

Editorial poliTICs 40

O Nupef tem o prazer de colaborar na publicação de artigos que debatem os resultados da consulta do CGI.br sobre regulação de plataformas digitais no Brasil, realizada em 2023.¹ A consulta originalmente recebeu um total de 1.336 contribuições de pessoas e organizações dos quatro setores que compõem o CGI.br (setor governamental, terceiro setor, setor empresarial e comunidade científica e tecnológica).

Em seguida, o CGI.br publicou um relatório de sistematização dos resultados, para apresentá-los de forma mais acessível e sintética.² No entanto, as extensas contribuições dadas pela sociedade brasileira permitem diversas interpretações e análises por metodologias diversas, que podem ser exploradas por novos autores.

Uma nova chamada de contribuições foi então realizada pelo CGI.br, em parceria com o Nupef, com o objetivo de publicar os textos melhor avaliados, em um processo de revisão por pares, na revista poliTICs. Das mais de 60 contribuições, 14 artigos foram aprovados e selecionadas para publicação. Devido ao tamanho para publicação impressa, a edição especial 40 da poliTICs será publicada em dois volumes com o apoio do CGI.br. Os artigos restantes seguirão sendo analisados pela equipe editorial da poliTICs para futura publicação conforme o interesse das/os autoras/es.

Este primeiro volume da edição especial traz os cinco trabalhos abaixo relacionados³:

Design enganoso e manipulativo em plataformas digitais – um mapeamento entre estudos empíricos e reflexões para regulação -- George Augusto Valença Santos

O artigo analisa os impactos de padrões antiéticos de design através de uma abordagem qualitativa. Foram realizadas duas oficinas com pesquisadores e estudantes de ensino superior para mapear tipos de design manipulativo e enganoso em plataformas de software. Os achados nas oficinas foram associados às contribuições sobre design antiético identificadas na consulta do CGI.br. Os resultados destacam a urgência de regulações e iniciativas de letramento digital específicas, além de práticas de design que priorizem o bem-estar dos usuários.

Significados do discurso de ódio: desafios regulatórios e disputas discursivas na visão de diferentes setores sociais -- Eliza Bachega Casadei, Gabriela Agostinho Pereira

O texto analisa a complexidade das dinâmicas que envolvem a regulação de discursos de ódio em plataformas digitais. São exploradas as respostas de três setores sociais na consulta do CGI.br (Comunidade Científica e Tecnológica, Terceiro Setor e Setor Empresarial). A análise destaca as distinções entre os setores acerca do entendimento do discurso de ódio e, conseqüentemente, as soluções regulatórias, revelando as tensões que permeiam o campo da regulação digital e os desafios para estabelecer consensos em torno do enfrentamento do discurso de ódio nas plataformas digitais.

Regulação da desinformação nas plataformas digitais e o contexto eleitoral brasileiro -- Giovanna Milanese Tavares, Leandro Léo Rebelo, Marina Giovanetti Lili Lucena.

O texto investiga a regulação das plataformas digitais no contexto eleitoral tendo em conta a sistematização da consulta do CGI.br. A partir de uma análise conceitual, busca-se compreender se as medidas de mitigação propostas pela consulta no contexto das eleições são inovadoras ou se já são previstas na legislação eleitoral brasileira.

¹<https://dialogos.cgi.br/documentos/debate/consulta-plataformas/>

²<https://cgi.br/publicacao/sistematizacao-das-contribuicoes-a-consulta-sobre-regulacao-de-plataformas-digitais/>

³No volume 01, foram selecionados os artigos que receberam, ao menos, um parecer com aprovação sem solicitação de modificações. Os demais artigos selecionados receberam dois pareceres recomendando aprovação, mas solicitando modificações. Trata-se de critérios adotados para fins de organização, os artigos não foram pontuados.

Atenção, dados e poder: uma análise crítica da infodemia e da desinformação -- Leandro Modolo, Leonardo Castro, Luis Gonçalves

A partir da leitura crítica da sistematização da consulta pública do CGI.br, este artigo explora os conceitos de infodemia, desinformação e noções associadas, tais como os sistemas de crenças, para reenquadrá-los em um contexto sócio-histórico mais amplo, em diálogo com os conceitos de economia da atenção e capitalismo de plataformas. Argumenta que a infodemia e a desinformação são aspectos determinantes da economia digital contemporânea, e não simplesmente falhas de mercado. O texto explora como a infodemia e a desinformação se entrelaçam com o desenvolvimento de inteligência artificial e a produção massiva de dados digitais no contexto do capitalismo de plataformas.

