado na monetização do engajamento, incentiva estruturalmente a amplificação de conteúdos prejudiciais.

Talvez a denúncia mais grave do relatório, contudo, refira-se ao que Khan designa como *rollback*, ou retrocesso sistemático das plataformas em seus compromissos com a integridade eleitoral. Desde 2022, observa-se um movimento coordenado de redução de equipes de confiança e segurança, particularmente em Meta, X e Google, acompanhado da reversão de políticas relevantes de moderação. Sob a liderança de Elon Musk, o X retirou-se integralmente da moderação de conteúdo em nome de uma concepção absoluta de liberdade de expressão; o YouTube reverteu políticas que limitavam conteúdos contestatórios dos resultados eleitorais norte-americanos de 2020; e a Meta, no início de 2025, encerrou abruptamente suas parcerias com *fact-checkers* profissionais.

No que diz respeito à inteligência artificial, Khan reconhece que a interferência da IA generativa nas eleições de 2024 não se materializou na escala anteriormente antecipada por analistas, advertindo, contudo, que tal constatação não autoriza complacência. A IA é compreendida em sua dupla face: por um lado, pode aprimorar campanhas políticas legítimas e acelerar a verificação factual, por outro, pode gerar conteúdo enganoso, amplificar desinformação e operar redes coordenadas de *bots*.

Para além dessas limitações, o relatório aponta problemas adicionais. Khan

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

registra que os influenciadores são percebidos por seus seguidores como portadores de uma 'voz autêntica' (§22) — característica que, somada à ausência de obrigações deontológicas descrita em §23, suscita questões sobre a assimetria entre a confiança que mobilizam e a responsabilidade a que estão submetidos. Nas recomendações dirigidas aos partidos políticos, a Relatora alerta para a utilização de influenciadores como '*proxies*' ou '*surrogates*' para discurso prejudicial e desinformação (§107), recomendando que candidatos e partidos se abstenham dessa prática, bem como do recompartilhamento e promoção desse tipo de conteúdo. Já nas recomendações dirigidas às empresas, Khan demanda que as plataformas assegurem que seus sistemas de recomendação não monetizem a disseminação de desinformação ou discurso de ódio por influenciadores (§114) — formulação que pressupõe a existência atual desse mecanismo de monetização.

O relatório A/HRC/59/50 (ONU, 2025) qualifica os influenciadores como proxies de discurso prejudicial, percebidos como "voz autêntica" e desobrigados de deveres deontológicos. O microssistema brasileiro já veda a contratação de influenciadores para promoção de candidatura sob o manto da liberdade de expressão (art. 28, §7º-A), exigindo que a publicidade paga se faça diretamente perante a plataforma. Sustenta-se, contudo, a necessidade de densificar o regime: dever de identificação inequívoca de conteúdo político remunerado, transparência da remuneração e responsabilização solidária da campanha contratante, alcançando o impulsionamento indireto e a publicidade política encoberta.

Em conclusão parcial, é possível afirmar que os sistemas computacionais que mediam o discurso político contemporâneo, sejam eles LLMs conversacionais ou algoritmos de recomendação em redes sociais, operam como agentes ativos de transformação cognitiva e comportamental, com viés sistêmico identificável, mecanismos de influência que escapam às defesas críticas tradicionais e efeitos persistentes que se acumulam tanto na arquitetura informacional quanto na estrutura das redes sociais dos usuários**.** A aparente divergência entre os achados, viés à esquerda no ChatGPT, tendência à Direita Radical em buscadores e em LLMs no contexto europeu e viés conservador em feeds de redes sociais, não

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

implica contradição, mas confirma a tese central: o vetor do viés é idiossincrático de cada sistema, função de sua arquitetura, de seu *corpus* e de seu *fine-tuning*, sendo a não neutralidade, e não a direção, o dado estruturalmente relevante.

