



RIO

PUC

Dissertação de Mestrado

Conexões entre a desinformação político-eleitoral e a moderação de conteúdo no Brasil:

Impactos sobre a democracia

Gabriela Pavão Pereira

Pontifícia Universidade Católica do Rio de Janeiro
Centro de Ciências Sociais
Departamento de Comunicação

Rio de Janeiro, 30 de setembro de 2025



Pontifícia
Universidade
Católica do
Rio de Janeiro

Dissertação de Mestrado

Conexões entre a desinformação político-eleitoral e a moderação de conteúdo no Brasil: Impactos sobre a democracia

Gabriela Pavão Pereira

Orientação: Professor Marcelo Alves dos Santos Junior

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Comunicação pelo programa de Pós-Graduação em Comunicação, no Departamento Comunicação

Rio de Janeiro, 30 de setembro de 2025



Pontifícia
Universidade
Católica do
Rio de Janeiro

Conexões entre a desinformação político-eleitoral e a moderação de conteúdo no Brasil:

Impactos sobre a democracia

Gabriela Pavão Pereira

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Comunicação Aprovada pela Comissão examinadora abaixo:

Professor Marcelo Alves dos Santos Junior

Orientador

Departamento Comunicação - PUC-RIO

Professor Arthur Cezar de Araujo Ituassu Filho

Departamento Comunicação - PUC-RIO

Professor Juliana Gagliardi de Araujo

Universidade Federal Fluminense

Rio de Janeiro, 30 de setembro de 2025



Pontifícia
Universidade
Católica do
Rio de Janeiro

Todos os direitos reservados. A reprodução, total ou parcial, do trabalho é proibida sem autorização da universidade, da autora e do orientador.

Gabriela Pavão Pereira

Graduou-se em Relações Internacionais na Universidade Federal do Rio de Janeiro (UFRJ) em 2021. Participou de congressos na área de comunicação e materialidades e do Grupo de Pesquisa em Comunicação, Dados e Tecnologias Digitais.

Ficha Catalográfica

Pereira, Gabriela Pavão

Conexões entre a desinformação político-eleitoral e a moderação de conteúdo no Brasil : impactos sobre a democracia / Gabriela Pavão Pereira; orientação: Marcelo Alves dos Santos Junior. – 2025.

67 f.; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Comunicação Social, 2025.

Inclui bibliografia

1. Comunicação Social – Teses. 2. Desinformação. 3. Moderação. 4. Regulação. 5. Plataformas. 6. Brasil. I. Santos Junior, Marcelo Alves dos. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Comunicação Social. III. Título.

CDD: 302.23



Pontifícia
Universidade
Católica do
Rio de Janeiro

Agradecimentos

Agradeço aos meus pais, Otavio e Flavia, por serem sempre meus maiores apoiadores e parceiros em qualquer situação e por sempre me incentivarem a ir mais longe, com muito amor e sabedoria. Aos meus avós e padrinhos, nada disso seria possível sem vocês.

Agradeço ao meu orientador Marcelo Alves, que me guiou nessa trajetória acadêmica que é o Mestrado, ensinou tanto e ainda ensinará diversos outros privilegiados como eu. Aos professores do PPGCOM PUC-Rio e ao professor Mauro Porto da Tulane University, agradeço pelos ensinamentos e trocas importantes.

Agradeço aos amigos de Mestrado, especialmente aos amigos do Grupo de Pesquisa Condado, sem os quais a trajetória de mestrado certamente teria sido muito mais solitária. Poder dividir reuniões, comemorações, conversas, eventos e laboratórios com vocês, é uma honra.

Ao meu parceiro especial de todos os momentos, Rômulo, por estar sempre ao meu lado e compartilhar sonhos.

Aos meus amigos (irmãos, na verdade) com os quais tenho compartilhado a vida faz um bom tempo e levo no meu coração. Obrigada por sempre estarem comigo e fazerem a diferença durante este processo.

Por fim, agradeço à CAPES pelas bolsas e pela oportunidade única de intercâmbio acadêmico que foram fundamentais para a minha dedicação à pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.



Pontifícia
Universidade
Católica do
Rio de Janeiro

Resumo

Pereira, Gabriela Pavão; Junior, Marcelo Alves dos Santos. **Conexões entre a desinformação político-eleitoral e a moderação de conteúdo no Brasil: impactos sobre a democracia.** Rio de Janeiro, 2025. 67p. Dissertação de Mestrado – Departamento de Comunicação Social da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

O presente trabalho dedica-se a apresentar uma revisão bibliográfica referente à análise do fenômeno da desinformação política-eleitoral e da moderação de conteúdo no contexto político-eleitoral brasileiro. Para tal, é traçado um breve histórico e utiliza-se o recorte temporal das décadas de 2020 e 2025. Aborda-se o papel essencial desempenhado pela moderação de conteúdo nas plataformas digitais no impacto da desinformação política, além do enfoque no combate à desinformação e no processo de regulação das plataformas digitais, em seus variados formatos. Para compreender a relação entre os temas abordados, propõe-se a realização de uma Revisão Sistemática de Literatura. A pesquisa não se limita a um campo científico específico, mas busca analisar as produções de diferentes epistemes, para chegar às conclusões.

Palavras-chave

Desinformação; Moderação; Regulação; Plataformas; Brasil; Política.

Abstract

Pereira, Gabriela Pavão; Junior, Marcelo Alves dos Santos (Advisor). **Connections between political-electoral disinformation and content moderation in Brazil: impacts on democracy** Rio de Janeiro, 2025. 67p. Dissertação de Mestrado – Departamento de Comunicação Social da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

The present work presents a bibliographic review regarding the analysis of the phenomenon of political disinformation and content moderation in the Brazilian political-electoral context. Considering that, a brief history is presented and the time frame from 2020 and 2025 is used. The crucial role played by content moderation in digital platform against the dissemination of political-electoral disinformation and as part of the process of digital platform regulation, through its various formats, is addressed. In order to understand the themes presented, a systematic literature review is proposed. The research is not limited to a specific scientific field, analyzing productions from different epistemologies in order to reach its conclusions.

Keywords

Disinformation; Moderation; Regulation; Platforms; Brazil; Politics.

Sumário

1	Introdução	10
2	Desinformação	14
2.1	O que é Desinformação?.....	14
2.2	O Impacto das Plataformas Digitais	16
2.2.1	A Desinformação em Contextos Eleitorais.....	20
2.2.2	Debate Jurídico Brasileiro sobre Desinformação	23
2.3	Desinformação e Democracia nas Plataformas Digitais.....	25
2.4	Governança nas Plataformas Digitais	28
2.4.1	Projeto de Lei (PL) 2630/20	30
3	Moderação de Conteúdo.....	33
3.1	O Que é Moderação de Conteúdo	33
3.2	Histórico da Moderação.....	37
3.3	Repertório da Moderação.....	39
4	Metodologia	43
4.1	Processo de Seleção.....	45
4.2	Processo de Avaliação	47
4.3	Limitações do Estudo	48
5	Resultados	50
5.1	Análise de Resultados.....	50
5.2	Discussão dos Resultados	52
5.3	Análise e Discussão dos Artigos	52
5.4	Discussão Nacional sobre a Desinformação Política	53
5.5	Discussão Nacional sobre a Moderação de Conteúdo e Plataformas	55
5.6	Discussão Nacional sobre Legislações	57
6	Considerações Finais	58
7	Referências Bibliográficas.....	61

Lista de Quadros

Quadro 1: Resultados de coletas por base	45
Quadro 2: Termos mobilizados na busca	47
Quadro 3: Mapeamento das áreas de conhecimento.....	50
Quadro 4: Estudos selecionados e respectivos anos de publicação	51

1 Introdução

O propósito desta introdução é apresentar os temas e especificidades que serão analisados ao longo da presente dissertação, além de explicitar a escolha de divisão de capítulos.

Trata-se de uma pesquisa que introduz as dimensões do problema da desinformação política nas plataformas digitais a partir da revisão de literatura acadêmica. São abordadas as características e ao impacto do fenômeno da desinformação político-eleitoral e à moderação de conteúdo no Brasil. Para tal, são trabalhados também conceitos indispensáveis ao estudo dos temas evidenciados, como: governança, plataformas digitais, democracia e política brasileira.

Na sociedade de plataformas, o fluxo de informações ocorre em tempo real — o que, embora facilite a pluralização de vozes e o acesso ao conhecimento, também contribui para o aumento da circulação de notícias falsas e enganosas, sobretudo durante períodos eleitorais. A desinformação, potencializada pelas plataformas digitais, tem se tornado uma questão, influenciando decisões políticas, comportamentos sociais e eleições.

Partindo do pressuposto que falta transparência sobre os critérios de regulação adotados nas plataformas mencionadas, o trabalho fortalece o argumento de que a ausência de clareza dificulta o combate à desinformação nas plataformas e afeta a integridade democrática.

O problema de pesquisa e os temas escolhidos são atuais. A pergunta de pesquisa é a seguinte: “Como a moderação de conteúdo nas plataformas digitais influencia a circulação e o impacto da desinformação política no Brasil, considerando períodos eleitorais?”. O problema de pesquisa em questão consiste em analisar detalhadamente a relação entre as políticas de moderação de conteúdo das plataformas digitais e o a propagação de desinformação política no ambiente virtual. A partir dessa análise, busca se entender a circulação da desinformação afeta a integridade democrática, sobretudo considerando contextos eleitorais. Considera-se como contexto, a contradição existente entre o discurso veiculado pelas plataformas e que se observa na prática.

O recorte temporal escolhido para a análise em questão foi o período recente de 5 anos da última década, de 2020 a 2025. Para a escolha desse recorte, considerou-se a evolução tecnológica das plataformas como meio de expressão e veiculação de

informações, o avanço das discussões e legislações sobre a moderação de conteúdo, eleições recentes e a ressignificação do conceito de desinformação.

A fim de compreender o tema abordado, utiliza-se o conceito de plataformas digitais, conforme elaborado por Tarleton Gillespie no seu trabalho *Custodians of the Internet* (2018), de governança no meio digital, além de uma extensa introdução à moderação de conteúdo e seus diversos formatos e modalidades.

O texto foi estruturado a partir da seguinte divisão: introdução; dois capítulos teóricos, abordando os conceitos de desinformação e moderação de conteúdo; um capítulo metodológico; um capítulo de resultados e conclusão. A seção de introdução contempla a estrutura da dissertação, além da apresentação dos temas que serão explorados, a justificativa de pesquisa e os objetivos.

O segundo capítulo traça, ao longo de seus subtítulos, um panorama sobre o tema da desinformação. Ao longo dele, realiza-se uma análise empírica das suas principais características. Para além delas, seu histórico no cenário político brasileiro e suas dinâmicas relacionadas às plataformas digitais.

Ao longo do capítulo teórico sobre desinformação, aborda-se primeiramente a desinformação política anteriormente à era digital, considerando que não é algo inédito, mas presente ao longo de diversos processos históricos. A ideia central do capítulo teórico é contextualizar algumas das principais discussões a respeito do tema.

Em seguida, relaciona-se desinformação e democracia nas plataformas digitais. Busca-se compreender como a desinformação política impacta a democracia a curto e longo prazo por meio das plataformas, contemplando estratégias, características e potencial de alcance.

Também é apresentado o conceito de governança e a sua aplicabilidade às plataformas digitais, além da definição do que é um programa de governança. Ademais, são explicadas as principais características do fenômeno da desinformação, a partir de aspectos sócio-políticos.

Já o terceiro capítulo, explora a fundamentação teórica sobre a moderação de conteúdo, reforçando a contribuição da mesma para o desmantelamento da infraestrutura de desinformação e contemplando seu propósito, principais características, repertório, legislações e contexto histórico.

Apresenta-se uma visão geral acerca das formas de moderação e dos responsáveis pela sua implementação, seja ela *soft* ou *hard*, manual ou automatizada; e uma visão

crítica, a partir do argumento de que a ausência de clareza dos critérios adotados dificulta o combate à desinformação.

O quarto capítulo apresenta a metodologia utilizada para a obtenção de resultados. Desta forma, descreve as etapas do protocolo da revisão sistemática da literatura, as bases de dados consultadas, critérios de seleção dos estudos e as etapas de seleção e análise dos artigos. A unidade de pesquisa deste estudo é composta por artigos científicos selecionados em bases de dados indexadas. A RSL é realizada de acordo com os parâmetros estabelecidos pelo *Guia Para Realizar uma Revisão Sistemática da Literatura*¹. Adota-se uma abordagem abrangente, de modo a incluir materiais relevantes, visando analisar e sintetizar obras já publicadas e conhecimentos sólidos.

Pretende-se analisar o conhecimento disponível e apresentar o vasto repertório de estudos de desinformação e de moderação de conteúdo online de forma inovadora em uma dissertação acadêmica brasileira. Busca-se, portanto, obter novas descobertas e reflexões acerca dos temas abordados.

O quinto capítulo, por sua vez, apresenta os resultados obtidos a partir da revisão sistemática de literatura, além de discutir os principais achados do estudo. A seção de discussão dos resultados não se limita a repetir as descobertas feitas a partir dos textos analisados durante a revisão sistemática. Durante a discussão de resultados, pretende-se interpretar, comparar e explicar os achados da Revisão Sistemática da Literatura à luz da literatura.

Por fim, no sexto capítulo, as conclusões são devidamente apresentadas, sintetizando o que foi exposto ao longo do texto para os leitores e indicando possíveis caminhos para novas pesquisas e investigações.

Como objetivos específicos, destaca-se, primeiramente, a compreensão de como a moderação de conteúdo impacta a desinformação política, e como essa dinâmica é amplificada durante os períodos eleitorais. A partir da análise e dos resultados encontrados, pretende-se apresentar um panorama contextualizado sobre a desinformação e as características da moderação de conteúdo no contexto político brasileiro. Como objetivos secundários, destaca-se o aprofundamento e a expansão do conhecimento existente acerca dos temas analisados elucidando o que já foi produzido sobre o tema

¹ OKOLI, Chitu. Guia para realizar uma revisão sistemática da literatura. Tradução de David Wesley Amado Duarte; Revisão técnica e introdução de João Mattar. *eaD em Foco*, 2019;9 (1): e748. DOI: <https://doi.org/10.18264/eadf.v9i1.748>

Como justificativa para a escrita da dissertação, pretende-se contribuir de forma relevante e enriquecedora para o campo de estudos da Comunicação e dos Estudos de Plataforma no Brasil. Ademais, destaca-se o propósito de apresentar a revisão e síntese de um extenso corpo de literatura sobre o tema à sociedade brasileira, possibilitando avanços e preenchendo possíveis lacunas que justificam a condução da revisão sistemática.

Aborda-se ainda o contexto político brasileiro e a relação existente entre democracia e desinformação, considerando que vem se intensificando ao longo da última década. Essa relação, da qual se falará adiante, é impulsionada pela evolução da tecnologia e pelas novas possibilidades de expressão criadas a partir das plataformas. A compreensão dessas dinâmicas e dos contextos histórico, político e social brasileiros é fundamental para analisar o fenômeno abordado e suas particularidades.

O crescimento da quantidade de estudos sobre desinformação, comunicação, política, moderação e plataformas digitais é fundamental para gerar um debate qualificado sobre desinformação no país, evitando generalizações e reducionismo. Tais estudos relevantes e atuais contribuem para o reconhecimento da complexidade dos temas. O intuito dessa dissertação é, dentro deste contexto, mostrar-se elucidativa e gerar repercussões e reflexões relevantes.

2 Desinformação

2.1 O que é Desinformação?

Este capítulo teórico propõe-se a apresentar o conceito de desinformação, a partir de uma análise das suas principais características, histórico e forma como se manifesta em contextos eleitorais. São abordadas discussões centrais ao campo dos estudos de desinformação, elaborando a respeito do seu significado e das suas dimensões político-culturais para além da comunicação.

Ao longo das próximas subseções, discute-se a desinformação para além da dualidade de “verdade” e “falsidade”, a fim de evitar uma reducionista, que não dá conta das possibilidades híbridas ou conteúdos enganosos por omissão. É preciso compreender a desinformação como um fenômeno para entender a seriedade do risco que representa para a sociedade para além do discurso padrão e, assim, estudá-la apropriadamente.

Pode-se definir a desinformação como um fenômeno sistêmico (Recuero, 2024). A desinformação, embora seja um fenômeno decorrente da comunicação, é um sistema capaz de desestruturar o sistema social, geralmente com o intuito de beneficiar um indivíduo, grupo político ou sistema. Pode tomar diversas formas: descrédito de valores consolidados, como a defesa da democracia; influência sobre o comportamento de indivíduos e obtenção de lucro econômico (Recuero, 2024). Entendê-la é fundamental para se pensar em como lidar com ela.