A questão negra na regulação das plataformas digitais: uma análise da consulta realizada pelo CGI.br -- Fernanda dos Santos Rodrigues Silva, Marco Antônio Alves, Mariana Alves Araújo Lopes

O texto analisa como a questão racial apareceu e foi abordada no debate da consulta. Partiu-se do pressuposto de que o marco regulatório brasileiro sobre plataformas digitais precisa ser explicitamente antirracista para impedir a reprodução do racismo estrutural. As principais conclusões indicam que houve avanço na interseção entre regulação de plataforma e o debate racial, mas ainda há dificuldades. O consulente, ao formular as questões, foi mais genérico, sem focar diretamente o racismo. Nas contribuições, sugeriu-se, dentre outros, uma moderação de conteúdo mais inclusiva, auditorias algorítmicas, relatórios sobre discriminação racial e a criação de uma autoridade regulatória com composição diversificada, incluindo representação racial. Apesar disso, o total de menções à questão racial foi baixo em relação ao número total de contribuições.

Design enganoso e manipulativo em plataformas digitais – um mapeamento entre estudos empíricos e reflexões para regulação

George Augusto Valença Santos

Resumo

A ascensão das plataformas digitais transformou as interações entre usuários e tecnologias, trazendo desafios éticos, como a disseminação de padrões de design enganosos e manipulativos. Essas práticas exploram vieses cognitivos e vulnerabilidades psicológicas, promovendo comportamentos viciantes, afetando a autonomia e comprometendo o bem-estar dos usuários – especialmente crianças e adolescentes. Apesar do avanço de taxonomias e estudos sobre o tema, ainda há lacunas em termos de regulamentação e diretrizes para enfrentar esse problema. Neste artigo, investigamos os impactos de padrões antiéticos de design por uma abordagem qualitativa. Realizamos duas oficinas com pesquisadores e estudantes de Ensino Superior para mapear tipos de design manipulativo e enganoso em plataformas de software. Em seguida, associamos esses achados às contribuições sobre design antiético identificadas na consulta sobre regulação de plataformas digitais do Comitê Gestor da Internet no Brasil (CGI.br). Os resultados destacam a urgência de regulações e iniciativas de letramento digital específicas, além de práticas de design que priorizem o bem-estar dos usuários. As reflexões finais apresentam um modelo que nos descreve esse cenário e nos permite traçar estratégias para promover uma reflexão crítica sobre escolhas nas plataformas, abordando também o conceito de padrões justos ou iluminadores de design. Acreditamos que este estudo contribui para discussões sobre um design responsável, oferecendo subsídios para responsabilização de empresas por práticas antiéticas e propondo diretrizes para tecnologias mais inclusivas.

Palavras-chave: *Dark Pattern*. Design Manipulativo. Plataformas. Regulação.

Introdução

Embora a ascensão das plataformas digitais, especialmente as de mídias sociais, tenha inaugurado uma nova fase na interação entre usuários e tecnologias, com outras formas de consumo de informações, a revolução digital traz consigo inúmeros desafios éticos e sociais. Entre as questões que mais afetam usuários de todas as faixas etárias e em diversos tipos de plataforma (de jogos a e-commerce) está a crescente adoção de padrões de design enganosos e manipulativos – o que era denominado como *dark patterns*, nomenclatura a ser evitada diante do caráter racializado (Jafari; Vassileva, 2024; Trzaskowski, 2024; Mildner, 2024). São inúmeros os tipos desses padrões, geralmente usados de forma combinada. Por exemplo, *Confirmshaming* (confirmar por vergonha) usa linguagem emocional e constrange usuários para que tomem decisões específicas, *Urgência Falsa* usa temporizadores para criar uma urgência artificial, enquanto *Rolagem Infinita* prolonga o tempo de uso (AEPD, 2024).

Diversos estudos demonstram a presença ubíqua desses padrões antiéticos de design não só nas interfaces digitais, mas nas funcionalidades em si de plataformas de mídia social, em particular (Chang; Seaborn; Adams, 2024; Leiser; Santos, 2023). De forma sutil, esse design explora vieses cognitivos e vulnerabilidades psicológicas de usuários para influenciar a tomada de decisões e o fluxo de informações oferecido à plataforma (Jafari; Vassileva, 2024; Trzaskowski, 2024). Com isso, tornam-se estratégias de design que põem em risco o bem-estar dos usuários, afetando a saúde mental (Mildner, 2024).

Crianças e adolescentes, em particular, são considerados mais vulneráveis a essas práticas de manipulação devido ao baixo letramento digital e à menor capacidade de identificar e resistir a táticas enganosas. Por exemplo, a TIC Kids Online 2024, pesquisa regular

conduzida pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br) vinculado ao Núcleo de Informação e Coordenação do Ponto BR (NIC.br), indicou apenas que 46% de crianças e adolescentes sabe ajustar as configurações de privacidade nas redes sociais¹, algo intencionalmente dificultado por padrões, como *Interferência Visual* e *Obstrução*, que ressaltam determinadas funcionalidades de interesse da plataforma e introduzem fricção (aumento do número de etapas para realização da ação) em outras, confundindo usuários ou limitando a descoberta de possibilidades relevantes, como ajustes para sua segurança – por exemplo, criar avisos de tempo máximo num aplicativo (Mildner, 2024).