A regulação da inteligência artificial constitui a inovação mais significativa do ciclo recente de reformas, materializada em microssistema específico introduzido pela Resolução nº 23.732/2024 (Brasil, 2024). O artigo 9º-B consagra o princípio da transparência algorítmica, ao impor ao responsável pela propaganda o dever de informar, de modo explícito, destacado e acessível, sempre que utilizar conteúdo sintético multimídia gerado por inteligência artificial para criar, substituir, omitir, mesclar, alterar a velocidade ou sobrepor imagens e sons. Trata-se de obrigação de rotulagem que opera como pressuposto de licitude da propaganda assistida por IA. A norma destaca expressamente que o uso de *chatbots*, avatares e conteúdos sintéticos como artifício para intermediar a comunicação de campanha com pessoas naturais submete-se à transparência algorítmica, vedada qualquer simulação de interlocução com a pessoa candidata ou outra pessoa real.

O artigo 9º-C institui vedação absoluta à utilização, na propaganda eleitoral, de conteúdo fabricado ou manipulado para difundir fatos notoriamente inverídicos ou descontextualizados com potencial para causar danos ao equilíbrio do pleito ou à integridade do processo eleitoral. A norma alcança frontalmente o fenômeno das *deepfakes* eleitorais[2], podendo configurar abuso dos meios de comunicação, com possível cassação do registro ou do mandato.

Os artigos 9º-D e 9º-E ampliam o eixo da responsabilização, deslocando-a da figura individual do candidato para as plataformas digitais. O artigo 9º-D estabelece o dever de cuidado dos provedores de aplicação de internet, exigindo-lhes medidas preventivas e mitigatórias contra a disseminação de conteúdos desinformativos. Já o artigo 9º-E institui a responsabilidade solidária, civil e administrativa, dos provedores que não removerem imediatamente, durante o período eleitoral, conteúdos com desinformação, discurso de ódio, ideologia nazista

[2] § 1º É proibido o uso, para prejudicar ou para favorecer candidatura, de conteúdo sintético em formato de áudio, vídeo ou combinação de ambos, que tenha sido gerado ou manipulado digitalmente, ainda que mediante autorização, para criar, substituir ou alterar imagem ou voz de pessoa viva, falecida ou fictícia (deep fake).

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

ou fascista, materiais antidemocráticos, racistas, homofóbicos ou misóginos. Em sequência, o artigo 9º-G opera como cláusula sistematizadora do enfrentamento à desinformação.

A Resolução nº 23.755/2026 (Brasil, 2026) ampliou substancialmente esse microssistema, instituindo o chamado *blackout* de IA, proibição de publicação, republicação e impulsionamento de novos conteúdos sintéticos durante as 72 horas anteriores e as 24 horas posteriores ao pleito, bem como a vedação à utilização de provedores de IA para ranquear, recomendar ou priorizar candidaturas, ainda que a pedido do próprio usuário.

Anote-se que o artigo 3º-B, incluído pela Resolução nº 23.671/2021 (Brasil, 2021) e com redação dada pela Resolução nº 23.732/2024 (Brasil, 2024), autoriza o impulsionamento pago durante a pré-campanha, desde que cumpridos cumulativamente quatro requisitos: contratação direta pelo partido ou pré-candidato com o provedor de aplicação (inciso I); ausência de pedido explícito de voto (inciso II); gastos moderados, proporcionais e transparentes (inciso III); e observância das regras aplicáveis ao impulsionamento durante a campanha (inciso IV). A exigência de contratação direta visa coibir a pulverização do gasto eleitoral por meio de intermediários não declarados.

Já o artigo 28, § 7º-A, regula o impulsionamento durante a campanha propriamente dita, permitindo-o exclusivamente para promover ou beneficiar candidatura, partido ou federação que contrata o serviço, sempre com identificação clara da origem do conteúdo[3]. Anote-se que a publicidade paga deve ser feita diretamente perante a plataforma, vedada a contratação de influenciadores para, sob o manto da liberdade de expressão, promover determinado candidato. Verifica-se, portanto, que o Tribunal Superior Eleitoral, no uso de suas atribuições, adota um modelo intervencionista nas manifestações digitais, de forma a buscar equilibrar as forças e evitar a manipulação apta a influenciar a vontade do cidadão.