Conforme apontam D’Andréa e Henn (2021), o fenômeno da desinformação não se trata de algo inédito, considerando sua presença em diversos processos históricos ao longo dos séculos. Como ressalta Albuquerque (2024), a desinformação como recurso para se obter vantagens não é um fenômeno novo, mas vem adquirindo novas dimensões. Fatores como a lógica de operação das plataformas e o uso de agentes não humanos contribuem para o crescimento da atividade desinformacional.

Ao longo da última década, uma quantidade crescente de disciplinas e campos do conhecimento têm se voltado para os estudos de desinformação, como Ciência da Computação, Direito, Comunicação, Ciências Sociais e Direito, por meio de abordagens multidisciplinares, transdisciplinares e interdisciplinares.

Para além do sentido denotativo, quando se pensa em desinformação é comum deparar-se com a noção de desordem informacional (Wardle; Derakhshan, 2017). Esta noção compreende três subcategorias. São elas as seguintes: desinformação, má informação e informação maliciosa ou malinformação.

Pensando nas dinâmicas próprias do fenómeno desinformacional, Wardle e Derakhshan (2017) propõem três categorias no relatório *A Multi-dimensional Approach to Disinformation - Report of the Independent High Level Group on Fake News and Online Disinformation*. A partir destas se examinaria os diferentes tipos de “desordem informativa”.

Primeiramente, há a informação errada (*misinformation*), quando uma informação falsa é partilhada, mas sem intenção de causar prejuízos. Já a desinformação (*disinformation*), consiste na informação falsa partilhada conscientemente com o intuito de causar danos a uma pessoa, grupo social, organização ou país. Há ainda a informação maliciosa (*mal-information*), nos casos em que uma informação baseada na realidade é partilhada para causar danos, usada para impor prejuízos a uma pessoa, organização ou país (Wardle; Derakhshan, 2017, p. 20; Dourado, 2020, p.44).

De acordo com o relatório, a desinformação seria utilizada para distorcer fatos, prejudicando uma ou mais pessoas. Essa manipulação ou mesmo alteração da realidade é veiculada nas redes sociais e consegue atingir uma parcela considerável da população.

A desinformação pode ser definida ainda como:

Falsidades ou distorções intencionais, frequentemente espalhadas como notícias, para impulsionar objetivos políticos como descreditar oponentes, perturbar debates políticos, influenciar eleitores, inflamar conflitos sociais existentes ou criar um cenário geral de confusão e paralisia informacional (Bennet; Livingston, 2021, p.3).

Um dos maiores desafios contemporâneos, internacional e nacionalmente, é a circulação da desinformação, sendo este um tema que permeia debates públicos, políticos e académicos. A definição de desinformação apresentada neste relatório é aquela abordada ao longo desta dissertação para tratar do fenómeno desinformacional, sendo aquela que pode ser identificada em contextos eleitorais e nas plataformas digitais.

2.2 O Impacto das Plataformas Digitais

Para compreender o cenário político brasileiro contemporâneo, é necessário introduzir o impacto das plataformas digitais. O tema será melhor explorado nos próximos subtítulos, mas, para os fins deste estudo, considera-se como plataformas digitais “sites e serviços que apresentam expressão pública, a armazenam e a oferecem na nuvem, organizam o acesso a isso por meio de busca e recomendação ou o instala em dispositivos móveis” (Gillespie, 2017, p.1).

O termo “plataforma” abrange diversas categorias que aparentemente não possuem relação entre si: YouTube, Instagram, Whatsapp, Telegram, Google Search, Apple App Store e X são apenas alguns exemplos. Estudar elementos digitais significa também ter em mente que a estrutura de cada plataforma é única, com dinâmicas e regras próprias. Assim, não são iguais ou completamente comparáveis.

A presente dissertação opta por focar no aspecto computacional do conceito de plataformas. Neste contexto, “aspecto computacional” é utilizado para referir-se à tecnologia das plataformas: geração de conteúdo, tratamento de dados e análise de plataformas digitais, como blogs, canais de vídeos e redes sociais.

As pessoas usam a internet principalmente para se conectarem umas com as outras e buscar entretenimento, além de usos voltados para trabalho e compras. Informar-se está em segundo plano (Altay; Berriche; Acerbi., 2023, p.3). Cada vez mais as plataformas se tornam infraestruturas cívicas indispensáveis, estendendo seus serviços a atividades sociais relevantes, como a comunicação, processos democráticos e acesso a conteúdos informativos.

As plataformas digitais tornaram-se indispensáveis para compreender o processo político brasileiro desde 2010. Considera-se 2010 um marco, uma vez que foi o ano no qual se autorizou a realização de campanhas políticas utilizando a internet no Brasil. A Resolução 23.191, expedida pelo Tribunal Superior Eleitoral (TSE) em 2009, direcionava-se à propaganda eleitoral nas eleições gerais de 2010. A lei permitiu o uso da internet para propaganda gratuita., tanto em sites quanto em blogs².

Os candidatos ou qualquer pessoa podem manter blogs, sites de mensagens instantâneas e páginas nas redes de relacionamento, como Facebook e X, durante o período eleitoral. Havia restrição para sites de pessoas jurídicas, com ou sem fins

² Fonte: Agência Senado

lucrativos, e sites oficiais ou hospedados por órgãos ou entidades da administração pública direta ou indireta da União, dos Estados, do Distrito Federal e dos Municípios. A implementação da lei impactou significativamente as dinâmicas eleitorais nos anos seguintes.

Alguns anos depois, houve o movimento conhecido popularmente como junho de 2013, que envolveu mobilizações em massa em diversas cidades brasileiras que aconteceram simultaneamente. Desde esse episódio, as mídias sociais tornaram-se uma variável relevante de análise para estudos políticos contemporâneos em nosso país” (Recuero; 2017; p.228).

A internet contribui significativamente para a pluralização de vozes, para a descentralização da comunicação e, conseqüentemente, para o aumento da participação política de indivíduos, organizações e movimentos sociais. O meio digital facilitou a veiculação de opiniões, o encontro de pessoas com posicionamentos e valores semelhantes, o surgimento de novos atores e o fluxo de distribuição de informação (Jungherr e Schroeder, 2021, p.3). Os espaços de circulação da informação são amplamente disputados por atores com interesses diversos em meio a um momento de crise epistemológica das instituições consolidadas (Oliveira, T., 2020, p.22) em torno da produção de verdade.

Contudo, simultaneamente, o ambiente virtual propicia a deflagração da desinformação, considerada uma das principais ameaças atuais à democracia, preocupando governos, legisladores, organizações da sociedade civil e indivíduos. O novo ambiente social e tecnológico no qual esse fenômeno se propaga atualmente traz à tona novos problemas e questões político-sociais.

Um fator essencial para compreender a deflagração da desinformação na sociedade brasileira é a crise de autoridade das instituições (Bennett; Livingston, 2021, p.4). Sem legitimidade, perdem força e controle da narrativa central no debate público, ou seja, falham em manter seu papel como *gatekeepers*. A falta de credibilidade associada a estas instituições gera então um vácuo que propicia o surgimento de novos atores e de fatos alternativos.

Cabe apontar também que a descentralização da comunicação não equivale à democratização da internet e que grandes empresas como Google, X, Facebook, Apple e Microsoft monopolizam e intermediam boa parte das interações online. A noção de que há uma profunda correlação entre a democracia e o meio digital surge a partir da difusão da “Ideologia Californiana” (Barbrook; Cameron, 1996 apud Ride; Wood, 2018), criada

no Vale do Silício na Califórnia (EUA) na década de 1990. A Ideologia Californiana pressupunha a internet era aberta, livre e acessível

Estes intermediários como Google, YouTube e Meta, são ainda hoje associados aos ideais e princípios atribuídos a essa corrente ideológica dos anos 1990. São eles: transparência, igualdade, progresso, potencialidade democrática, determinismo tecnológico antiestatal e neutralidade. A chamada Ideologia Californiana combinava os valores culturais propostos pelo movimento hippie ao zelo empreendedor das indústrias de alta tecnologia do Vale do Silício. Dessa forma, resultando em um discurso liberal sobre o potencial emancipatório do crescimento das tecnologias de informação.

Apesar de ter se originado na costa oeste norte-americana, essa visão foi propagada em escala mundial em paralelo à expansão da internet. Livros, revistas, programas televisivos, websites e discursos políticos contribuíram para esse fenômeno. Assim, tornou-se uma corrente ideológica popular na época.

Hoje, a Ideologia Californiana é percebida majoritariamente com ceticismo e críticas, sobretudo em razão de diversos crimes virtuais e mal-uso da internet com o passar dos anos. São alguns deles: cyberbullying; uso indevido de dados; propagação de informações falsas; incitação à violência, dentre outros.

Entretanto, mesmo no cenário crítico atual descrito, a segmentação direcionada ou segmentação nichada (*microtargeting*), vem ganhando força. É comum a customização de anúncios voltados para públicos específicos por meio de *embeds*³ oferecidos por Google, Meta e X. Um exemplo foi o uso destes para a estratégia digital de campanha de Donald Trump (Mello, 2020), algo que contribuiu para o destaque de Donald Trump nas eleições presidenciais de 2016 dos Estados Unidos.

Durante a campanha do candidato, houve uma segmentação, alcançando milhares de usuários sem que eles soubessem com o que haviam consentido. Além dessa segmentação, houve uma campanha massiva de espalhamento de desinformação e notícias enganosas nas redes sociais, relacionadas a promessas de campanhas, teorias conspiratórias e ataques a outros candidatos.

Após o escândalo em questão, o tema da desinformação ganhou destaque como pauta e passou a ser amplamente discutido por meio de fóruns internacionais, novas legislações, artigos acadêmicos e debates governamentais. O julgamento dos envolvidos

³ *Embeds* são mensagens que contém e/ou mídias de terceiros adicionadas pelos usuários. Podem assumir diferentes formatos, como: postagens, infográficos ou podcasts, por exemplo.

no ocorrido ganhou as manchetes internacionais e originou discussões sobre privacidade, uso indevido de dados, transparência e necessidade de maior regulação do ambiente digital.

Conforme aponta Jonas Valente (Valente. In: Maurício; Saback, 2024, p.136), o crescimento do problema e as suas consequências, tanto para a democracia quanto para a vida dos cidadãos, geraram uma necessidade de soluções e respostas em diversos países. Assim, criou-se uma nova onda de demandas por soluções regulatórias eficazes que contemplassem e solucionassem efetivamente os problemas gerados pela desinformação.

Em relação ao contexto brasileiro, é preciso ser cauteloso com a abordagem utilizada para se falar sobre a desinformação no cenário sócio-político digital do país. Apesar de haver o uso contínuo de estratégias desinformativas antiéticas e tentativas de manipulação da opinião pública, é preciso ter o cuidado de não tratar os cidadãos brasileiros como seres apolíticos altamente influenciáveis. Falta forte evidência empírica que sugira ou comprove a desinformação como um fator decisivo para a construção do posicionamento político.

Como apontam Jungherr e Schroeder:

Ainda que assumamos que a desinformação tem uma difusão muito mais ampla do que a evidência empírica atual, está longe de ser claro que a desinformação convence as pessoas de posições políticas que elas não tinham anteriormente. (Jungherr e Schroeder, 2021, p.3).

Os autores defendem que os indivíduos frequentemente já estão predispostos a acreditar em um determinado discurso por questões de alinhamento político, ou seja, por se identificarem com ele ou por compartilharem valores e crenças de um determinado candidato.

Ademais, compartilhar e curtir um conteúdo não evidencia necessariamente identificação. Há outras razões motivadoras, como: socializar, ironizar, expressar ceticismo, rir ou expressar indignação. Não é possível medir o impacto da desinformação apenas a partir de métricas de engajamento (Altay; Berriche; Acerbi., 2023, p.6).

2.3 A Desinformação em Contextos Eleitorais

Na América Latina e, mais especificamente, no Brasil, a difusão de notícias falsas por parte de representantes políticos e meios de comunicação sensacionalistas não é uma novidade exclusiva da era da internet. No Brasil, o uso de contas falsas como tática para aumentar a visibilidade pública e desviar a atenção de certos assuntos, por exemplo tem sido observado desde 2010 (Dourado, 2020).

É imprescindível compreender o contexto brasileiro e sua articulação com processos históricos, sócio-políticos e culturais para entender a dinâmica do fenômeno da desinformação no país. Há diversos episódios de desinformação ao longo da história, seja voltada para desqualificar forças opositoras, favorecer candidatos ou para favorecer movimentos políticos. (Valente, 2024). O discurso de ódio, o apelo ao medo e o sensacionalismo são incorporados como elementos das narrativas que influenciam a opinião pública ao longo do tempo. A polarização política contribui para a criação de um clima favorável à propagação e até mesmo a validação de informações falsas (Dourado, 2020).

No que diz respeito à esfera política, a desinformação está presente de diversas formas para além do meio digital. São algumas delas: fraudes de urnas e resultados de eleições; mentiras inventadas para danificar a imagem de candidatos de oposição ou tentar moldar a opinião pública; notícias fictícias para sustentar regimes políticos, conforme ocorrido durante a ditadura militar; corrupção; omissão de dados.

Entretanto, com o crescimento das plataformas digitais, a implementação da lógica algorítmica de consumo aplicada aos conteúdos digitais, associada à economia da atenção (Bentes, 2021), vem adquirindo força nos últimos anos. A viralização e a monetização de conteúdos, dentre outras estratégias, conferiram um novo peso ao fenômeno da desinformação política.

Surgiram e ganharam popularidade fóruns e comunidades, utilizados com os seguintes fins desinformativos por grupos e atores políticos: novas narrativas digitais; difusão de informações equivocadas ou distorcidas e de inverdades, devido ao potencial de alcance ampliado; ataques ao sistema eleitoral; impulsionamento de conteúdo político

fora do período eleitoral; fortalecimento de teorias falsas; ataques online e difusão de discursos de ódio; *trolls*⁴; criação de *bots*⁵ ou *sock puppets*⁶.

Pensando no contexto político-social brasileiro, a desinformação, sobretudo em períodos eleitorais, constitui uma ameaça. Esta é utilizada como estratégia política para impactar o posicionamento dos indivíduos, por meio da supressão de votos, do incentivo à polarização e à inflamação de grupos extremistas, ainda que não seja possível comprovar empiricamente que molde posições políticas de forma determinante.

As plataformas, devido ao seu potencial de alcance, são um amplo espaço de discussão democrática, para busca e circulação de informações e emissão de opiniões. Paralelamente, é propícia à difusão de informações falsas, que podem causar danos à democracia. Esta dualidade é um fator levado em conta na escolha de canais utilizados pelos políticos brasileiros.

Nacionalmente, eventos como as eleições presidenciais de 2018 também contribuíram para o impulsionamento da discussão. Em meio ao contexto da eleição de Jair Bolsonaro em 2018 e da pandemia de covid-19 entre 2020 e 2022, a agenda de pesquisa sobre a desordem informacional se consolidou no Brasil.

Nas eleições presidenciais de 2018, na qual foi eleito o então candidato Jair Bolsonaro, foi possível observar o impacto do uso das redes sociais para fins desinformativos. Dentre as diversas formas, destaca-se: o espalhamento de notícias falsas; a propagação de discurso de ódio; o uso de bots e perfis falsos; disparos de mensagens automatizadas de forma ilegal via aplicativos de mensagens.

A polarização política no ano de 2018 foi instigada pela ansiedade acerca da prisão ou possibilidade de candidatura de Luís Inácio Lula da Silva à presidência da República, somado a acontecimentos no mesmo ano que intensificaram o clima de agitação. Episódios como o assassinato da vereadora Marielle Franco e de seu motorista, Anderson Gomes, no mês de março e a greve dos caminhoneiros, no fim de maio contribuíram para a radicalização da disputa político-ideológica (Dourado, 2020). Neste contexto, a

⁴ Contas falsas de plataformas digitais gerenciadas por indivíduos, que buscam desestabilizar alvos pré-determinados em postagens, fóruns e comunidades virtuais.

⁵ Programas de computador que executam tarefas de forma automatizada e pré-definida, simulando o comportamento de um usuário humano. Podem ser executados de forma independente, sem necessidade de ação humana.

⁶ Contas falsas criadas para elogiar, defender ou apoiar uma pessoa ou organização, influenciando a opinião pública.

polarização política contribuiu para a deflagração do movimento de difusão de notícias falsas nas plataformas digitais.