A exposição contínua de usuários a esses padrões promove comportamentos viciantes, catalisados por notificações e números (como curtidas, encaminhamentos, novos seguidores), entre outros, que demonstram suposto progresso virtual e estimulam a economia da atenção. Assim, tem-se um design antiético que molda comportamentos, aumenta o tempo gasto, reduz a capacidade de autorregulação e leva à erosão da confiança nas plataformas – que buscam se beneficiar, em vez de primar pelo melhor interesse dos usuários (Obi *et al.*, 2022; Chamorro; Lallemand; Gray, 2024; Jafari; Vassileva, 2024).

Neste artigo, avaliamos os impactos negativos do design antiético bem como a regulação e a promoção do design responsável de plataformas digitais. Para isso, conduzimos dois estudos de campo sobre o tema com pesquisadores e estudantes de Ensino Superior. Neles, os participantes mapearam padrões de design enganoso e manipulativo em mídias sociais, além de refletir sobre alternativas justas de design. Em seguida, combinamos os achados empíricos com reflexões feitas por diversos setores da sociedade brasileira na consulta do CGI.br sobre regulação de plataformas digitais no Brasil, adotando um processo de síntese temática dos dados. A partir dessa metodologia diversa, associando visões do campo às de inúmeros setores presentes na consulta, buscamos apoiar a elaboração de diretrizes para tratamento do design antiético.

Conceitos Básicos

Padrões enganosos e manipulativos de design derivam do termo em inglês *deceptive* ou *manipulative patterns*. São estratégias de design inseridas em soluções de software a partir do conhecimento profundo de empresas desenvolvedoras sobre o comportamento humano, como desejos dos usuários. Assim, é possível construir funcionalidades enganosas que não são do interesse do público e/ou que manipulam suas decisões². Dezenas desses padrões foram mapeados na literatura de Ciência da Computação e Interação Humano-Computador. Aqui, trazemos alguns a partir da ontologia de Harry Brignull e equipe em seu *site* (Brignull; Leiser, 2023), que lista padrões considerados comuns ou frequentes desde o surgimento do conceito em 2010. Apresentamos essa lista, com descrições (Figura 1), cuja abordagem, com exemplos, segue nos parágrafos seguintes.

Figura 1 –. Lista e descrição dos principais padrões de design enganoso ou manipulativo.

¹ TIC Kids Online Brasil 2024 - <https://cetic.br/pt/pesquisa/kids-online/>

² Antes, tais padrões eram denominados *dark patterns* ou padrões obscuros de design, terminologia que tem sido evitada nos últimos anos por seu cunho racializado.

Padrões de Design Enganoso - descrição

Prevenção de comparação

O usuário tem dificuldade para comparar produtos porque características e preços são combinados de maneira complexa ou porque é difícil encontrar informações essenciais.

Confirmshaming

O usuário é emocionalmente manipulado para fazer algo que de outra forma não teria feito.

Anúncios disfarçados

O usuário acredita erroneamente que está clicando em um elemento da interface ou conteúdo nativo, mas na verdade é uma propaganda disfarçada.

Falsa escassez

O usuário é pressionado a concluir uma ação porque recebe uma indicação falsa de oferta ou popularidade limitada.

Prova social falsa

O usuário é induzido a acreditar que um produto é mais popular ou confiável do que realmente é, porque foram mostrados comentários, depoimentos ou mensagens de atividades falsas.

Urgência falsa

O usuário é pressionado a concluir uma ação porque é apresentada uma falsa limitação de tempo.

Ação forçada

O usuário deseja fazer algo, mas é obrigado a fazer algo indesejável em troca.

Redação complicada

O usuário é induzido a realizar uma ação em erro devido à apresentação de linguagem confusa ou enganosa.

Dificuldade para cancelar

O usuário acha fácil se cadastrar ou assinar, mas quando quer cancelar tem muita dificuldade.

Despesas ocultas

O usuário é seduzido por um baixo preço anunciada. Depois de investir tempo e esforço, eles descobrem taxas e cobranças inesperadas quando chegam ao caixa.

Assinatura oculta

O usuário está inadvertidamente inscrito em uma assinatura recorrente ou plano de pagamento sem divulgação clara ou seu consentimento explícito.

Interrupção (Nagging)

O usuário tenta fazer algo, mas é persistentemente interrompido por solicitações para fazer outra coisa que pode não ser do seu interesse.

Obstrução

O usuário se depara com barreiras ou obstáculos que dificultam a conclusão de suas tarefas ou o acesso às informações.