Não obstante o avanço significativo representado pelo microssistema de regulação da inteligência artificial, e as recentes incursões na limitação da interface

[3] Art. 29. É vedada a veiculação de qualquer tipo de propaganda eleitoral paga na internet, excetuado o impulsionamento de conteúdos, desde que identificado de forma inequívoca como tal e contratado exclusivamente por partidos políticos, federações, coligações, candidatas, candidatos e representantes

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

conversacional e de ranqueamento trazidas pela Resolução nº 23.755/2026, o núcleo duro da disciplina normativa permanece fortemente orientado à rotulagem do conteúdo sintético multimídia (imagens, sons e vídeos). A redação do caput do artigo 9º-B é eloquente ao referir-se expressamente ao âmbito audiovisual, denotando um recorte focado na integridade sensorial da informação.

Dessa opção regulatória decorre uma insuficiência substantiva no enfrentamento da complexidade inerente às modalidades emergentes de utilização da inteligência artificial generativa em contextos eleitorais, notadamente aquelas estruturadas em torno da persuasão automatizada baseada em texto e diálogo. Quatro pontos merecem especial atenção do regulador.

É verdade que o artigo 28, §1º-C veda expressamente aos provedores de aplicação o uso de IA para emitir opiniões, indicar preferências eleitorais ou favorecer/desfavorecer atores políticos. Essa opção, em nosso entender, contempla o *chatbot IA*. Contudo, a redação do dispositivo permite que, ao invés da indicação direta de um voto, o sistema exerça o mesmo papel de forma subliminar ou por meios transversos, induzindo sem estabelecer expressamente uma preferência ou recomendação. Essa lacuna é tecnicamente confirmada por Fisher et al., para quem o mero reconhecimento do viés não o atenua de forma suficiente para neutralizá-lo, o que torna a proibição de indicação explícita de voto insuficiente diante da persuasão implícita mensurada empiricamente. Ademais, coloca-se a questão de como estabelecer uma forma eficaz de fiscalização da aplicação desse dispositivo, sobretudo para sistema de IA que não permitam a transparência do algoritmo utilizado.

Anote-se que, embora a redação do § 1º-C, II do art. 28 alcance expressamente o favorecimento direto ou indireto, inclusive por respostas automatizadas, persiste um déficit estrutural de efetividade: a opacidade dos LLMs e a ausência de obrigações específicas de auditoria algorítmica e explicabilidade tornam a fiscalização do dispositivo praticamente inviável. A norma proíbe, mas não estabelece os mecanismos técnicos e procedimentais que permitiriam detectar e comprovar a violação. Trata-se, portanto, de norma com elevada densidade proibitiva, mas baixa densidade fiscalizatória. Uma proibição sem instrumentos

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

efetivos de verificação.

Embora a técnica legislativa corretamente privilegie o sistema em detrimento do nome da tecnologia, os LLMs apresentam características singulares, como a capacidade de geração textual em escala, a suscetibilidade à produção de falsidades factuais não intencionais ("alucinações") e, principalmente, a maleabilidade discursiva que permite a modulação fina e hiperpersonalizada de mensagens. Tais características demandam mecanismos de governança algorítmica e supervisão humana contínua que vão muito além da simples obrigação de rotulagem.

A constatação dessas insuficiências não importa, evidentemente, em desconsideração do mérito do esforço regulatório empreendido pelo Tribunal Superior Eleitoral, que se posicionou na vanguarda internacional ao disciplinar o uso da inteligência artificial. Importa, antes, no reconhecimento do caráter dinâmico, iterativo e provisório que a regulação tecnológica necessariamente assume. A velocidade com que a inteligência artificial generativa diversifica suas táticas de microdirecionamento impõe ao regulador o dever de revisão periódica do arcabouço normativo, sob pena de esvaziamento da efetividade protetiva.