No que diz respeito às campanhas de desinformação, tem-se visto ao longo dos últimos anos de atividade política online a adesão a contas falsas, automatizadas ou gerenciada por humanos. São contas criadas com o intuito de simular movimentos sociais espontâneos, por meio de ataques direcionados a adversários políticos, de notícias falsas, de rumores, do impulsionamento de hashtags e do questionamento de uma determinada afirmação (Dourado, 2020).

Como aponta Claire Wardle, “cada vez mais, vemos a armamentização do contexto e o uso de conteúdo genuíno, mas conteúdo que é distorcido e reformulado” (Wardle; 2020). Durante os processos eleitorais brasileiros observa-se a desinformação tanto direcionada contra candidatos quanto contra o processo eleitoral.

Conforme aponta Tatiana Dourado (2020), episódios presentes na agenda pública de 2018 durante o período pré-eleitoral, como a morte de Marielle Franco, a greve dos caminhoneiros e o contexto anterior à prisão de Lula, contribuíram significativamente para a definição e delimitação do que constituem informações falsas no cenário nacional. Já em 2022, as eleições presidenciais foram marcadas pela propagação de informações falsas. Estas informações se mostraram mais sofisticadas do que aquelas que circularam durante as eleições de 2018, com o uso de *bots* e *deepfakes*.

Os problemas enfrentados mencionados acima continuam sendo pautas centrais no debate político brasileiro de 2026. O debate mantém-se marcado por uma comunicação progressivamente descentralizada e ambientada nas plataformas de mídias sociais e pelos efeitos gerados sobre o comportamento político-social.

Entretanto, além de iniciativas independentes ou associadas a veículos jornalísticos, a partir do começo da gestão de Luiz Inácio Lula da Silva, em janeiro de 2023, o tema da integridade da informação passou a ser uma das prioridades do governo, sobretudo nas relações internacionais. Foram assinadas declarações, lançadas iniciativas e redigidas propostas advindas da presidência brasileira no G20. Assinou-se a Declaração Global sobre Integridade da Informação, em conjunto com a Bélgica e o Canadá, e a Declaração de Intenções Conjunta sobre Integridade da Informação e Combate à Desinformação e houve o lançamento de uma iniciativa voltada para a integridade da informação relacionada a questões climáticas.

A discussão política que acontece na esfera pública, contribui para a formação da opinião pública. É sustentada por processos comunicacionais e trocas informativas que envolvem cidadãos, meios de comunicação e governos.

2.4 Debate Jurídico Brasileiro sobre Desinformação

Visando o combate à desinformação política, o governo brasileiro busca reduzir o seu impacto ao instituir medidas por meio de leis. Dentre elas, destaca-se o Marco Civil da Internet (Lei 12.965/2014), a Lei 14.192/2021, o PL 2630/2020, as regras de propaganda atualizadas pelo Tribunal Superior Eleitoral (TSE) e a autarquia Autoridade Nacional de Proteção de Dados (ANPD) vinculada ao Ministério da Justiça e Segurança Pública.

O Marco Civil da Internet, considerado a “Constituição da internet brasileira” (Hartmann.; Curzi.; Zingales.; Almeida., 2023, p.13), estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. O Marco Civil foi elaborado a partir de consulta pública, passando por um extenso processo de discussão, edição e aprovação. Pessoas, organizações, empresas e órgãos governamentais tiveram a sua parcela de contribuição para a elaboração do documento. Com a adoção do MCI, a legislação brasileira sobre o ambiente digital se equiparou às diretrizes internacionais sobre preservação dos direitos humanos.

Cabe ressaltar, entretanto, que o Marco Civil, necessita de atualizações urgente que contemplem as questões e demandas do cenário digital brasileiro atual. Há, por exemplo, limitações à responsabilização de intermediários. Provedores de aplicações da internet, ou seja, empresas responsáveis pelas plataformas digitais, não poderiam ser totalmente responsabilizados, de acordo com essa regra, salvo em casos excepcionais. Neste contexto, o PL 2630/2020, que será comentando mais a frente, contribui para o processo de atualização.

A Lei 14.192/2021, por sua vez, estabelece normas para prevenir, reprimir e combater a violência política contra a mulher. Assim, criminaliza a divulgação de fatos ou vídeos com conteúdo inverídico durante o período de campanha eleitoral com o objetivo de prejudicar candidatas.

Esta lei altera o Código Eleitoral (Lei nº 4.737/1965), a Lei dos Partidos Políticos (Lei nº 9.096/1995) e a Lei das Eleições (Lei nº 9.504/1997). Considera-se, para esse fim,

violência política contra a mulher toda ação, conduta ou omissão com a finalidade de impedir, obstaculizar ou restringir direitos políticos da mulher.

Analisando outro aspecto, há iniciativas exercidas por órgãos governamentais que contribuem ativamente para o debate Jurídico Brasileiro sobre Desinformação. De acordo com documento redigido pelo TSE, plataformas como Meta (Facebook, Instagram e WhatsApp), X, Google, YouTube e TikTok se comprometem, desde o período eleitoral de 2022, a utilizar a tecnologia à sua disposição para divulgar e conferir credibilidade a fatos verídicos sobre o processo eleitoral, dando prioridade à divulgação de dados oficiais.

Há ainda a iniciativa Coalizão para Checagem, criada a partir do “Plano Estratégico do TSE para combate à desinformação nas Eleições 2022”. O objetivo da iniciativa é apurar a veracidade das informações que circulam na internet e identificar notícias falsas sobre as eleições para contestá-las de por meio de respostas reais e de qualidade, a fim de:

Garantir o direito de eleitoras e eleitores de participar de eleições livres e justas com igualdade de oportunidades entre os diferentes candidatos, partidos e correntes políticas na disputa, a partir de um processo eleitoral legítimo, com higidez e credibilidade pública. (Portal do TSE, 2022).

O TSE atua ainda em parceria com agências de checagem, apoiado pelos Tribunais Regionais Eleitorais. Assim, o TSE une-se a agências como Agence France-Presse (AFP), Lupa, Aos Fatos, Boatos.org, Comprova e Fato ou Fake, para agir na linha de frente do combate digital à desinformação.

A parceria contribui para desmentir, por exemplo, boatos sobre urnas eletrônicas pré-programadas, veracidade de resultados de eleições, mensagens suspeitas sem embasamento e comportamentos e falas atribuídos erroneamente a candidatos.

Em 2018, o contexto de desordem, disparos automatizados de mensagens e a circulação de notícias falsas sobre o processo eleitoral durante as eleições presidenciais brasileiras, fortaleceu o cenário da checagem de fatos. Então, a Justiça Eleitoral Brasileira estabeleceu parcerias com checadores de fatos no período entre 2018 e 2020, a fim de deter, reduzir o alcance e mitigar o impacto da desinformação relacionada ao processo eleitoral (Lelo, 2022).

Em 2020, o Tribunal Superior Eleitoral lançou o projeto conjunto “Coalizão para Checagem - Eleições 2020”, voltado para a checagem de informações nas Eleições de 2020. O projeto foi realizado em parceria com nove agências de checagem, sendo elas:

AFP, Agência Lupa, Aos Fatos, Boatos.org, Comprova, E-Farsas, Estadão Verifica, Fato ou Fake e UOL Confere.

Propostas como estas mencionadas acima, reforçam o compromisso das instituições judiciárias e governamentais com a estipulação de mecanismos de governança no ambiente digital.

2.5 Desinformação e Democracia nas Plataformas Digitais

O termo “plataformas”, ainda se baseando no conceito elaborado pelo autor Tarleton Gillespie (2010), pode assumir diversos significados, a partir de perspectivas diversas, como sociológicas, de negócios ou de estudos de redes. A sua definição depende do enfoque, podendo ser abordada pelos campos de Estudos de Mídias, Comunicação, Marketing, Ciência da Computação, Direito e Sociologia, por exemplo.

Diversas das plataformas existentes foram criadas e desenhadas por indivíduos e organizações inspirados pela promessa de uma internet pautada na liberdade de expressão, estendida à participação e conexão social entre as pessoas (Gillespie, 2018). Assim como jornais, estúdios de cinema, gravadoras e editoras, buscam posicionar-se de uma forma estratégica no cenário econômico-cultural digital, ganhando centralidade como ambiente de acesso, distribuição e transmissão de informações (Dourado, 2020).

Ao providenciar acesso instantâneo e sem fronteiras ao armazenamento de informações, às facilidades como calendários, notas, serviços de mensagens e ao conteúdo em diversos formatos, as plataformas buscam se tornar relevantes e, eventualmente, indispensáveis. Podemos dividir o poder das plataformas em três formas: o poder político-econômico para influenciar ou prever decisões; o *soft power* que funciona por meio da atração e da cooptação; e o *platform power*, ou seja, o poder da plataforma exercido pelas grandes empresas de tecnologia. (Nielsen e Ganter 2022, p.20).

A terceira forma de poder seria exclusiva das plataformas. Este poder compreenderia: a definição de padrões; o estabelecimento e o rompimento de conexões e a operação simultânea em mais de um domínio. Por meio desta última, demonstraria a dimensão da sua influência em comparação àquela exercida por outros canais midiáticos.

Mediante este quadro apresentado, torna-se inviável separar as plataformas online da vida offline para analisá-las objetivamente. Ambas estão intrinsecamente conectadas, afinal, vivemos uma “cultura de conectividade” baseada na “socialidade plataformizada”,

segundo José van Dijck (Nielsen e Ganter, 2022, p.20). As empresas responsáveis pelas plataformas tornaram-se entrelaçadas à vida real contemporânea, da política à cultura, às relações de trabalho e ao lazer (Gorwa, 2019).

Desde questões como privacidade à violência, a sociedade moderna se depara com uma tecnologia, que, além de definir dinâmicas de relacionamento no ambiente virtual, impacta diretamente vidas reais. A partir dos algoritmos, as plataformas também fazem uma curadoria personalizada de conteúdos e informações “(...) eles não só nos ajudam a encontrar informações, mas nos fornecem meios para saber o que há para ser conhecido e como fazê-lo; a participar dos discursos sociais e políticos e de nos familiarizarmos com os públicos dos quais participamos” (Gillespie, 2018, p. 97).

O discurso sobre uma cultura de conectividade implica no ideal de que as redes sociais são necessárias, indispensáveis e vantajosas, sobrepondo-se a questões como a necessidade de se adotar filtros ou discutir privacidade e inviabilizando o debate de temas urgentes que tangem às plataformas. Entender a extensão do poder de influência exercido pelas plataformas e a sua centralidade na cultura de conectividade é fundamental para dimensionar a sua relação com a democracia no século XXI.

As mídias sociais permitem a construção de redes sociais online (Silva, 2012), propiciando a criação de comunidades, que produzem e consomem conteúdo. A dinâmica de funcionamento das plataformas visa privilegiar a comunicação interpessoal e, desta forma, a troca de interações entre usuários com interesses afins.

Considerando o seu potencial de alcance e rápida transmissão de informações, as plataformas são um espaço favorável à discussão política. Seja para a busca de informações, compartilhamento de ideias, convocação para eventos e manifestações ou para a emissão de opiniões, há um fluxo contínuo de informações.

Crescem bolhas nas diferentes plataformas que, por vezes, não se comunicam entre si, dando lugar, assim, a um debate virtual polarizado, mas fragmentado e a criação de um ambiente favorável à propagação e validação de informações falsas e narrativas enviesadas. Os espaços virtuais transformam-se então em campos abertos à disseminação de informação. Tal transformação tem potencial de favorecer a difusão de informações falsas e a construção de narrativas enviesadas tendenciosas, que podem causar danos

severos à democracia e à credibilidade de jornalistas, antes considerados os únicos *gatekeepers*⁷ da informação; de representantes políticos e de instituições governamentais.

No Brasil, a militância da extrema-direita financia o funcionamento de uma indústria desinformativa, que se apropria de uma circulação e construção de narrativas nas plataformas digitais. O uso de mídias sociais, a polarização política e a proeminência de informações falsas em circulação podem, combinados, pôr em xeque a qualidade da democracia (Tucker et al., 2018). Esses três fatores, junto com engajamento político, meios de comunicação tradicionais e comportamento político, contribuem para um quadro de confusão generalizada sobre os acontecimentos públicos e sobre o funcionamento democrático.

O papel do jornalismo no ambiente digital vem tornando-se crescentemente associado à checagem de fatos, resultando na criação de diversas agências de checagem de fatos ao longo dos últimos anos. As agências de checagem, como Agence France-Presse (AFP), Lupa, Aos Fatos, Boatos.org, Comprova e Fato ou Fake, atuam em parceria com o TSE, apoiada pelos Tribunais Regionais Eleitorais, na linha de frente do combate digital à desinformação, desmentindo, por exemplo, boatos sobre urnas eletrônicas pré-programadas, veracidade de resultados de eleições, mensagens suspeitas sem embasamento e comportamentos e falas atribuídos a candidatos.

Há uma relação direta entre a desinformação e a democracia nas plataformas digitais, sobretudo no que diz respeito à soberania. Mentiras, ferramentas e estratégias utilizadas para a produção e difusão de informações falsas ou distorcidas representam perigo para a democracia. Organizações e plataformas ao redor do mundo circulam teorias conspiracionistas, fatos sem comprovação ou fonte confiável, além de narrativas fabricadas (Bennett; Livingston, 2021, p.3).

A desinformação é um fenômeno que se transforma e adquire novas características com a cultura digital da sociedade de plataformas, ou seja, a partir da plataformização. Entender as suas dinâmicas e materialidades é uma forma de criar mecanismos técnicos, legais, educacionais, informacionais para combatê-la.

Há uma relação direta entre a desinformação e a democracia nas plataformas digitais, sobretudo no que diz respeito à soberania nacional. Para além das mentiras em si, as ferramentas e estratégias utilizadas para a produção e difusão de informações falsas

⁷ Conceito que descreve a função do profissional que define o que será noticiado, de acordo com a linha editorial, valor-notícia e outros critérios.

ou distorcidas representam perigo. Organizações e plataformas ao redor do mundo circulam teorias conspiracionistas, “fatos alternativos”⁸ sem comprovação, além de narrativas e incidentes fabricados (Bennett; Livingston, 2021, p.3).

Se antes era possível facilmente desmentir informações falsas e conduzir processos de checagem, o surgimento de novos artifícios como a Inteligência Artificial possibilita a criação de conteúdos cuja veracidade é dificilmente distinguível para o cidadão comum, apenas mediante a análise de especialistas. A importância da regulação das mídias e da moderação de conteúdo neste contexto será trabalhada ao longo do próximo capítulo.

2.6 Governança nas Plataformas Digitais

O termo governança, inicialmente, era utilizado como sinônimo de articulação governamental, ou seja, da capacidade de um governo de criar e reforçar regras e de entregar serviços (Fukuyama, F., 2013, p. 4). Nos anos 1990, entretanto, a governança assumiu um novo significado, mais próximo daquele conhecido atualmente. O conceito, antes limitado à articulação do governo, passou a englobar noções amplas de organização, estruturação e regulação da vida em sociedade. Governança passou a significar uma complexa rede de interações entre atores e comportamentos (Gorwa., 2019, p.4).

Seguindo essa lógica, a moderação de conteúdo, as diretrizes de plataforma, termos de serviço e interface de websites constituem mecanismos de governança, para reger o ambiente virtual, garantir o cumprimento de normas pré-estabelecidas e promover a cooperação entre usuários. Essa governança pode ser tanto a nível individual (Gillespie, 2015), quanto coletivo, a partir das políticas e restrições regulatórias da plataforma (Gorwa, 2019). O espaço pode ser dividido entre plataformas, que arquitetam o ambiente virtual; usuários, que atuam e determinam padrões de comportamentos e governos, que estipulam legislações para as interações online (Helberger, Pierson e Poell, 2018).

Assim, a governança nas plataformas pode ser definida como o conjunto complexo de sistemas, procedimentos, instituições, mecanismos e legislações responsáveis por definir as responsabilidades das plataformas e orientar os conteúdos

⁸ Expressão usada pela assessora presidencial de Donald Trump, Kellyanne Conway, ao ser questionada sobre o número de pessoas que estava na cerimônia de posse do republicano em janeiro de 2017.

produzidos por usuários e atividades hospedados nelas (Gillespie, 2017; Alves, 2022). Esta governança da internet envolve diversos atores, regras, práticas e situações,

A cultura digital contemporânea pode ser definida a partir da plataformização, da dataficação e da performatividade algorítmica (Lemos, 2020). As plataformas digitais, para além de propiciar a veiculação de conteúdos e informações e de possibilitar trocas entre usuários, são responsáveis por regular, monitorar e coordenar as atividades que ocorrem no seu espaço. Essas estabelecem métodos de governança (Alves, 2022).