Sustenta-se, portanto, a necessidade de aprimoramento do microssistema regulatório mediante a incorporação de dispositivos que densifiquem a regulação sobre: (a) obrigações de auditoria contínua dos modelos de linguagem para mitigar vieses e "alucinações" que possam desinformar o eleitorado; (b) parâmetros rigorosos para a interação conversacional automatizada, impondo limites técnicos à hiperpersonalização persuasiva e ao tratamento de dados pessoais inferidos por IA; e (c) estabelecimento de um regime de responsabilidade e limites de gastos específicos para o uso de agentes virtuais autônomos. A implementação dessas diretrizes contribuiria para elevar o princípio da transparência algorítmica.

A persuasão hiperpersonalizada depende de coleta, inferência e tratamento de dados pessoais. Incide, portanto, a Lei nº 13.709/2018 (LGPD), notadamente quanto a dados sensíveis (art. 5º, II), categoria que abrange convicção política, e a dados inferidos — perfis políticos deduzidos por IA a partir de comportamento, ainda que não declarados. A construção de perfis políticos para fins de

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

microdirecionamento eleitoral deve observar os princípios de finalidade, adequação, necessidade e transparência (art. 6º), exigindo base legal específica e, dada a sensibilidade, relatório de impacto à proteção de dados. Defende-se, em consequência, a vedação ao tratamento de dados sensíveis inferidos para personalização persuasiva eleitoral, salvo consentimento qualificado e específico.

5. Considerações Finais

À luz do percurso desenvolvido, é possível concluir que a formação da vontade política do eleitor não pode mais ser compreendida a partir de uma concepção simplificada de autonomia racional plena. A literatura filosófica e neurocientífica examinada indica que a tomada de decisão humana resulta de interação complexa entre processos conscientes e inconscientes, suscetíveis à influência de fatores contextuais, ambientais e informacionais.

Quando esse dado é projetado sobre o ecossistema digital contemporâneo, marcado por inteligência artificial conversacional, curadoria algorítmica, recomendação personalizada de conteúdo e circulação massiva de informações em redes sociais, torna-se evidente que tais tecnologias não operam como meros instrumentos neutros de comunicação, mas como estruturas capazes de interferir, ainda que de modo sutil, persistente e escalável, na formação da opinião e na orientação do comportamento político.

Os estudos empíricos analisados reforçam essa conclusão ao demonstrar que a exposição prévia a determinados conteúdos, a amplificação algorítmica assimétrica, o enviesamento de modelos de linguagem e a autoridade percebida das interfaces conversacionais podem deslocar percepções, prioridades e decisões mesmo quando o usuário não identifica claramente o viés ao qual está submetido. Em contexto eleitoral, esse risco assume especial gravidade, pois a interferência sobre a autonomia informacional do eleitor não afeta apenas interesses individuais, mas compromete a própria integridade do processo democrático e a igualdade de oportunidades entre candidaturas. Por isso, a tutela jurídica do tema não pode limitar-se ao combate pontual de conteúdos

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

manifestamente ilícitos.

Nesse cenário, sustenta-se que o ordenamento brasileiro deve avançar para uma disciplina normativa mais específica sobre o uso de *chatbots* e sistemas baseados em *Large Language Models* (LLMs) em período eleitoral. Não se trata de defender a proibição absoluta desses instrumentos, providência que seria excessiva e potencialmente incompatível com a liberdade de expressão e com a liberdade de iniciativa. O tratamento mais adequado é o de uma regulação intensiva fundada na transparência, na precaução e na proteção da autonomia do eleitor.

O mais relevante é não somente impedir a interferência direta por meio de sugestões objetivas e claras para determinado voto ou posicionamento político. O fundamental é dar efetividade ao disposto no Código Eleitoral (art. 242), não permitindo o emprego de meios publicitários destinados a criar, artificialmente, na opinião pública, estados mentais, emocionais ou passionais. Ou seja, impedir que *chatbots e LLMs* em resposta a pesquisas sobre temas sensíveis (pleito eleitoral, orientação política, histórico de candidatos) forneçam informações que, apesar de não recomendarem o voto direto, contenham vieses relevantes de forma a criar uma influência no eleitor por meio indireto e até mesmo inconsciente. Ademais, as "alucinações" de IA são deveras deletérias nesse tema, na medida em que informações fáticas distorcidas podem ser fornecidas como realidades.

Como exemplo do nosso raciocínio, podemos citar o caso de Deborah Lipstadt (1993), que dedicou toda uma obra para demonstrar os equívocos de teorias negacionistas do Holocausto. Note-se que, quando tratamos de negacionismo, nos referimos não a boatos, mas a obras com caráter científico capazes de induzir um raciocínio ou serem vendidas como verdades decorrentes de pesquisas científicas, a exemplo das manifestações de David Irving (1977) sobre o Nazismo e o Holocausto. Uma "alucinação" de IA ou mesmo um algoritmo com viés poderia, nesse tema, simplesmente formular resposta baseada na obra de David Irving e ignorar a contestação de Deborah Lipstadt, o que, a princípio, tornaria verdadeira uma informação historicamente incorreta, sem que o usuário tenha qualquer pista de que está diante de uma narrativa controvertida e refutada

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

pela historiografia consolidada.

As medidas propostas submetem-se ao teste de proporcionalidade. Quanto à finalidade legítima, visam proteger a integridade do pleito, a igualdade de oportunidades e a autonomia informacional do eleitor — bens de status constitucional. Quanto à adequação, deveres de transparência, auditoria e responsabilização são aptos a mitigar a influência opaca empiricamente demonstrada. Quanto à necessidade, optou-se pelo meio menos restritivo: rejeita-se a proibição absoluta dos instrumentos, privilegiando-se deveres procedimentais e de transparência, que oneram menos a liberdade de expressão. Quanto à proporcionalidade em sentido estrito, distinguem-se três intensidades de intervenção: restrições de conteúdo (reservadas a deepfakes e desinformação, art. 9º-C), que exigem risco concreto à integridade do pleito; deveres de transparência e identificação (menos gravosos); e obrigações de auditoria e responsabilização (incidentes sobre a arquitetura do sistema, não sobre o conteúdo do discurso). Essa graduação preserva o núcleo da liberdade de expressão política, restringindo não o que se diz, mas o modo opaco e automatizado pelo qual se busca induzir a vontade.

Para conferir efetividade às vedações, propõe-se um conjunto operacional de instrumentos, hierarquizados conforme, ao menos, quatro obrigações: (i) identificação visível e permanente da natureza artificial do interlocutor; (ii) aviso prévio sobre riscos de viés, persuasão e inexatidão factual; (iii) disponibilização de informações mínimas sobre critérios de funcionamento relevantes para a interação política, inclusive eventual personalização com base em dados do usuário; e (iv) supervisão humana e regime claro de responsabilização da campanha, partido, coligação ou terceiro contratante pelo conteúdo veiculado. O ponto fulcral não é a recomendação direta nem mesmo a indireta, ambas já formalmente vedadas pelo art. 28, § 1º-C, II da Resolução nº 23.610/2019[4], mas sim a construção de mecanismos efetivos de detecção, auditoria e *enforcement* dessa vedação, hoje

[4] Registre-se que os eixos 'direto/indireto' (art. 28, §1º-C, II) e 'explícito/implícito' não se confundem: a vedação alcança o favorecimento político-eleitoral identificável, ainda que indireto, mas não a influência subliminar ou o viés estrutural que moldam a percepção sem configurar tomada de posição a favor de candidatura determinada. É nesse interstício que se situa o objeto desta análise.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

praticamente inexistentes. Sem instrumentos técnicos e procedimentais de verificação, a proibição corre o risco de operar apenas no plano simbólico.