Em 2015, durante a conferência internacional RightsCon, o documento denominado Princípios de Manila, também conhecido como Princípios sobre Responsabilidade dos Intermediários, foi criado por organizações da sociedade civil. O intuito da sua criação foi estabelecer parâmetros e diretrizes para as comunicações online e regulamentar a reponsabilidade legal dos intermediários no que diz respeito a conteúdos postados por terceiros.

Alguns dos principais pontos levantados no documento são: a necessidade de transparência sobre políticas existentes e decisões relacionadas à restrição de conteúdo; a proteção dos direitos humanos na Internet; a importância de se evitar a remoção excessiva de conteúdos, para não interferir na liberdade de expressão. De acordo com o primeiro princípio dos Princípios de Manila, “os intermediários que devem ser protegidos por lei da responsabilização por conteúdos produzidos por terceiros” (Manila Principles, 2015, p.2). Tal princípio reforça a ideia de se contribuir para um ambiente virtual seguro sem abusos por parte de autoridades privadas e governo.

Entretanto, uma vez que as plataformas digitais são consideradas intermediárias, sem mecanismos externos de *accountability*⁹ ou processos eleitorais democráticos, recebem críticas sobre a sua legitimidade para exercer e estabelecer mecanismos de governança no ambiente digital. Para além das plataformas, há outras instâncias diretamente e indiretamente responsáveis pelo estabelecimento da governança. Outros exemplos de mecanismos de governança são a moderação de conteúdo e as agências de checagem.

Investe-se ainda em canais oficiais e dispositivos de denúncia nos websites, como o aplicativo Pardal, utilizado para denúncias de propaganda eleitoral irregular e corrupção (propaganda, disparo em massa e envio de mensagens); páginas e linhas telefônicas disponibilizadas para denúncias de crimes eleitorais e o Sistema de Alertas de

⁹ Termo que pode ser traduzido como prestação de contas ou responsabilização.

Desinformação Eleitoral (SIADE), voltado para denúncias de propaganda eleitoral irregular.

As sanções e punições aplicadas dizem respeito não somente aos criadores de conteúdos falsos, mas também a quem o compartilha, quem o replica. Dentre as punições aplicáveis, pode-se mencionar: aplicação de multas monetárias; suspensão do usuário, ordem de retirada do conteúdo em questão.

As plataformas são intermediárias entre usuários; entre usuários e público e entre cidadãos e a órgãos de aplicação da lei, legisladores e governo. A adoção desta posição “intermediária” afeta diretamente a governança das plataformas, assim como as discussões em torno da conveniência e das responsabilidades associadas à mesma (Gillespie, 2017).

Embora as plataformas se comprometam publicamente a combater o fenômeno desinformativo, a indústria da desinformação está estrategicamente relacionada à lógica de operação destas plataformas e ao modelo de negócios das *big techs*¹⁰. Além de defenderem uma liberdade de expressão absoluta no ambiente digital, promovem a busca desenfreada por engajamento e *likes*, o ranqueamento de algoritmo, a criação de redes de incentivos e a monetização relacionada à veiculação de conteúdos extremos.

Não há real interesse das plataformas em alterar suas políticas, sob a alegação de que mecanismos de regulação mais eficientes cerceariam a liberdade de expressão dos criadores. Há uma visão de que comprometeria o modelo de operação e rentabilidade da plataforma.

2.7 Projeto de Lei (PL) 2630/20

O Projeto de Lei (PL) 2630/20, apelidado popularmente de “PL das Fake News” foi idealizado para “instituir a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet”. O documento é um marco relevante para o fomento e a atualização do debate acerca das plataformas digitais e da regulação no Brasil. Atualmente, tramita sua versão reformulada de 2022, após consultas públicas do projeto

¹⁰ São as grandes empresas de tecnologia que dominam o mercado e que atuam em diversos ramos da inovação e da tecnologia.

aprovado pelo Senado, que traz importantes contribuições relacionadas à transparência (D'Almonte; Santos, 2023).

O intuito da criação do PL 2630/20, conforme consta nos dez capítulos da proposição aprovada pelo Senado Federal e apresentada à Câmara de Deputados, é promover a regulação e a responsabilização em condutas online, estabelecendo medidas de combate à desinformação por meio do estabelecimento de regras para o uso da internet no Brasil e obrigações de transparência às plataformas. Conforme definido no artigo 6º, um dos principais objetivos do documento é “proteger a liberdade de expressão, o acesso à informação e fomentar o livre fluxo de ideias na internet e resguardar os usuários de práticas fraudulentas” (Brasil, 2022).

Como justificativas para sua implementação, é citada em documentos oficiais do governo brasileiro, a abordagem dos seguintes tópicos: liberdade de expressão, garantia de direitos fundamentais, acesso às comunicações, fortalecimento do processo democrático, moderação de conteúdo na internet, publicidade na rede, responsabilidade das plataformas digitais, atuação do Poder Público nas redes, criação de um Conselho de Transparência e Responsabilidade na internet e mecanismos de autorregulação para as plataformas.

Ademais, o Projeto de Lei propõe-se a alterar o Marco Civil da Internet (Lei nº 12.965, de 23 de abril de 2014) e a Lei nº 10.703, de 18 de julho de 2003, dispõe sobre cadastramento de usuários de telefones celulares pré-pagos. Segundo a atualização mais recente do processo, referente ao mês de abril de 2025, o PL está pronto para ser levado como pauta no plenário.

A abordagem de conteúdos de propaganda eleitoral também é mencionada. Tais conteúdos devem ser identificados, assim como o gasto feito pelo candidato ou partido com o impulsionamento, tempo de veiculação e técnicas de perfilamento. O quarto capítulo, especificamente, aborda o papel do Poder Público.

Dentre os principais pontos abordados no texto do projeto, pode-se destacar: a transparência em relação aos conteúdos publicitários e impulsionados; a transparência nas regras das plataformas; os procedimentos aplicáveis na intervenção ativa de contas e conteúdos de terceiros.

A tramitação e aprovação do projeto é polêmica, considerando a polarização existente na sociedade em relação aos tópicos de “censura no ambiente virtual” e de defesa de uma suposta liberdade de expressão. As próprias nomenclaturas do PL, chamado de PL das Fake News pela imprensa tradicional e pela esquerda, que o

defendem, e PL da Censura pelas plataformas digitais e a direita, evidencia essa polarização (D’Almonte; Santos, 2023).

Às vésperas da votação da PL nº 2630 (Brasil, 2022), *big techs* como Google e Meta se uniram a veículos da direita, como o Brasil Paralelo, para a veiculação de anúncios contra o projeto de lei, infringindo as suas próprias regras de condutas e restrições à publicidade (D’Almonte; Santos, 2023). Ainda em 2023, o Telegram enviou mensagem a todos os seus usuários criticando o Projeto de Lei 2630/2020 e alertando sobre o seu “perigo” e censura para a democracia. A mensagem continha frases como “A democracia está sob ataque no Brasil” e “Veja como esse projeto de lei matará a internet moderna se for aprovado com a redação atual” (Telegram, 2023).

É evidente, assim, que a regulamentação das plataformas digitais vai de encontro aos interesses econômicos das plataformas, que desejam manter seu status com intermediárias, sem responsabilização ou imposição de restrições e regras envolvendo a questão da transparência. A resistência e a postura de vilanização da regulamentação por parte das plataformas contribui progressivamente para o aumento da disseminação da desinformação e, conseqüentemente, para o enfraquecimento da democracia.

Entretanto, esta não é a única questão controversa a ser debatida. Outro ponto controverso está relacionado à própria estrutura proposta pelo projeto. O texto do PL estabelece um modelo regulatório de autorregulação regulada para o Conselho de Transparência proposto. Este arranjo não favorece o estabelecimento do Conselho como órgão público supervisor independente, algo que pode vir a facilitar a possibilidade de privatização da regulação (Hartmann.; Curzi.; Zingales.; Almeida., 2023, p.185)

Como apontado ainda em posicionamento emitido pelo Centro de Tecnologia e Sociedade da Fundação Getúlio Vargas (2021, p.19), o Conselho de Transparência proposto no PL também não impõe uma obrigação de desenvolvimento contínuo do aprimoramento técnico de obrigações legais nem de prestação de contas a nível público.

O documento representa um marco relevante para o fomento e a atualização do debate acerca da regulação das plataformas digitais no Brasil, sendo importante acompanhar a sua evolução.

3 Moderação de Conteúdo

3.1 O Que é Moderação de Conteúdo

O extenso debate sobre a moderação de conteúdo nas plataformas digitais vem adquirindo uma centralidade cada vez maior ao longo dos últimos anos. Ao longo da última década, pudemos observar o surgimento de empresas, agências, grupos de pesquisa, cursos, especializações profissionais, fóruns e *think tanks* especializados neste serviço, além do lançamento de publicações acadêmicas sobre o assunto (SILVA, 2012, p.41).

O tema, em razão da sua grande relevância, envolve diferentes atores e a mobilização da sociedade civil, em escala mundial. A moderação tem impacto direto sobre a vida e os direitos dos cidadãos, sendo essa uma das razões pela qual é essencial debatê-la e analisá-la. Caracteriza-se como uma das políticas relacionadas a redes sociais mais contestadas, tendo sido analisada de diferentes perspectivas (Gillespie et al., 2020) ao longo dos últimos anos.

Estuda-se a moderação para além do seu papel relacionado à prevenção de danos e de reforço de leis e direitos. É essencial compreender o que são as plataformas efetivamente e analisar como impactam sutilmente a vida pública (Gillespie, 2018). Todas as plataformas estão sujeitas à moderação, sendo obrigadas a servir a determinados conjuntos de regras e atuar como mediadores de disputas, intérpretes de leis e árbitros, mediante critérios pré-estabelecidos.

O propósito deste capítulo teórico é abordar um breve histórico e a discussão teórica sobre o tema. Ao longo dos próximos subtítulos serão apresentados: o conceito de moderação de conteúdo em si; perspectivas regulatórias; o repertório da moderação e as suas principais características. Este capítulo reforça ainda a contribuição fundamental da moderação de conteúdo para o desmantelamento da infraestrutura de desinformação, ao analisar o seu repertório e as funções que assume no cenário político brasileiro.

Para além de responder tais questões, esta dissertação apresenta conceitos que auxiliam na compreensão das características do sistema de regulação das plataformas. A partir da compreensão das políticas de moderação de conteúdo, torna-se possível reformar o sistema regulatório digital de forma eficiente

A partir dos conceitos apresentados no capítulo anterior, pode-se argumentar que a tecnologia simultaneamente molda e é moldada por discussões na internet (Wright; Street, 2007). Assim, envolve a atuação de órgãos públicos, empresas e instituições, sejam elas governamentais ou não-governamentais. Afinal, a regulação sobre plataformas é moldada ativamente pela(s) interação(ões) entre diversos atores, públicos ou privados.

No cenário de um espaço virtual marcado cada vez mais pela pluralidade de vozes, tanto na esfera pública quanto na privada, circula um fluxo crescente e ininterrupto de informações. A fim de organizar esse fluxo constante, surge a necessidade de se criar regras e mecanismos que facilitem a regulação do ambiente virtual e atendam aos requisitos impostos pela legislação nacional, no caso, a legislação brasileira, para o meio virtual.

A moderação de conteúdo é a governança do conteúdo criado por usuários, contemplando revisão e monitoramento, garantindo que este corresponda a diretrizes e normas para conteúdos no ciberespaço. Na era digital, a moderação de conteúdo online surge como mecanismo de governança das plataformas (Grimmelmann, 2015). Ainda que em graus e formas diferentes, é realizada por todas as plataformas digitais (Gillespie, 2018). Afinal, ao mesmo tempo que as plataformas favorecem uma rápida difusão de informações, criam espaços para comportamentos incorretos e desinformação (*misinformation*)¹¹ (Singal, et al., 2022).

Comportamentos incorretos, segundo Singal, são aqueles que contemplam uso de linguagem explícita, conteúdos depreciativos ou desrespeitosos; já o termo desinformação, é utilizado como sinônimo de informações falsas ou imprecisas. Apesar desse propósito, o principal motivo pelo qual as plataformas moderam conteúdo é manter a sua imagem para seus usuários, anunciantes e o público em geral (Gillespie, 2018).

É importante para a manutenção do modelo econômico sob o qual operam proporcionar um ambiente que corresponda às expectativas dos seus usuários e anunciantes e aos seus valores propagados. Envolver-se em debates políticos polarizados poderia trazer dificuldades práticas de alinhamento e alienar parte da base de usuários (Douek, 2021; p.775).

A moderação à qual o texto se refere é aquela realizada nas plataformas digitais, relacionadas à postagem, notificação, publicização e veiculação de conteúdos online. Podemos dizer que esse monitoramento das mídias sociais é simultaneamente “tanto um

¹¹ Tradução livre para os termos *uncivil behavior* e *misinformation*.

mercado da comunicação digital quanto um conjunto de ferramentas e técnicas empreendidas por diferentes profissionais e setores de empresas” (Silva, 2012).

As plataformas atuam como intermediárias entre usuários e conteúdos, intermediando o discurso online. Assim, assumem um papel central na estrutura informacional da internet, como fontes de informação e discussão e amplificadores de discursos. Elas vêm crescendo e se desenvolvendo significativamente ao longo das últimas décadas, e adquirindo poder em uma escala inédita.

Considerando sua relevância e centralidade, foi preciso estabelecer normas, garantias de conformidade com leis locais e internacionais, além de mecanismos que possibilitassem a prestação de contas e a filtragem de conteúdos considerados nocivos, prejudiciais aos usuários, falsos ou ilegais. Tais regras e mecanismos são essenciais como medidas de regular, limitar e fiscalizar o poder crescente das plataformas e a forma como são usados.

Cada plataforma possui suas respectivas particularidades e diretrizes (*guidelines*), criadas para estabelecer parâmetros e regras de conduta e comportamento dos usuários no ambiente virtual. Por meio desse mecanismo, as plataformas tornam-se capazes de regular e orientar o conteúdo que circula no seu espaço.

O teor da moderação pode diferir de uma plataforma para outra, a depender dos critérios utilizados, prioridades e necessidades de cada rede. Isto torna difícil o estabelecimento de uma padronização, apesar de a moderação, em todo caso, obrigatoriamente respeitar a legislação vigente do país no qual se estabelece. Além disso, a pluralidade linguística e modal das plataformas representa desafios (Recuero, 2024). Segundo Recuero (2024), a subjetividade da compreensão das regras das plataformas é significativa, o que dificulta a eficácia do modo.

Outra questão que dificulta exercer apropriadamente a atividade da moderação de conteúdo é a relutância dos representantes das plataformas em se envolverem efetivamente e assumirem responsabilidade por uma moderação de conteúdo mais rígida (Curzi; Hartman; Almeida; Zingales, 2022). Esses representantes alegam que regimes regulatórios muito rígidos poderiam interferir direta ou indiretamente no fluxo de informações.

Esta questão é uma das pautas centrais do debate sobre liberdade de expressão na internet, que mobiliza a sociedade. A discussão gira em torno da seguinte dualidade: a moderação de conteúdo cerceia a liberdade de expressão absoluta dos usuários, ao determinar o que é correto ou incorreto e o que está de acordo?

Buscando responder a esta pergunta, estudiosos do campo de estudos de mídias e tecnologia, lideranças da sociedade civil e parlamentares que advogam a favor da importância da moderação, trazem argumentos sólidos para a necessidade de se regular as tecnologias. Alguns destes argumentos são: a atribuição de papéis aos devidos atores; a responsabilização das plataformas como intermediárias; a consolidação de direitos e obrigações em consonância com a legislação sobre o ambiente virtual e com a Constituição; a prevenção e a solução de problemas atuais e emergentes de forma justa e eficiente.

Recentemente, o CEO da Meta, Mark Zuckerberg, retomou esse debate em público, quando justificou sua decisão de encerrar o programa de checagem de fatos sob o pretexto de combater censura e checadores de fatos politicamente tendenciosos. A situação oposta também pode ocorrer em casos nos quais as plataformas recebem incentivos significativos para intensificar a prática da remoção de conteúdo. Nesses casos, podem acontecer bloqueios excessivos (prática chamada de *overblocking*) dos conteúdos veiculados dentro dessas plataformas. (Curzi; Hartman; Almeida; Zingales, 2022).

A moderação de conteúdo vem revelando-se um recurso cada vez mais indispensável ao meio digital. Atuando de maneira complementar à legislação que trata da regulação do espaço virtual, ela tem um papel fundamental no desmantelamento da estrutura de desinformação, no reforço de leis já existentes e na checagem de fatos (*fact checking*), prática que atesta a veracidade de informações. Não há uma única forma de se realizar a moderação de conteúdo, uma vez que a atividade assume diferentes categorias e formas, que serão abordadas ao longo dos próximos subtítulos do capítulo.

Um ponto frequente de críticas sobre a atividade da moderação é a falta de transparência das plataformas. Frequentemente, a atividade é tratada de forma secreta pelas empresas privadas responsáveis pelas plataformas, sob pretexto de proteger a identidade dos seus trabalhadores moderadores, a sua propriedade tecnológica ou para atribuir menor responsabilidade para os moderadores.

Tal opacidade quanto aos critérios adotados para qualquer decisão (restringir, editar, rotular, suspender ou remover um conteúdo) dificulta o combate à desinformação e gera confusão para os usuários, que não compreendem como as regras são criadas nem a sua validade. Neste contexto, esta falta de transparência pode, conseqüentemente, acabar afetando a integridade democrática.

3.2 Histórico da Moderação

Cada plataforma desenvolveu suas próprias políticas de moderação de conteúdo, a partir dos seus próprios critérios e prioridades. Embora hoje tais políticas já estejam formuladas e suas principais regras consolidadas por meio de documentos, guias e diretrizes, não foi um processo “natural”. Afinal, a responsabilidade das plataformas de implementar e fiscalizar a moderação de conteúdo, era algo novo.

Até então, no período que compreende o início dos anos 90 e início dos anos 2000, “muitas novas empresas da internet pensavam sobre si mesmas como empresas de softwares” (Klonick, 2018). A maior parte destas empresas não se considerava efetivamente responsável pelos efeitos do discurso veiculado online nem pela sua intermediação. Seu papel como facilitadoras exigia menor compromisso, envolvimento e, conseqüentemente, significava menor responsabilização, necessidade de transparência ou prestação de contas.

Até 2010, a discussão sobre moderação de conteúdo restringia-se à restrição da liberdade de expressão em regimes autoritários e no seu papel como parte do processo de apuração jornalística. A preferência por uma abordagem mais leve era então favorável ao modelo operacional e comercial das plataformas, afinal, essas são redes construídas a partir da premissa de promoção da liberdade de expressão dos usuários, baseadas nos princípios da Ideologia Californiana.

A prática da moderação online da forma como a conhecemos hoje ganhou força no início dos anos 2000, período no qual surgiram Facebook (2004), YouTube (2005) e X - originalmente Twitter (2006).

Como aponta o artigo *The Emerging Science of Content Labeling: Contextualizing Social Media Content Moderation* (2020), de início, a abordagem adotada consistia em uma política de baixa interferência, priorizando a manutenção da autonomia do usuário para postar o conteúdo que desejasse.

Nessa época, tais plataformas se autopromoviam como espaços abertos, livres e participativos, se comparadas à mídia de massa. (Keulenaar, E.; Burton, A.; 2021) e eram amplamente percebidas como altamente inspiradas por instituições liberais democráticas (Klonick, 2018).

Adotava-se então uma política não-intervencionista para a moderação, pautada na ideia de que as plataformas, como intermediárias, deveriam promover a conexão entre pessoas e a difusão de informações, sem intervir sobre questões políticas ou morais e

condutas. Eram impostas limitações aos usuários, aplicadas para viabilizar a pluralidade de discursos e consideravam as decisões e opiniões desses. (Keulenaar.; Burton; Kesjes, 2021, P.2.)

Embora, até então, não se tratasse de uma questão pública coletiva, já havia críticas na década de 2010 relacionadas à alta quantidade de dados pessoais aos quais grandes plataformas, como o Facebook e a Google, tinham acesso. Alguns anos depois, as plataformas se viram diante de acusações de incentivo à propagação do discurso de ódio e da desinformação, além da ausência de transparência dos seus mecanismos. Assim, foram pressionadas a reformular suas políticas e aprimorar a fiscalização, o que resultou na adoção de um modelo de moderação mais ativo e rígido, semelhante ao modelo vigente hoje.

Com o passar dos anos, houve uma mudança no discurso das plataformas acerca da moderação, marcada pela ruptura com a noção da moderação leviana até então (Gillespie, 2013). Os mecanismos se tornaram mais ativos e complexo, com termos de uso mais robustos e moderação mais rígida.

Foi preciso atualizar as leis que, inicialmente, não foram idealizadas com as especificidades das plataformas digitais em mente. Boa parte das políticas de moderação aplicadas às plataformas foram criadas voltadas para uma categoria mais ampla de serviços online. (Gillespie, 2018).

Após situações como a distorção de informações para fins políticos em eleições; pressão crescente para remover conteúdos desinformativos; o aumento considerável de práticas enganosas e escândalos sobre vazamentos de dados, as plataformas passaram a reivindicar cada vez mais para si mesmas o papel de moderadoras. O número de usuários nas plataformas havia crescido consideravelmente, o que implicou em um impacto maior do espaço midiático sobre a mobilização do discurso midiático.

Na década de 2020 o uso de sistemas automatizados de moderação para sinalização, filtragem, suspensão e *downranking*¹² de conteúdo tornou-se mais comum, caracterizando uma abordagem mais interventora em comparação à anterior. Eventos como as eleições presidenciais dos Estados Unidos em 2016, o agravamento de técnicas desinformativas, novas leis transnacionais para o ambiente virtual, as eleições brasileiras de 2018 e 2022 e o 8 de janeiro, contribuíram para a implementação dessa modalidade de moderação mais agressiva.

¹² Prática que consiste na redução da visibilidade de um conteúdo.

Cada plataforma estabelece suas diretrizes conforme suas necessidades e demandas, mediante pressão social dos próprios usuários, de anunciantes e de organizações. Não houve um movimento unificado sobre a determinação de quais regras deveriam constar na moderação de conteúdo.

Por esse motivo, há diferenças significativas entre abordagens de um mesmo tema em plataformas diferentes. Em razão destas diferenças, surgem também opiniões distintas sobre como atuar em relação a um determinado tipo de conteúdo e pedidos de mudança nas diretrizes vigentes, por exemplo.

3.3 Repertório da Moderação

A fim de se realizar uma análise aprofundada acerca da moderação de conteúdo, é preciso abordar o seu repertório. A moderação desdobra-se em diferentes formatos e nos diferentes modelos de negócios, cada qual com as suas respectivas características e peculiaridades.

Primeiramente, de acordo com Gillespie (2019), a moderação pode ter uma abordagem *soft* ou *hard*, a depender da forma como é implementada e dos objetivos pretendidos. A moderação *soft*, de caráter mais temporário, contempla invisibilidade algorítmica, rótulos, notas e advertências relacionadas a um determinado conteúdo considerado problemático. Já a moderação *hard*, de caráter mais permanente, implica em medidas mais facilmente perceptíveis e drásticas, como a remoção de um determinado conteúdo e a suspensão de contas.

A moderação pode ser realizada de forma manual, por meio da supervisão e do processo de filtragem de dados realizado por moderadores humanos ou de forma automatizada, conduzida por filtragens de dados mecanizadas, realizada por meio de ferramentas ou softwares.

No caso da moderação automatizada, sistemas de inteligência artificial são treinados para analisar e gerenciar de forma otimizada conteúdos que possam violar os padrões estabelecidos pela comunidade. Entretanto, moderadores humanos trazem à tona fatores como a compreensão baseada em contexto, repertório sociocultural e a habilidade de tomada de decisões a partir de uma percepção humana, de forma que uma máquina não conseguiria.

As plataformas possuem equipes próprias de moderadores (trabalhadores responsáveis pela moderação de conteúdo) mas também podem contratar equipes de análise e avaliação de conteúdo de empresas terceirizadas que seguem as regras determinadas pela plataforma.

Esta “terceirização da moderação de conteúdo” (Ahmad, S., 2023, p.115) por parte das empresas responsáveis pelas plataformas foi um fator que contribuiu para a criação de “cadeias de valor globais de moderação de conteúdo” (Ahmad., 2023, p.115). O trabalho mal remunerado, repetitivo e mecânico implicado na moderação frequentemente é realizado por esses moderadores contratados.

Nestes casos, frequentemente, há uma divisão de tarefas atribuídas às equipes das próprias empresas responsáveis pelas plataformas e às equipes terceirizadas: condução de treinamentos; desenvolvimento das políticas de moderação atribuídas às empresas; análise dos conteúdos e resolução de conflitos atribuídos aos funcionários terceirizados.

Existem ainda programas de verificação de fatos conduzidos por colaboradores certificados (*third party fact-checking program*) das plataformas, que contribuem para a governança das mesmas. O programa da Meta, por exemplo, até janeiro de 2025 firmava parceria com organizações de checagem de fatos independentes de cerca de 80 países em mais de 60 idiomas. O foco da iniciativa eram casos virais, principalmente aquelas comprovadamente falsas com potencial de enganar as pessoas.

Ademais, de acordo com Grimmelmann (2015), a moderação de conteúdo online compreende não apenas a remoção ou suspensão de conteúdo, mas também a forma como ele chega aos usuários. Segundo o autor, há quatro modelos de operação: exclusão, precificação, organização e criação de regras.

Para o autor, a moderação pode ser categorizada em duas fases: *ex-ante* (pré-moderação) e *ex-post* (pós-moderação). A primeira, geralmente automatizada, antecede as publicações e determina quais conteúdos serão permitidos, estimulados e vistos nas plataformas; A última, por sua vez, também chamada de moderação reativa, pode ser realizada por moderadores humanos ou sistemas automatizados, e é acionada como resposta a acontecimentos para editar, limitar alcance, corrigir ou remover uma postagem.

Dentre as práticas de moderação de conteúdo existentes, há aquelas mais utilizadas, observadas de forma mais frequente nas plataformas, e aquelas com menor visibilidade. Compreendendo as diferentes técnicas, é possível interpretar a gravidade de cada tipo de conteúdo.

Há a checagem de fatos (*fact-checking*), prática que vem ganhando força nos últimos anos, executada principalmente de forma institucionalizada por veículos, agências de checagem especializadas e jornalistas. Essa propõe-se a combater a atividade desinformativa por meio de fontes confiáveis; da análise de dados; da emissão de informações corretivas que possam desmascarar notícias e alegações falsas ou enganadoras (Bateman; Jackson; 2024; p.38) e da verificação da veracidade de alegações.

Há o uso de rótulos (*labels*), aplicados pelas plataformas a conteúdos ou contas de usuários (definindo, por exemplo, se uma conta é verificada ou se é patrocinada pelo governo). Auxiliam os usuários a entenderem se uma informação é confiável ou não. O conceito de *labeling* utilizado neste capítulo refere-se àquele explicado no relatório e guia de políticas, *Countering Disinformation Effectively: An Evidence-Based Policy Guide* (2024). A criação dos rótulos (*labeling*) pode ser definida como a inserção de contexto relevante para informar ou influenciar como um conteúdo é visto, sem que se precise necessariamente recorrer à checagem de fatos. Assim, o propósito do rótulo é contextualizar uma notícia para quem a visualiza ou lê, “informar sobre a informação”.

Os rótulos, que podem ser positivos ou negativos, deixam a avaliação final a critério do usuário. Podem assumir diferentes formas: categorizar um vídeo ou canal de comunicação como “verificado”; confirmar a identidade de um indivíduo; advertir; rotular um conteúdo como desinformativo; informar sobre um conteúdo; restringir a atividade de uma conta que viole diretrizes da plataforma; encorajar o usuário a seguir boas práticas de informação e buscar fontes variadas, além de apontar o uso de Inteligência Artificial.

Conforme apontam Bateman e Jackson (2024), diversos recursos de design também podem ser considerados rótulos digitais. Dentre eles, elementos gráficos (quadros explicativos, pop-ups que oferecem informações adicionais ao serem clicados, mudança de fonte) e a exibição de dados sobre engajamento em uma notícia, por exemplo.

Estes rótulos podem contribuir para alterar a percepção de um usuário sobre um determinado conteúdo, ainda que não seja algo certo. A fim de medir o impacto de fatores como o tamanho do texto e o uso de elementos gráficos na eficiência da mensagem transmitida sobre a construção da percepção do leitor, é preciso estudar de forma mais detalhada e aprofundada o tema.

Há a aplicação de selos e distintivos (*badges*), que podem ser utilizados para destacar uma determinada conta, indicando o seu status ou afiliação. No caso dos usuários do microblog X, por exemplo, as cores do símbolo de checagem (*checkmark*) apontam o

status da conta: a cor azul indica assinatura premium ativa; a dourada, a conta oficial de uma organização; a cinza, a conta de um governo ou oficial. Esses selos e distintivos são aplicados pela própria plataforma, de acordo com a categorização proposta.

Cabe apontar que um rótulo ou distintivo não significa necessariamente confiabilidade. No caso do X, a verificação da conta pode ser comprada mediante assinatura premium. Isso implica em casos como o ocorrido em 2024, no qual uma conta falsa passando-se pelo ex-presidente estadunidense norte-americano George W. Bush “tuitou” sobre bombardear o Estado de Israel. Inicialmente, a ameaça gerou repercussão na mídia, por se tratar de uma conta verificada. O caso é apenas um dentre outros que simbolizam a confusão gerada pela compra do selo de verificação (*checkmark* azul).

Existem ainda as estratégias de *counter-messaging*. A prática consiste no combate à desinformação por meio da criação de mensagens e narrativas alternativas. Partindo da premissa de que a lógica por si só pode não ser suficiente para mudar as opiniões de indivíduos e impactá-los, a estratégia utilizada se estende para além de fatos e lógica (Bateman; Jackson, 2024).

A eficácia do *counter-messaging* depende de certos fatores variáveis. A adoção da abordagem mais adequada para o público-alvo desejado pode ser decisiva: um público moderado, por exemplo, dificilmente acreditará em narrativas extremistas. Cabe apontar ainda que intensificar um debate e criar narrativas alternativas pode gerar comoção, mas não necessariamente mitigar os efeitos da desinformação.

Outra técnica aplicada é o *shadow banning*, que restringe, de forma temporária ou permanente, a visibilidade do conteúdo de um usuário na plataforma sem notificar o usuário. O usuário continua apto a postar e a responder, mas sem que o seu conteúdo apareça para os seus seguidores. A técnica é aplicada com o intuito de proteger usuários, em casos nos quais o conteúdo de um usuário é considerado inapropriado, produzido por bots ou viola as diretrizes de uma plataforma.

Como medidas reativas mais extremas, há a remoção do conteúdo e a suspensão temporária ou definitiva de uma conta. Tais medidas são aplicadas como consequência em casos excepcionais de violação das políticas de moderação pré-definidas pela plataforma em questão.

4 Metodologia

Os dois capítulos teóricos da presente dissertação apresentaram e os conceitos de “desinformação” e de “moderação de conteúdo”. O referencial teórico, além de trazer os fundamentos teóricos, contribuiu para a contextualização da pergunta de pesquisa.

O capítulo de Metodologia visa detalhar a metodologia utilizada para embasar a pesquisa deste estudo. Assim, ao longo desta seção, serão explicados os procedimentos, etapas básicas, considerações e avaliações aplicáveis para a obtenção de resultados.

O estudo apresenta uma revisão sistemática de literatura para investigar o estado da pesquisa nacional sobre a desinformação política e a moderação de conteúdo nas plataformas digitais no Brasil.

Existem ferramentas desenvolvidas para auxiliar a aplicação dos requisitos necessários às revisões sistemáticas, como cursos e guias práticos, de modo a garantir a qualidade dos resultados. Dada a relevância da construção, esse estudo segue os parâmetros de estruturação e o desenvolvimento de uma RSL estabelecidos pelo curso “Revisão Sistemática e Meta-análise”¹³ e pelo artigo “Guia Para Realizar uma Revisão Sistemática de Literatura” de Chitu Okoli (2019).

Além de simplesmente relatar os procedimentos, a revisão de literatura visa destacar possíveis novas descobertas e contribuições e confere importância à qualidade da pesquisa e dos resultados obtidos.

Mais especificamente, trata-se de uma “revisão de literatura de tese” (Okoli, 2019). Nessa modalidade, a revisão de literatura é realizada como um capítulo de uma tese de pós-graduação, neste caso uma dissertação.