Cumpre reconhecer, contudo, uma tensão relevante entre a primeira obrigação proposta (identificação visível da natureza artificial do interlocutor) e a evidência empírica produzida por Fisher et al. (2026), segundo a qual o reconhecimento explícito do viés do modelo não atenua seu efeito persuasivo sobre o usuário. Esse achado, de elevado impacto regulatório, indica que a mera rotulagem opera como condição necessária, mas insuficiente, de proteção da autonomia do eleitor. Daí decorre que o núcleo efetivo da tutela não reside no aviso, mas nas obrigações estruturais de transparência sobre critérios de funcionamento e personalização e, sobretudo, na supervisão humana qualificada com responsabilização clara dos agentes envolvidos. A rotulagem cumpre função informacional e de devido processo comunicativo; a efetividade protetiva, todavia, depende de instrumentos que atuem sobre a arquitetura do sistema, e não apenas sobre a percepção do usuário.

Fora do período eleitoral, embora a intensidade regulatória possa ser menor, subsiste o dever de proteção contra formas opacas de manipulação cognitiva. Operadores de sistemas de LLM, plataformas de *feed* de notícias e demais agentes responsáveis por algoritmos de recomendação devem ser juridicamente vinculados à adoção de esforços efetivos para impedir o uso de mensagens subliminares, técnicas de manipulação encoberta e expedientes de indução não transparente do comportamento político.

Do mesmo modo, o conteúdo político deve ser tratado como categoria informacional sensível, a exigir deveres acrescidos de cuidado, auditoria e mitigação de riscos, especialmente para prevenir *deepfakes*, desinformação sintética, enviesamento sistemático de recomendações e personalização abusiva da persuasão. Nesse campo, a neutralidade algorítmica talvez não seja sempre empiricamente alcançável; todavia, quando inexistente, sua ausência deve ser claramente informada ao usuário, de modo que este saiba que está diante de sistema que seleciona, prioriza ou organiza conteúdo segundo critérios não neutros.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

Em síntese, a democracia contemporânea passou a depender não apenas da liberdade formal de expressão, mas também da preservação das condições materiais e cognitivas para uma deliberação política minimamente autônoma. Se algoritmos, modelos de linguagem, sistemas de recomendação e redes de influência possuem capacidade concreta de moldar preferências, reduzir resistência crítica e amplificar assimetrias, o Direito Eleitoral e o Direito Digital precisam responder a esse novo cenário com instrumentos normativos compatíveis com a sofisticação do problema. O desafio não está em interditar a tecnologia, mas em submetê-la a parâmetros de transparência, neutralidade possível, informação qualificada, responsabilidade e proteção da autonomia do eleitor, de modo a assegurar que a inovação não se converta em instrumento silencioso de captura da vontade política.

Cumpre, por fim, registrar os limites do presente estudo e indicar frentes de investigação que permanecem em aberto. O trabalho desenvolveu-se em chave predominantemente teórico-normativa, sem submeter as soluções regulatórias propostas a teste empírico de eficácia, o que constitui limitação inerente à proposta. Ademais, a literatura mobilizada, sobretudo os estudos sobre persuasão e viés em modelos de linguagem, foi produzida em contextos majoritariamente anglófonos, o que, apesar de não invalidar seu uso, recomenda cautela na transposição direta de seus achados ao eleitorado brasileiro, cuja dinâmica informacional possui especificidades culturais, linguísticas e institucionais relevantes.

Abrem-se, assim, ao menos três frentes promissoras de pesquisa futura: (i) avaliação empírica da efetividade da Resolução TSE nº 23.755/2026 a partir dos dados produzidos no pleito de 2026; (ii) realização de estudos experimentais sobre persuasão de LLMs com eleitorado de língua portuguesa; e (iii) desenho de instrumentos técnicos e procedimentais de auditoria algorítmica aplicáveis ao contexto eleitoral, sem os quais a vedação normativa de influência indireta corre o risco de operar apenas no plano simbólico.

Referências

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

BOURDIEU, Pierre. **Sobre a televisão**. Rio de Janeiro: Jorge Zahar Editor, 1997.