Originalmente aplicada na análise dos resultados quantitativos de estudos da área de saúde, a revisão sistemática de literatura também pode ser aplicada às ciências humanas. No presente contexto foi aplicada de forma qualitativa. O intuito de utilizá-la é auxiliar na compreensão de diferenças, lacunas e novos caminhos possíveis a serem seguidos por novas pesquisas (Okoli, 2019).

As revisões sistemáticas utilizam protocolos específicos que permitem sua reprodução por outros pesquisadores e a obtenção de evidências científicas válidas. Ao se buscar artigos e/ou livros que auxiliem no desenvolvimento da revisão, há um consenso

¹³ Curso disponibilizado na plataforma de aprendizado Coursera “Revisão Sistemática e Meta-análise”, disponível em: <https://www.coursera.org/learn/revisao-sistematica>

sobre a necessidade de se definir um protocolo detalhado antes da realização da coleta de dados (Donato, Donato, 2019).

O processo de revisão sistemática de literatura (RSL) foi conduzido de forma imparcial por uma única pessoa, por meio de etapas padronizadas. Essas etapas serão devidamente introduzidas e explicadas ao longo da seção.

A realização das revisões sistemáticas permite uma avaliação cuidadosa e abrangente da literatura já existente. Como afirmam Donato e Donato (2019), por tratar-se de um processo sistemático, há foco na estrutura, na organização e na documentação.

Aplicou-se etapas a fim de garantir maior rigor teórico e evitar a possibilidade de haver vieses de seleção e aferição. Foram utilizadas, neste estudo secundário, palavras-chave e descritores em um processo de síntese para responder à pergunta de pesquisa.

Considera-se como o ponto de partida da realização de revisão sistemática de literatura a formulação de uma pergunta norteadora. A pergunta orienta a pesquisa, o processo de busca e as etapas do processo. A pergunta de pesquisa que guia o projeto, conforme estabelecido na introdução, é a seguinte: “Como a moderação de conteúdo nas plataformas digitais influencia a circulação e o impacto da desinformação política no Brasil, considerando períodos eleitorais?”.

Um dos primeiros passos desta RSL foi a definição das bases de indexação dos trabalhos científicos utilizadas para a busca, recuperação e seleção dos trabalhos. Assim, a pesquisa foi realizada a partir de dois bancos de dados de acesso aberto, sob delimitação temporal específica (2020-2025). São elas o Google Acadêmico e a *Social Science Research Network* (SSRN). A escolha das bases foi feita a partir da análise dos resultados iniciais e considerando a relevância das bases em questão para as áreas da Comunicação e dos Estudos de Plataforma, sobretudo considerando o elemento da língua portuguesa.

As bases de dados consultadas, assim como os critérios de busca e de seleção, inclusão e exclusão de documentos são explicitados ao longo deste estudo. Do mesmo modo, o processo de análise realizado é descrito, assim como as suas limitações.

A partir da pergunta de pesquisa e dos elementos-chave para o estudo, foram definidos os descritores “desinformação política” e “moderação de conteúdo”. Buscou-se transformar os elementos da pergunta de pesquisa em termos de busca.

Em relação ao termo de busca utilizado, optou-se pela utilização de operadores booleanos para a obtenção de resultados. O conectivo escolhido foi “AND”, para que apenas documentos contendo todos os termos pesquisados fossem encontrados. Assim,

definiu-se como expressão final de busca: "desinformação política" AND "moderação de conteúdo".

O Quadro 1, abaixo, apresenta a quantidade de resultados encontrados a partir das coletas em cada base de dados.

Quadro 1: Resultados de coletas por base

Base	Resultados Encontrados	Total
Google Acadêmico	37	41 produções encontradas
SSRN	4	

Fonte: elaboração própria

4.1 Processo de Seleção

O processo de seleção ocorreu em três etapas sucessivas: leitura dos títulos, leitura dos resumos e, por fim, a leitura das palavras-chaves dos textos elegíveis, com aplicação final dos critérios estabelecidos. Os termos de busca foram analisados nos resumos e palavras-chaves dos documentos, de forma a garantir que não houvesse vieses no estudo. Tais termos foram buscados em língua portuguesa.

Para esta escolha linguística, foi considerada a intenção de se obter apenas resultados diretamente relacionados ao contexto político-social brasileiro. Não houve restrições quanto à localidade dos autores, apenas ao idioma de busca. Outro critério adotado para a análise foi considerar apenas produções disponibilizadas com acesso livre online como elegíveis para a análise.

Os mecanismos de busca foram configurados para incluir resultados a partir dos resumos de artigos e livros das áreas de ciências políticas, comunicação, ciências sociais, estudos de plataforma, ciência da informação, ciências humanas e direito. Estas foram as áreas consideradas relevantes para a etapa da pesquisa exploratória. Foram incluídos livros, teses, dissertações e artigos como materiais elegíveis.

Definiu-se o recorte temporal de cinco anos para a análise, contemplando, assim o período compreendido dezembro de 2020 e dezembro de 2025. O período foi definido com o intuito de contemplar as produções mais recentes sobre os temas analisados. Dessa forma, torna-se possível abranger maior repertório de moderação de conteúdo, legislações

mais atualizadas relacionadas à desinformação e regulação das plataformas digitais, além das repercussões das eleições presidenciais de 2018 e 2022.

A busca da bibliografia foi realizada de forma ampla, com o objetivo de sintetizar os estudos prévios potencialmente relevantes para os fins da pesquisa. Foram buscadas pesquisas que apresentassem resultados consolidados e submetidas a processos rigorosos da comunidade científica (Kitchenham et al., 2009).

Para tal, fez-se um processo rigoroso de coleta e análise, a partir de estudos já disponíveis. Os estudos selecionados compuseram o corpus final da análise. Por meio da revisão, tornou-se possível apontar evidências científicas consistentes sobre as lacunas de conhecimento dos temas investigados em pesquisas realizadas anteriormente. São eles: desinformação político-eleitoral e moderação de conteúdo.

Considerou-se a busca da literatura concluída ao verificar que a pesquisa repetida das palavras-chave em datas diferentes, atingiu os mesmos resultados. Neste primeiro momento, a busca nas bases de dados resultou em 41 estudos para triagem. Deste total, 17 estudos foram excluídos na etapa de seleção prática a partir de leitura de títulos e resumos. Esses estudos foram excluídos por não atenderem aos critérios de inclusão adotados. Assim, 24 estudos foram selecionados para leitura na íntegra e, posteriormente, avaliação de qualidade.

Alguns critérios de inclusão/exclusão foram adotados a priori da coleta dos dados. Alguns deles são os critérios de ano, língua de produção, alguns textos poderiam ser excluídos pelos filtros dos próprios sites. A definição destes critérios teve como objetivo a redução do conjunto de artigos para um número gerenciável e a manutenção de apenas estudos efetivamente relevantes para a pesquisa.

Em relação aos critérios de exclusão definidos durante a busca da bibliografia, foram estipulados: pesquisas que fogem do escopo do cenário político brasileiro e não têm incidência sobre a questão de pesquisa; pesquisas fora da delimitação das áreas designadas; estudos não disponibilizados inteiramente ou parcialmente online, ou seja, podem ser visualizados apenas mediante acesso privado; pesquisas fora do recorte de tempo analisado.

Artigos que não informavam no título as palavras-chave designadas ou a desinformação política ou a moderação de conteúdo como principal objeto, foram considerados dentro do escopo somente se ao longo do texto debatiam explicitamente essas ideias. Artigos que mencionam “desinformação” ou “moderação” sem aprofundar-se ou relacioná-la ao cenário brasileiro podem aparecer na coleta inicial de artigos, mas

não englobam os dados designados para análise. Além disso, foi verificado que não havia artigos duplicados selecionados.

O trabalho do processo de seleção foi realizado com o auxílio de planilhas eletrônicas, neste caso, Microsoft Excel. O detalhamento desses procedimentos e análises referentes a essa etapa estão disponíveis no Quadro 2 a seguir:

Quadro 2: Termos mobilizados na busca

Base de dados	Termos buscados	Termos relacionados aceitos no título	Operadores booleanos
Google Acadêmico	“desinformação política”; “moderação de conteúdo”	“desinformação”; “desinformação eleitoral”; “democracia”; “eleições”; “plataformas”; “regulação”	AND
SSRN	“desinformação política”; “moderação de conteúdo”	“desinformação”; “desinformação eleitoral”; “democracia”; “eleições”; “plataformas”; “regulação”	AND

Fonte: elaboração própria

Durante a seleção prática, os artigos encontrados não foram analisados de forma aprofundada. Nesta etapa, foram adotados critérios objetivos e práticos para a seleção ou a exclusão de artigos.

Em seguida, prosseguiu-se para a extração de dados. A extração de informações consistiu na reunião das informações qualitativas, ou seja, dos conceitos teóricos, relações e explicações relativos aos temas estudados, para fomentar a discussão e obter-se resultados.

4.2 Processo de Avaliação

Após a seleção dos artigos, buscou-se avaliar criticamente a sua qualidade metodológica enquanto produção científica. Reconhece-se, conforme explicado acima, que a seleção prática não exclui necessariamente artigos de baixa qualidade. Considerando também que podem ocorrer falhas, foi realizada uma revisão manual dos dados coletados.

Assim, foi preciso examinar novamente a lista de artigos selecionados de forma mais cuidadosa, garantindo que os estudos contemplados atendam a vários padrões de qualidade. Durante a avaliação, buscou-se priorizar os artigos de acordo com sua qualidade. Estudos que não possuem alta qualidade metodológica ou formato incompatível foram desclassificados após o processo de avaliação. Incompatibilidade, neste caso, aplica-se a formatos diferentes daqueles adotados em artigos redigidos pautados pelas normas ABNT. Pôsteres ou apresentações em slides, por exemplo, foram desconsiderados.

A etapa final de inclusão contou então com a seleção de um total de 17 estudos para compor o processo de avaliação, síntese dos dados e redação dos resultados. Buscou-se documentar e relatar em detalhes a conduta adotada em cada etapa, a fim de garantir precisão e imparcialidade.

Para definir a elegibilidade dos artigos, os referidos títulos e resumos foram lidos integralmente para entender a existência, ou não, de um debate e/ou proposição de combate à desinformação. Assim, sete publicações foram enquadradas como não elegíveis.

Após a avaliação final dos textos completos, 17 estudos foram excluídos por motivos como inadequação ao objetivo da pesquisa, ausência de dados empíricos ou não atendimento aos critérios metodológicos estabelecidos. Dessa forma, 17 estudos compuseram a amostra final da revisão sistemática.

Após a avaliação de qualidade e da metodologia, se prosseguiu para a avaliação do risco de viés. Cada estudo foi analisado frente aos critérios metodológicos previamente estabelecidos, a fim de identificar a presença de quaisquer vieses de seleção e aferição. Após a avaliação da qualidade metodológica e dos riscos de evidências dos estudos, prossegue-se para a redação e discussão dos resultados.

4.3 Limitações do Estudo

Reconhece-se, conforme introduzido no início do capítulo, que a pesquisa, assim como qualquer revisão sistemática, apresenta limitações.

Primeiramente, pode-se apontar a restrição às bases indexadas analisadas, escolhidas devido à sua abrangência e relevância. Produções não indexadas podem ter sido, portanto, desconsideradas. O recorte temporal de cinco anos adotado também pode

ser considerado uma limitação de trabalho, uma vez que pesquisas anteriores ou posteriores ao período analisado não são analisadas.

Ademais, embora a revisão sistemática tenha sido conduzida de forma imparcial, foi realizada por uma única pesquisadora. Assim, não houve o envolvimento de uma equipe de revisores ou uma dupla de revisores independentes que pudesse comparar os resultados encontrados.

A situação descrita não invalida o estudo, mas pode ser considerada uma limitação metodológica. Visando reduzir um possível viés de seleção e interpretação, foram utilizadas estratégias. Dentre estas, destaca-se o uso de critérios de inclusão e exclusão bem definidos e o registro detalhado e completo do processo de busca.

Apesar de haver limitações inerentes ao estudo, há total consistência metodológica no protocolo adotado. Obtém-se, então, um panorama estruturado acerca da produção científica brasileira sobre os temas da desinformação política e da moderação de conteúdo no cenário brasileiro.

Após a apresentação dos caminhos metodológicos, serão apresentados os resultados da análise dos dados. Além dos resultados, serão discutidas as futuras para no enfrentamento ao fenômeno desinformativo explicitados pela revisão sistemática de literatura.

5 Resultados

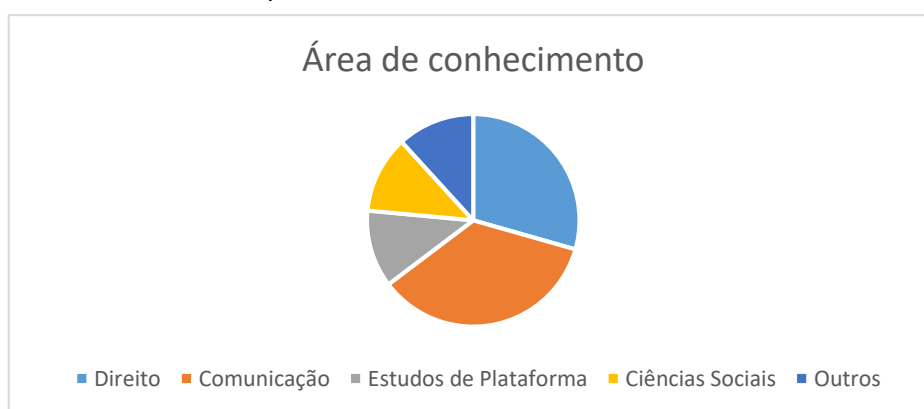
5.1 Análise de Resultados

Ao longo deste capítulo, os resultados da revisão sistemática realizada serão analisados e discutidos. Nesta primeira seção, os resultados obtidos serão brevemente introduzidos e descritos. O intuito deste capítulo é apresentar uma visão abrangente e ao mesmo tempo sintética das descobertas a partir dos artigos relevantes considerados.

Em relação à análise dos dados, algumas questões iniciais podem ser observadas antes do aprofundamento da discussão, com seus potenciais e limites. A partir dos resultados alcançados, haverá uma subseção para discussão

Os 17 estudos selecionados apresentam diversidade quanto ao contexto geográfico, área de interesse e período de publicação entre os anos 2020 e 2025, conforme explicado anteriormente. Considerando os estudos, artigos e livros considerados relevantes, as áreas de interesse que surgem com maior frequência são a Comunicação e o Direito. Conforme exibido no gráfico quantitativo descritivo abaixo, as discussões pertencentes aos dois campos orientam a maior parte das discussões do capítulo de resultados. Tal centralidade da discussão atribuída aos campos do direito e da comunicação, indica as tendências metodológicas dos estudos brasileiros sobre a regulação das plataformas digitais.

Quadro 3: Mapeamento das áreas de conhecimento



Fonte: elaboração própria

Ao aplicar filtros de valor na planilha, pôde-se observar o crescimento exponencial do interesse acadêmico nos temas de estudo ao longo dos últimos anos,

sobretudo a partir de 2023. Dentre os 17 estudos selecionadas, a maioria foi publicada nos últimos três anos: oito foram publicados no ano de 2025 e quatro foram publicados no ano de 2024.

Esse crescimento pode estar diretamente relacionado ao aumento da desinformação política no Brasil ao longo do período analisado e à repercussão das eleições de 2022. É válido apontar que os resultados encontrados demonstram a constância do interesse da comunidade científica brasileira ao longo do período analisado.

Abaixo, no Quadro 4, é apresentada a lista de artigos selecionados após as avaliações, junto às suas respectivas datas de publicação. Além dos títulos dos estudos, há uma coluna apresentando a sua data de publicação. Este quadro foi feito a partir de uma tabela criada por meio do software de planilhas eletrônicas Microsoft Excel.

Quadro 4: Estudos selecionados e respectivos anos de publicação

TÍTULO	ANO
Controle Via Agência em Plataformas Algorítmicas	2020
A desinformação nas plataformas digitais e a “política de informações enganosas sobre a covid-19” do Twitter: desafios e perspectivas da moderação de conteúdo	2023
Desinformação política: análise dos posts de Eduardo Bolsonaro nas eleições de 2022	2024
Estratégias desinformacionais nas eleições: o uso das redes sociais por Pablo Marçal em 2024	2025
Desinformação eleitoral em plena campanha: nuances de pós-verdade e populismo no coração da democracia brasileira.	2024
Desinformação e processo eleitoral brasileiro: enfrentamento à luz da liberdade de expressão e informação	2023
A rede da desinformação: sistemas, estruturas e dinâmicas nas plataformas de mídias sociais	2024
"Cadê o código-fonte?": a urna eletrônica no centro da campanha de desinformação na eleição presidencial brasileira de 2022	2025
O hexágono da desinformação: uma proposta de framework para pesquisas empíricas	2025
Democracia em risco: <i>fake news</i> , censura e a procuradoria nacional da união de defesa da democracia	2025
Desinformação: impactos das <i>fake news</i> na democracia brasileira	2023
Fascistização, <i>fake news</i> e atos antidemocráticos no brasil: considerações sobre a necessidade de responsabilização e punição ...	2024
Regulação das redes sociais e das plataformas digitais no brasil (2010–2025: governança digital em perspectiva comparada	2025
Como as plataformas se apropriam de práticas jornalísticas de checagem: um olhar para a comunidade de notas	2025
A guerra da (des)informação: o papel das <i>fake news</i> durante o processo eleitoral presidencial de 2018	2025
Direito à liberdade de expressão versus direito à informação: em busca da construção da cidadania e da compreensão esclarecida na democracia brasileira	2021
A ciência no enfrentamento da desinformação: uma revisão sistemática	2025

Tais resultados orientam a discussão, tanto geral quanto específica, apresentada nesse capítulo. Eles possibilitam a criação das categorias principais que compõem os subtítulos descritos a seguir.

5.2 Discussão dos Resultados

Ao longo desta seção, os artigos pertinentes serão devidamente apresentados e discutidos, assim como os resultados da revisão sistemática realizada. O intuito é oferecer uma visão abrangente e, ao mesmo tempo, resumida dos artigos apontados como relevantes. Assim, serão discutidas as principais características temáticas, as similaridades e concordâncias teóricas similares entre os artigos e as potenciais controvérsias encontradas, a partir de uma abordagem qualitativa.

Todos os 17 estudos selecionados foram publicados entre 2020 e 2025. Desta forma, a busca reflete o panorama sócio-político vigente até o fim de 2025, deixando de necessariamente refletir particularidades do ano de 2026. Assim, cabe ressaltar que a pesquisa não é necessariamente aplicável a questões específicas de 2026, uma vez que foi realizada anteriormente a esse período.

5.3 Análise e Discussão dos Artigos

Ao longo desta seção será discutido o conjunto de artigos considerados pertinentes para a pesquisa, informando sinteticamente suas principais características temáticas, tendências teóricas comuns e controvérsias. As discussões dos artigos serão agrupadas em torno de dois focos de pesquisa quanto à desinformação política e a à moderação de conteúdo: a) a desinformação política no Brasil; b) a discussão nacional sobre moderação de conteúdo nas plataformas digitais e c) discussão sobre legislações.

Todos os autores escolhidos abordam a temática de pesquisa da desinformação política e a moderação de conteúdo. Cada um a explora a partir de uma perspectiva diferente, considerando que há diferenças metodológicas entre os estudos e ao contexto acadêmico em que foram conduzidos.

5.4 Discussão Nacional sobre a Desinformação Política

Albuquerque (2024) aponta o fato de que a desinformação como recurso para se obter vantagens não é um fenômeno novo. A discussão sobre desinformação se torna relevante no cenário sócio-político brasileiro devido às suas novas dimensões. Há outros fatores descritos pelo autor que ampliam o impacto da atividade desinformativa, como a lógica de operação das plataformas e o uso de agentes não humanos, como bots, para manipular a opinião pública.

Recuero (2024) define a desinformação como um fenômeno sistêmico complexo. A autora a classifica no estudo como um fenômeno decorrente da comunicação, capaz de gerar caos e dano, e, de forma geral, desestruturar o sistema social. Segundo Recuero, geralmente é praticada com o intuito de beneficiar algum indivíduo, grupo ou sistema.

Silva (2025) e Recuero (2024) defendem nos seus textos que a desinformação gera consequências de impacto na sociedade (Recuero, 2024). Ao gerar crises epistêmicas, a desinformação afeta a estabilidade e gera efeitos como o aprofundamento da polarização, quanto para crescimento da ameaça à democracia.

Santos (2024) definiu no seu texto a desordem informacional como todo conteúdo publicado com a intenção de gerar enganos e, conseqüentemente, causar confusão ambiente digital e, mais especificamente nas plataformas.

A apresentação da desordem informacional como fenômeno sistêmico está em consonância com as definições apresentadas ao longo do capítulo de desinformação. A crise de autoridade das instituições, por exemplo, é uma consequência direta da desinformação e, simultaneamente, contribui para a sua deflagração na sociedade brasileira. A falta de credibilidade, como apontaram Benett e Livingston (2021), gera então um vácuo de autoridade que propicia o surgimento de novos atores e narrativas alternativas.

Dentre os atores políticos mencionados nos textos analisados, alguns nomes se destacam. São mencionados alguns dos principais atores e candidatos da política brasileira nos últimos anos, como Pablo Marçal, Eduardo Bolsonaro, Fernando Haddad e Jair Bolsonaro.

Lima (2025) faz referência aos dois últimos para abordar o impacto da desinformação política sobre as eleições presidenciais de 2018. Silva (2025) analisa as estratégias desinformativas utilizadas por Pablo Marçal durante a campanha eleitoral para a prefeitura de São Paulo em 2024. Santos (2024), alude à desinformação política

veiculada nas redes sociais como estratégia de campanha de Eduardo Bolsonaro nas eleições presidenciais de 2022.

Já Albuquerque (2024) e Castro (2020) citam Jair Bolsonaro e seus aliados políticos como, respectivamente, causadores do debate sobre desinformação no Brasil e grupo como montados por correntes políticas e campanhas eleitorais. Luiz Inácio Lula da Silva também é citado por Vicari e Iasulaitis (2025), considerando que foi adversário de Jair Bolsonaro nas eleições presidenciais de 2022 e vencedor.

Vicari e Iasulaitis (2025), por sua vez, demonstraram como as eleições presidenciais brasileiras de 2022 foram marcadas por uma intensa campanha de desinformação protagonizada pela direita, sobretudo nas redes sociais. Ambas apontam esse quadro como parte um fenômeno internacional contemporâneo associado à ascensão da extrema-direita em diversos territórios. Para reforçar seu ponto, aludem às campanhas de desinformação do *Brexit*, e das eleições presidenciais de Donald Trump nos Estados Unidos, em 2016.

Pode-se observar, em consonância com as semelhanças entre as informações descritas acima, que durante períodos eleitorais há maior impacto da desinformação política. As campanhas eleitorais, tanto estaduais quanto federais, vêm sendo marcadas pela presença crescente de estratégias desinformativas. Neste contexto, fala-se no fenômeno da desinformação político-eleitoral, explicado anteriormente ao longo dos capítulos teóricos.

Além das eleições presidenciais de 2018 e 2022, Recuero cita o episódio de tentativa de golpe de Estado conhecido como “8 de janeiro”, ocorrido em 8 de janeiro de 2023. O evento composto por apoiadores do então candidato derrotado Jair Bolsonaro foi motivado por narrativas desinformativas. As narrativas criadas envolviam boatos falsos sobre fraude eleitoral e descrédito de instituições democráticas governamentais, como os tribunais eleitorais.

Vicari e Iasulaitis (2025) também apontam no seu texto como a veiculação de conteúdos que afirmavam que o código-fonte das urnas eletrônicas teria configurado uma fraude eleitoral influenciou a opinião de parte dos cidadãos brasileiros. A desconfiança fomentada por Bolsonaro e parlamentares bolsonaristas teria motivado a recusa em aceitar o resultado da disputa presidencial, o que, eventualmente, incitou manifestações de caráter antidemocrático. Tais manifestações culminaram nos ataques ao Congresso Nacional e ao Supremo Tribunal Federal (STF) no dia 8 de janeiro de 2023.

Batista (2023) traz como enfoque da sua pesquisa os eventos das eleições presidenciais brasileiras de 2018 e 2022. Parte-se do pressuposto de que desinformação não contribui para o debate coletivo, além de comprometer a integridade democrática caso não seja refreada pela regulamentação das plataformas digitais, seja por iniciativa do Estado ou das próprias plataformas digitais. Este achado sugere trazer a regulação e a moderação de conteúdo nas plataformas digitais como medida de combate à desinformação política, no lugar de efetivamente criminalizar a prática.

Recuero (2024) também sugere a regulação e a moderação de conteúdo como formas de combate aos sistemas desinformativos, além da prática da checagem de fatos. Recuero e Ferrarezi (2023) sugerem como alternativa o letramento digital e informacional. Entretanto, reconhece-se que, no Brasil, as iniciativas públicas de fomento à educação midiática ainda são tímidas, em contraste com os imensos desafios que assolam o país no que diz respeito à desinformação.

Canela (2025) indica o letramento como possibilidade para legisladores, políticos, educadores e acadêmicos ao fomentar debates sobre a implementação da educação contra a desinformação no país. Já Batista (2025), defende que há necessidade de regulamentação das plataformas de redes sociais para que haja maior transparência e responsabilidade no combate a conteúdos falsos e enganosos e a quem os difunde.

Albuquerque (2024) menciona a dificuldade de delimitação do conceito de informação como um fator que afeta o combate efetivo à atividade desinformativa. Há dificuldade de se conceituar o que é desinformação de forma ampla, para além da definição que contempla o espalhamento de notícias falsas ou enganosas.

Silva (2025), de forma similar, defende que, se a desinformação opera como um sistema complexo e adaptativo, são necessárias ferramentas sofisticadas e integradas para compreendê-la e combatê-la. Para se conceituar a desinformação, os autores defendem que é preciso reconhecer a sua complexidade, superando a visão limitada do fenômeno como um simples conjunto de conteúdos falsos.

5.5 Discussão Nacional sobre a Moderação de Conteúdo e Plataformas

Foram analisadas as plataformas digitais abordadas em cada artigo de modo a compreender quais as que concentram a atenção dos pesquisadores. Dentre as plataformas

digitais mencionadas nos textos analisados, algumas se destacam e são mencionados na maior parte dos textos analisados. Identificou-se o predomínio do chamado GAFAM¹⁴. O Facebook foi mencionado mais frequentemente (acima de seis vezes), em seguida o X, o Google e o YouTube (acima de três vezes). O nome Meta também foi mencionado como um sinônimo da junção entre Facebook, Instagram, WhatsApp, Messenger e Threads.

Santos (2024) aponta as gigantes da tecnologia (*big techs*), como o Google e a Meta como ferramentas centrais no espalhamento da desinformação em contextos políticos, especialmente em temporadas eleitorais. Batista (2023), por outro lado, afirma que as plataformas vêm sendo estimuladas pela sociedade a atuar em parceria com as instituições democráticas para a defesa da democracia.

Trata-se de parcerias. Santos menciona: o Programa de Enfrentamento à Desinformação em conjunto com o TSE; a criação de rótulos no Facebook e no Instagram para divulgar informações oficiais sobre o sistema de votação; a implementação de um canal de denúncias para informar à Meta sobre conteúdos de desinformação eleitoral; a criação de uma página com as tendências de pesquisa decorrentes do *Google Search* sobre as eleições, com a divulgação do Relatório Google Trends Eleições 2022.

Conforme relatado no subtítulo anterior, a maior parte dos autores analisados descreve a moderação de conteúdo como medida importante de combate aos sistemas desinformativos.

Batista (2023) traz a regulação e a moderação de conteúdo nas plataformas digitais como essenciais. Castro (2020), por sua vez, argumenta que a moderação pode ser eficiente a nível individual, mas ser contornada por grupos organizados. Estes grupos, conhecendo suas diretrizes, podem recorrer a artifícios como contas falsas, robôs e discursos de ódio velados em meio à defesa de correntes políticas e campanhas eleitorais. Recuero, em contraponto, alerta sobre a necessidade de atualização da moderação de conteúdo, considerando que o sistema desinformativo está em constante adaptação.

Ferrarezi (2023) aponta que o debate acerca da responsabilização dos intermediários divide especialistas. De acordo com seu texto, há quem defenda que manter as plataformas como intermediárias impede o enfrentamento da desinformação e quem defende que pode ser perigoso concentrar mais responsabilidade nas plataformas, permitindo que tenham autonomia sobre quais informações podem ou não circular.

¹⁴ Acrônimo para as cinco maiores empresas de tecnologia dos EUA, que dominam o mercado mundial: Google (Alphabet), Apple, Facebook (Meta), Amazon e Microsoft.

Entretanto, é importante que se leve em conta a perspectiva oferecida por Tarleton Gillespie (2017), argumentada no marco teórico. Gillespie argumenta que manter a classificação das plataformas como “intermediárias” afeta diretamente a governança das mesmas. Discute-se sobre a conveniência de isentá-las de assumir maior responsabilidade.

Concilia os pontos de vista apresentados acima, pode-se inferir que é necessária a responsabilização das plataformas, mas que é possível também assegurar a liberdade de expressão e segurança no ambiente virtual. Esse poderia vir a ser um caminho para abordar a regulação das plataformas no futuro.

5.6 Discussão Nacional sobre Legislações

Em relação à legislação nacional, o principal documento mencionado é o Marco Civil da Internet. A partir dos textos, alguns autores criticam a discrepância entre a legislação desatualizada (o Marco Civil da Internet), e a realidade brasileira atual.

Fala-se, assim como abordado nos capítulos teóricos da presente dissertação, sobre a necessidade de haver efetivamente a responsabilização das plataformas digitais, que deixaram de ser apenas intermediárias. Canela (2025) argumenta pela revisão do Marco Civil da Internet, de forma a estabelecer limites regulatórios para a proliferação de conteúdos nocivos à sociedade, como a desinformação.

Outras legislações nacionais citadas nos textos analisados, consideradas relevantes, são a Lei Geral de Proteção de Dados (LGPD) e o PL nº 2.630/2020, apontado por Silva (2025).

Ferrarezi apurou que, no período entre 2018 e 2022, foram apresentados 100 Projetos de Lei (PL) relacionados aos temas de “*fake news*” e “desinformação”, sendo 44 deles anexados ao PL 2630/20. A quantidade de projetos demonstra que os temas despertam crescente interesse público.

6 Considerações Finais

A presente dissertação buscou contribuir para o debate sobre o impacto da disseminação da desinformação sobre a democracia brasileira e explicar as dinâmicas da moderação de conteúdo. Este estudo não representa um ponto final, mas um convite à continuidade da investigação sobre a desinformação e a moderação de conteúdo. As considerações feitas ao longo desta seção baseiam-se na análise conduzida ao longo dessa dissertação por meio da RSL e da fundamentação teórica.

O objetivo da revisão sistemática de literatura conduzida foi refletir acerca da literatura nacional sobre os temas explicados acima. Assim, desta forma, tornou-se possível identificar tendências comuns e lacunas, além de sugerir novas ideias a serem exploradas.

Foram produzidos dois capítulos teóricos, evidenciando a relação existente entre os temas da desinformação e da moderação de conteúdo ao longo dos últimos anos. Entende-se que a desinformação política afeta a integridade democrática nas plataformas digitais e que a moderação de conteúdo, se ineficaz, pode fomentar a disseminação da desinformação e contribuir para um quadro de crise político-social. Como consequências do quadro de crise, há a desestabilização de instituições, a veiculação de informações enganosas e o estímulo a atos antidemocráticos.

Assim, foi possível relacionar e apresentar um repertório extenso à academia brasileira, assim como contribuir para a ampliação da discussão epistemológica sobre desinformação, moderação e governança das plataformas digitais.

Para compreender e combater a desinformação, é preciso enxergá-la como um sistema complexo que opera de forma adaptativa e é capaz de desestabilizar a sociedade. Apenas entendendo que é necessário superar a visão do fenômeno como um conjunto de conteúdos falsos, torna-se possível combatê-la.

Conforme observado, o fenômeno da desinformação tem raízes para além do espaço virtual. A desinformação política não é uma novidade na história da política brasileira nem em períodos de eleição. Entretanto, nos últimos anos vem adquirindo novas características, que a elevaram a uma posição de centralidade em debates atuais.

Primeiramente, seu potencial foi amplificado pelo contexto global político de campanhas de desinformação, fenômeno associado à direita observado, por exemplo, com Donald Trump nos Estados Unidos, o Brexit no Reino Unido e Jair Bolsonaro no Brasil.

Há fatores que ainda representam obstáculos para o alcance da integridade democrática. O contexto histórico e político-social no qual o Brasil está inserido, que facilita a desinformação antes mesmo da existência de redes sociais; eventos que contribuem para o crescimento da polarização política.