BRASIL. Supremo Tribunal Federal (1. Turma). **Habeas Corpus 225180 AgR**. Relator: Ministro Alexandre de Moraes, 27 de março de 2023. Diário da Justiça Eletrônico, Brasília, DF, 3 abr. 2023. Disponível em: https://portal.stf.jus.br. Acesso em: 28 abr. 2026.

BRASIL. Tribunal Superior Eleitoral. **Resolução nº 23.671, de 14 de dezembro de 2021**. Altera a Res.-TSE nº 23.610, de 18 de dezembro de 2019, que dispõe sobre propaganda eleitoral, utilização e geração do horário gratuito e condutas ilícitas em campanha eleitoral. Brasília, DF: Tribunal Superior Eleitoral, 2021. Disponível em: https://www.tse.jus.br/legislacao/compilada/res/2021/resolucao-no-23-671-de-14-de-dezembro-de-2021. Acesso em: 4 maio 2026.

BRASIL. Tribunal Superior Eleitoral. **Resolução nº 23.732, de 27 de fevereiro de 2024**. Altera a Res.-TSE nº 23.610, de 18 de dezembro de 2019, dispondo sobre a propaganda eleitoral. Brasília, DF: Tribunal Superior Eleitoral, 2024. Disponível em: https://www.tse.jus.br/legislacao/compilada/res/2024/resolucao-no-23-732-de-27-de-fevereiro-de-2024. Acesso em: 4 maio 2026.

BRASIL. Tribunal Superior Eleitoral. **Resolução nº 23.755, de 2 de março de 2026**. Altera a Resolução nº 23.610/TSE, de 18 de dezembro de 2019, que dispõe sobre propaganda eleitoral. Brasília, DF: Tribunal Superior Eleitoral, 2026. Disponível em: https://www.tse.jus.br/legislacao/compilada/res/2026/resolucao-no-23-755-de-2-de-marco-de-2026. Acesso em: 4 maio 2026.

COELHO, Jonas Gonçalves. Livre-arbítrio e a relação mente e cérebro em Benjamin Libet. **Principia**, v. 18, n. 1, p. 153-174, 2014. Disponível em: http://hdl.handle.net/11449/135483. Acesso em: 11 jan. 2023.

DAMIÃO, Íris; ALMEIDA, Paulo; FRANCO, João; SANTOS, Nuno; MAGALHÃES, Pedro C.; GONÇALVES-SÁ, Joana. **Evidence of political bias in search engines and language models before major elections**. arXiv, 2026. DOI: arXiv:2603.23474v1 [cs.CY]. Disponível em: https://arxiv.org/abs/2603.23474. Acesso em: 8 maio 2026.

ESTADOS UNIDOS DA AMÉRICA. Suprema Corte. **Nebraska Press Association v. Stuart**. 427 U.S. 539. Washington, DC: Suprema Corte dos Estados Unidos, 1976.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

ESTADOS UNIDOS DA AMÉRICA. Suprema Corte. **Richmond Newspapers, Inc. v. Virginia**. 448 U.S. 555. Washington, DC: Suprema Corte dos Estados Unidos, 1980.

FISHER, Jillian et al. Biased LLMs can influence political decision-making. **arXiv**, 2026. Preprint. Disponível em: https://arxiv.org/abs/2410.06415. Acesso em: 4 maio 2026.

GAUTHIER, Germain; HODLER, Roland; WIDMER, Philine; ZHURAVSKAYA, Ekaterina. The political effects of X's feed algorithm. **Nature**, v. 652, p. 416-423, 9 abr. 2026. DOI: 10.1038/s41586-026-10098-2. Disponível em: https://www.nature.com/articles/s41586-026-10098-2. Acesso em: 1 maio 2026.

HACKENBURG, Kobi et al. The levers of political persuasion with conversational artificial intelligence. **Science**, Washington, DC, v. 390, n. 6777, eaea3884, 4 dez. 2025. DOI: 10.1126/science.aea3884. Acesso em: 1 maio 2026.