Outra limitação é o modelo operacional das plataformas, que incentiva a manutenção da desinformação. Enquanto estas possuem Termos de Uso robustos que englobam a desinformação e, mais especificamente, a desinformação eleitoral, adotam modelos que, na prática, priorizam a lógica algorítmica, o que poderia conferir maior visibilidade a conteúdos desinformativos.

Não há clareza também quanto aos critérios adotados pelas plataformas para regulação e moderação, algo que dificulta um combate mais efetivo à desinformação nas plataformas e a produção de estudos sobre a questão.

Entretanto, há iniciativas que vem ganhando força nos últimos anos, como as estratégias de moderação de conteúdo. A moderação de conteúdo vem se mostrando cada vez mais importante para o desmantelamento da estrutura da desinformação. Fatores como investimento em recursos, parcerias e em desenvolvimento tecnológico, além do uso da automatização para dar conta de respostas mais eficientes em tempo real, podem ser determinantes para minar a disseminação da desinformação.

A partir dos resultados, pode-se concluir que a regulação das plataformas digitais é um tema que vem adquirindo crescente destaque nos campos de conhecimento do Direito e da Comunicação no Brasil. A crescente produção de matérias nos últimos dois anos, conforme evidenciado pela RSL, revela que é um tema de interesse público.

Ademais, a lógica de operação das plataformas digitais, utilizadas para campanhas políticas eleitorais, modificou significativamente as características e o impacto da desinformação política. Reconhece-se que eventos como eleições, sobretudo as eleições presidenciais de 2018 e 2022 são marcados pelo aumento expressivo da veiculação da desinformação política nas plataformas.

Conclui-se que as plataformas digitais precisam ser responsabilizadas para além do papel de intermediárias, dada a sua relevância para a fomentação da desinformação política. Entende-se que a responsabilização poderia ser benéfica para reduzir o alcance e os efeitos da desinformação política.

Dentre outras possíveis medidas de combate à desinformação política, são mencionados o letramento digital e a maior atenção às estratégias utilizadas para burlar a moderação de conteúdo.

É preciso associar essa responsabilização a uma legislação brasileira mais atualizada e eficiente, que seja capaz de abranger a rápida evolução da desinformação política e a relevância das plataformas. É necessário que se atualize a legislação nacional relacionada à internet, para que se equipare às novas normas internacionais adotadas e para que se possa atender as novas demandas, necessidades e urgências que surgiram em decorrência dos impactos da evolução tecnológica observada nos últimos anos.

É preciso também que haja um enfoque na moderação de conteúdo, a fim de contribuir para a e garantir que haja regras eficazes e efetivamente governança das plataformas. Afinal, a moderação de conteúdo ineficaz contribui para a deflagração da atividade desinformacional político-eleitoral. É imprescindível para assegurar governança estimular a mobilização conjunta de agências de checagem, órgãos governamentais, organizações, grupos de pesquisa, veículos midiáticos e *think tanks* especializados.

7 Referências Bibliográficas

AHMAD, S. *Who moderates my social media?* Locating Indian workers in the global content moderation practices. In C. Strippel, S. Paasch-Colberg, M. Emmer & J. Trebbe (Eds.), *Challenges and perspectives of hate speech analysis*, p. 74-88, 2023.

AHMAD, Sana. *Who moderates my social media?* Locating Indian workers in the global content moderation practices. In Strippel, C. et al (orgs) *Challenges and Perspectives of Hate Speech Research*, v. 12, p. 111-125, 2023.

ALBUQUERQUE, A. 2021. “As Fake News e o Ministério da Verdade Corporativa.” *Revista Eptic* 23 (1), p.124–141.

ALENCAR, M.; AQUINO, M. C. *Algoritmos e desinformação: automação da checagem na América Latina: Automation of Checks in Latin America. Animus. Revista Interamericana de Comunicação Midiática*, v. 21, n. 47, 2023. Disponível em: <<https://periodicos.ufsm.br/animus/article/view/72211>>. Acesso em: 02 de março de 2025.

ALTAY, Sacha; BERRICHE, Manon; ACERBI, Alberto. *Misinformation on misinformation: Conceptual and methodological challenges. Social Media+ Society*, v. 9, n.1, 2023. Acesso em: janeiro de 2025.

ALVES, Marcelo. *Governança nas Plataformas: a Moderação de Vídeos do Youtube da Eleição de 2022 à Insurreição Golpista*. In: *Anais Do 32º Encontro. Anual da Compós*, Campinas, Galoá, 2023. Disponível em: <<https://proceedings.science/compos/compos-2023/trabalhos/governanca-nas-plataformas-a-moderacao-de-videos-do-youtube-da-eleicao-de-2022-a?lang=pt-br>>. Acesso em: fevereiro de 2025.

BATEMAN, J. JACKSON, D., 2024. *Countering Disinformation Effectively: An Evidence-Based Policy Guide*, CEIP: Carnegie Endowment for International Peace. United States of America. <<https://coilink.org/20.500.12592/rjdfszt>> Acesso em: 15 de julho 2025.

BATISTA, Amanda Simões da Silva. *Desinformação e processo eleitoral brasileiro: enfrentamento à luz da liberdade de expressão e informação*, 2023.

BENTES, Anna. *Quase um tique: economia da atenção, vigilância e espetáculo em uma rede social*. Rio de Janeiro: Ed. UFRJ, 2021.

BRASIL. *Lei nº 12.965, de 23 de abril de 2014*. Brasília, DF: Planalto, 2014. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato20112014/2014/lei/112965.htm>. Acesso em: 10 de fevereiro de 2025.

BRASIL. *Lei nº 14.192, de 4 de agosto de 2021*. Estabelece normas para prevenir, reprimir e combater a violência política contra a mulher; e altera a Lei nº 4.737, de 15 de julho de 1965 (Código Eleitoral), a Lei nº 9.096, de 19 de setembro de 1995 (Lei dos Partidos Políticos), e a Lei nº 9.504, de 30 de setembro de 1997 (Lei das Eleições). Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/114192.htm>. Acesso em: 10 de fevereiro de 2025.

BRASIL. [Câmara dos Deputados]. *Projeto de Lei nº 2630, de 2020*. Ementa: Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Disponível em: <<https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735>> ou Câmara. Acesso em: 10 de fevereiro de 2025.

CANELA, Mateus Ferrari. *Bacharelado em Ciências Sociais*.

CASTRO, J. C. L. *Controle via agência em plataformas algorítmicas*. *Galáxia*, São Paulo (SP), n. 44, p. 144-157, maio/agosto de 2020.

CRESTANI, Pillar Cornelli; OLIVEIRA, Rafael Santos de. *A desinformação nas plataformas digitais e a “política de informações enganosas sobre a covid-19” do Twitter: desafios e perspectivas da moderação de conteúdo*. *Revista de Estudos Jurídicos UNESP*, v. 27, n. 45, p. 5, 2023.

CURY LUIZ, Thiago. *Desinformação eleitoral em plena campanha: nuances de pós-verdade e populismo no coração da democracia brasileira*. *Estudos em Comunicação*, n. 38, 2024.

CURZI, Yasmin; HARTMANN, Ivar; ALMEIDA, Clara; ZINGALES, Nicolo. *Moderação de conteúdo online: contexto, cenário brasileiro e suas perspectivas regulatórias*. São Paulo: Alameda, 2022.

DE KEULENAAR E., BURTON A. G., KISJES I. (2021). *Deplatforming, demotion and folk theories of big tech persecution*. *Fronteiras—Estudos Midiáticos*, 23(2), p. 118-139. Disponível em: <<https://doi.org/10.4013/fem.2021.232.09>>. Acesso em: janeiro de 2025.

DONATO, Helena e DONATO, Mariana (2019) “Etapas na condução de uma revisão sistemática”, em *Acta Médica Portuguesa*, Vol. 32, Nº 3.

DOURADO, Tatiana Maria Silva Galvão. *Fake news na eleição presidencial de 2018 no Brasil*. 2020. Disponível em: <https://repositorio.ufba.br/bitstream/ri/31967/1/Tese_Tatiana%20Dourado.pdf>. Acesso em: 10 de junho de 2025.

Duke Reporters' Lab. (2025). *2025 census: Fact-checkers persevere as politicians, platforms turn up heat*. Reporters' Lab, 2025. Disponível em: <<https://reporterslab.org/2025/06/19/fact-checkers-persevere-as-politicians-platforms-turn-up-heat/>>. Acesso em: janeiro de 2025.

D'ALMONTE, E. F., & SANTOS, A. O. (2024). *Regulamentação das plataformas digitais: entre a soberania digital e o transnacionalismo*. *E-COMPÓS*, 27. Disponível em: <<https://doi.org/10.30962/ecomps.2876>>. Acesso em: 10 de fevereiro de 2025.

D'ANDREA, C. F. de B.; HENN, R. *Desinformação, plataformas, pandemia: um panorama e novos desafios de pesquisa*. *Revista Fronteiras - estudos midiáticos*, v. 23, n. 2, 2021.

FERRAREZI, Renata Abibe. *Desinformação: Impactos das Fake News na Democracia Brasileira*. 2023.

FUKUYAMA, Francis. *What is governance? Governance - An International Journal of Policy, Administration, and Institutions*, v. 26, n. 3, p. 347-368, 2013.

GILLESPIE, T. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, 2018.

GILLESPIE, T. et al. *Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates. Internet Policy Review*, v. 9, n. 4, 21. 2024.

GILLESPIE, T. *The politics of 'platforms'. New Media & Society*, v.12, n.3, p. 347-364, 2010. Disponível em: <<https://doi.org/10.1177/1461444809342738>>. Acesso em: outubro de 2024.

GILLESPIE, T., AUFDERHEIDE, P.; CARMÍ, E.; GERRARD, E.; GORWA, R.; MATAMOROS-FERNÁNDEZ, A.; ROBERTS S.; SINNREICH, A.; WEST, S. *Expanding the debate about content moderation: scholarly research agendas for the coming policy debates. Internet Policy Review*, v.9, n.4, 2020. Disponível em: <<https://policyreview.info/articles/analysis/expanding-debate-aboutcontent-moderation-scholarly-research-agendas-coming-policy>>. Acesso em: 18 de dezembro de 2024.

GORWA, R. *What is platform governance? Information, Communication & Society*, 22(6),854–871,2019. Disponível <<https://doi.org/10.1080/1369118X.2019.1573914>> Acesso em: 10 de janeiro de 2025.

GRAVES, L, NYHAN, B., REIFLER, J. *Understanding Innovations in Journalistic Practice: A Field Experiment Examining Motivations for Fact-Checking, Journal of Communication*, Volume 66, Issue 1, February 2016, p. 102–138. Disponível em: <<https://doi.org/10.1111/jcom.12198>>. Acesso em: 01 de fevereiro de 2025.

GRAVES, L. *Boundaries not drawn. Journalism Studies*, p. 1-19, 2016. Disponível em: <<https://doi.org/10.1080/1461670X.2016.1196602>>. Acesso em: 01 de abril de 2025.

GRIMMELMANN, James. *The virtues of moderation. Yale JL & Tech.*, v. 17, p. 42, 2015.

HELBERGER, N., PIERSON, J., & POELL, T. (2017). *Governing online platforms: From contested to cooperative responsibility. The Information Society*, 34(1), p. 1–14. Disponível em: <<https://doi.org/10.1080/01972243.2017.1391913>>. Acesso em: 12 de janeiro de 2025.

JUNGHERR, A., SCHROEDER, R. *Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy*. *Social Media + Society*, v.7 n.1, 2021. Disponível em: <<https://doi.org/10.1177/2056305121988928>>. Acesso em: 18 de novembro de 2024.

KLONICK, Kate. *The New Governors: The People, Rules, and Processes Governing Online Speech*. *Harvard Law Review*. N. 131, 2018.

LELO, T. *Assessing the consistency of fact-checking in political debates*. *Journal of Communication*, v. 73, n. 6, pp. 587–600, 2023. Disponível em: <<https://doi.org/10.1093/joc/jqad028>>. Acesso em: 20 de julho de 2025.

LELO, T., and F. PACHI FILHO. *Credibilidade em disputa no jornalismo digital: Estratégias discursivas das agências de fact-checking brasileiras vinculadas aos conglomerados de mídia*, 2012. *Estudios sobre el Mensaje Periodístico* 27, p. 531–541.

LEMO, André. Plataformas, Dataficação e Performatividade Algorítmica (PDPA): desafios atuais da cibercultura. In: PRATA, Nair; PESSOA, Sônia Caldas (Orgs.). *Fluxos comunicacionais e crise da democracia*. São Paulo: Intercom, 2020, p. 107-116

LIMA, Teo Vieira et al. *A guerra da (des) informação: o papel das fake news durante o processo eleitoral presidencial de 2018*, 2025.

MANTZARLIS, A. 2015. *Introducing Poynter's International Fact-Checking Network*. *Poynter*, October 21. <http://www.poynter.org/2015/fact-checkers-of-the-world-unite/379716/>.

MELLADO, MOREIRA, LAGOS, E HERNANDÉZ. “*Comparing Journalism Cultures in Latin America: The Case of Chile, Brazil and Mexico*.” *International Communication Gazette* v.74 n;1, p. 60–77, 2012. Acesso em: 01 de junho de 2025.

MENA, P. *Principles and Boundaries of Fact-checking: Journalists' Perceptions*, *Journalism Practice*, v.13, n.6, p. 657-672, 2019. Disponível em: <<https://doi.org/10.1080/17512786.2018.1547655>> Acesso em: 01 de junho de 2025.

MORROW, SWIRE-THOMPSON, POLNY E WIHBEY. *The Emerging Science of Content Labeling: Contextualizing Social Media Content Moderation*, 2020. Disponível em: <<http://dx.doi.org/10.2139/ssrn.3742120>>.

NESBITT, T. (2022). W. Lance Bennett and Steven Livingston (eds.), *The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States*. *Journal of Communication Inquiry*, v.46 n.2, p. 225-229, 2022. Disponível em: <<https://doi.org/10.1177/01968599211041106>>.

OKOLI, C.; DUARTE, T. por: David W. A.; MATTAR, R. técnica e introdução: João. *Guia Para Realizar uma Revisão Sistemática de Literatura. EaD em Foco*, [S. l.], v. 9, n. 1, 2019. DOI: 10.18264/eadf.v9i1.748. Disponível em: <<https://eademfoco.cecierj.edu.br/index.php/Revista/article/view/748>>. Acesso em: 18 de dezembro de 2025.

Princípios de Manila. *Declaração dos Princípios de Manila sobre a Liberdade de Expressão*. Publicado por FORUM-ASIA, Article 19, AMARC, 2015.

RASMUS Kleis Nielsen; GANTER, Sarah Anne. *The Power of Platforms: Shaping Media and Society*, Oxford, Reino Unido: Oxford University Press, 2022, p. 272.

RECUERO, Raquel. *A rede da desinformação: Sistemas, estruturas e dinâmicas nas plataformas de mídias sociais*. Porto Alegre: Sulina, 2024.

SANTOS, Nathielly Maria Lima dos. *Desinformação Política: Análise dos Posts de Eduardo Bolsonaro nas Eleições de 2022*, 2024.

SILVA, Mariana Gomes da. *Estratégias desinformacionais nas eleições: o uso das redes sociais por Pablo Marçal em 2024*, 2025.

SILVA, Maurílio Luiz Hoffmann da. *O hexágono da desinformação: uma proposta de framework para pesquisas empíricas*, 2025.

SILVA, Tarcízio. (2012). *Para Entender o Monitoramento de Mídias Sociais*.

SINGHAL, MOHIT, et al. "SoK: Content moderation in social media, from guidelines to enforcement, and research to practice." *IEEE 8th European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2023.

TSUZUKI, Camila. ALVES, Marcelo. *Análise dos termos de uso das plataformas para publicidade digital e conteúdo*. Instituto Democracia em Xequê, 2024.

VICARI, Isabella; IASULAITIS, Sylvia. " Cadê o código-fonte?": a urna eletrônica no centro da campanha de desinformação na eleição presidencial brasileira de 2022. *Aurora. Revista de Arte, Mídia e Política*, v. 17, n. 51, p. 6-33, 2025.

WARDLE, Claire. *Guia Essencial da First Draft para Entender a Desordem Informacional*. First Draft, 2020.

WRIGHT, S., & STREET, J. (2007). *Democracy, deliberation and design: the case of online discussion forums*. *New Media & Society*, v.9, n.5, p. 849-869, 2007. Disponível em: <<https://doi.org/10.1177/1461444807081230>>. Acesso em: 20 de dezembro de 2024.