HILLS, Thomas T. Neurocognitive free will. Proceedings of the Royal Society B: Biological **Sciences**, v. 286, n. 1902, art. 20190510, 2019. DOI: 10.1098/rspb.2019.0510.

IRVING, David. **Hitler's war**. New York: Viking Press, 1977.

LIBET, Benjamin. Unconscious cerebral initiative and the role of conscious will in voluntary action. **The Behavioral and Brain Sciences**, n. 8, p. 529-566, 1985.

LIPSTADT, Deborah E. **Denying the Holocaust**: the growing assault on truth and memory. New York: Free Press, 1993.

MELLO, Sebastian Borges de Albuquerque. Da psicanálise à neurociência: do fim ao fim da culpabilidade na doutrina ibérica? Uma visão crítica. **Conpedi Law Review**, v. 1, p. 72-100, 2015.

MERTON, Robert; LAZARFELD, Paul. Comunicação de massa, gosto popular e a organização da ação social. In: LIMA, Luiz Costa. **Teoria da cultura de massa**. Rio de Janeiro: Paz e Terra, 2000. p. 109-131.

MOTOKI, Fabio; PINHO NETO, Valdemar; RODRIGUES, Victor. **More human than human**: measuring ChatGPT political bias. Public Choice, v. 198, n. 1, p. 3-23, 2024. DOI: https://doi.org/10.1007/s11127-023-01097-2.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

MUDRIK, Liad et al. Free will without consciousness? **Trends in Cognitive Sciences**, v. 26, n. 7, p. 555–566, 2022. DOI: 10.1016/j.tics.2022.03.005.

NAÇÕES UNIDAS. Conselho de Direitos Humanos. **Freedom of expression and elections in the digital age**: report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan. Genebra: ONU, 11 jun. 2025. Doc. A/HRC/59/50. Disponível em: https://www.ohchr.org/. Acesso em: 1 maio 2026.

PEREIRA, Claudio José Langroiva; FURTADO, Otavio Morais Alves de Souza Oliveira e. Sociedade, violência e segurança pública: investigações criminais e vieses cognitivos. In: SILVA, Marco Antonio Marques da (org.). **Sociedade, Educação e Violência**. São Paulo: Quartier Latin, 2020. p. 91-106.

RUVA, Christine L.; GUENTHER, Christina C. From the shadows into the light: how pretrial publicity and deliberation affect mock jurors' decisions, impressions, and memory. **Psychology Sarasota Manatee Campus Faculty Publications**, n. 12, 2015. Disponível em: https://digitalcommons.usf.edu/psy_facpub_sm/12. Acesso em: 28 abr. 2026.

SORAL, Wiktor et al. Exposure to hate speech deteriorates neurocognitive mechanisms of empathy. **Scientific Reports**, v. 13, art. 5119, 2023. Disponível em: https://www.nature.com/articles/s41598-023-31146-1. Acesso em: 21 abr. 2026.

SCHURGER, A., Hu, P. B., Pak, J., & ROSKIES, A. L. What is the readiness potential? **Trends in Cognitive Sciences**, 25(7), 558–570, 2021.

THALER, Richard H.; SUNSTEIN, Cass R. **Nudge**: the final edition. New Haven: Yale University Press, 2021.

VASWANI, Ashish et al. Attention is all you need. In: ADVANCES INNEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS), 30., 2017,Long Beach. Proceedings [...]. [s.l.]: Curran Associates, 2017.p. 5998-6008.

YE, Jinyi; LUCERI, Luca; FERRARA, Emilio. Auditing political exposure bias: algorithmic amplification on Twitter/X during the 2024 U.S. Presidential Election. In: **ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY**, 2025, Athens. Proceedings [...]. New York: Association for Computing Machinery, 2025. p. 2349-2362. DOI: 10.1145/3715275.3732159.

REMUNOM
ISSN 2178-6925

Received:01/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/8cbpz192
Pages: 1-40

Disponível em: https://doi.org/10.1145/3715275.3732159. Acesso em: 1 maio 2026.