Pré-seleção

É apresentada ao usuário uma opção padrão que já foi selecionada para ele, a fim de influenciar sua tomada de decisão.

Inclusão sorrateira (Sneaking)

O usuário é atraído para uma transação sob falsos pretextos, porque informações pertinentes são ocultadas ou demoram a ser apresentadas a ele.

Interferência visual

O usuário espera ver informações apresentadas de forma clara e previsível na página, mas elas estão ocultas, obscurecidas ou disfarçadas.

Fonte: Adaptado de Brignull; Leiser (2023).

Interferência Visual pode ser um conceito considerado clássico em termos de más práticas de design de software. Ele cria a chamada “fricção” na interface para aumentar a dificuldade para realizar uma tarefa, com aumento do número de cliques ou de telas a acessar, escondendo informações deliberadamente. Como exemplos, há textos pequenos e de baixo contraste, interfaces complexas ou dados relevantes (por exemplo, privacidade) ocultados em menus inesperados pelo usuário. Plataformas como Instagram e Facebook têm seções ou abas, como “dados pessoais”, “senha e segurança” e “informações e permissões”, sem formas de gerenciar dados pessoais; assim, a Meta evita que bilhões de usuários acessem tais recursos com facilidade, dado o benefício de ampla coleta e processamento de dados externos para seu negócio. Já Pré-seleção e Obstrução buscam, respectivamente, influenciar usuários com escolhas pré-selecionadas ou fazer com que enfrentem barreiras para terminar ações.

Para trazer mecanismos adicionais e ainda mais recentes, como forma de estimular a busca por outras ontologias na literatura, introduzimos o padrão Captura de Atenção (*Attention Capture*). Ele é um pilar de redes sociais, assim como um subtipo ou derivação do padrão Investimento Social (*Social Investment*), que retém usuários a partir de recompensas constantes e se configura como o pilar de mídias sociais. Recursos como visualizações, reações, comentários e amigos/seguidores concretizam esse padrão e vinculam usuários à plataforma, incutindo a ideia de que apenas o uso constante de plataformas de mídia social (Instagram, Facebook, Youtube ou TikTok, por exemplo) evita a perda de “progressos” (Roffarello; De Russis, 2022). Desse modo, eles contribuem para que essas plataformas pareçam indispensáveis à rotina de usuários e usuárias (Número [...], 2022). Esse círculo vicioso também é mantido pelos padrões Reprodução Automática (*Autoplay*) e Puxar para Atualizar (*Pull-to-refresh*): o primeiro ativa o recurso “próximo vídeo”, mantendo usuários assistindo a conteúdos na plataforma, com vídeos reproduzidos automaticamente, ao passo que o segundo consiste na rolagem infinita (quando usuários não visualizam o fim de uma página), que carrega conteúdo na parte inferior continuamente (Roffarello; De Russis, 2022).

O design antiético afeta particularmente crianças via técnicas de engajamento e exploração de vulnerabilidades psicológicas para maximizar o tempo gasto e a produção de conteúdo,

com visitas e interações contra a sua vontade (Lukoff, 2021). Com o uso de técnicas para criar recompensa virtuais, elas experimentam variações nos níveis de dopamina de forma semelhante à observada em pessoas com dependência de jogos de azar ou máquinas caça-níqueis (Roffarello; De Russis, 2022). Por exemplo, o padrão Investimento Social compromete a autonomia e afeta o bem-estar de *kid influencers*, alterando sua percepção de controle racional sobre o tempo gasto na plataforma, o tipo de conteúdo que criam e a frequência de produção. Nesse sentido, essas são formas de induzir os usuários ao engano, mantendo-os presos à rede social e direcionando decisões.

Diante do aumento de tempo na plataforma e conseqüente crescimento no total de dados oferecidos sobre a experiência digital, a empresa consegue traçar o perfil comportamental do público. Essa telemetria permite modelar funcionalidades com base em emoções, valores, traços de personalidade e opiniões, explorando a compreensão do estado psicológico de usuários (Crepax; Mühlberg, 2022). Logo, usuários manipulados seguem nutrindo a plataforma pelo medo de perder recompensas ou ficar para trás, numa típica situação de *Fear of Missing Out* (FoMO) (medo de não conseguir acompanhar atualizações e eventos), com receio de ficarem de fora da tecnologia ou se desenvolverem em outro ritmo (Sousa; Oliveira, 2023).

Metodologia

Este artigo busca responder a seguinte questão de pesquisa: como podemos desenvolver e implementar medidas eficazes que englobem educação, regulamentação e design ético para proteger os usuários dos impactos negativos do design enganoso e manipulativo em plataformas de mídia social? A fim de construirmos um entendimento abrangente do problema e delinear as soluções eficazes que promovam uma experiência digital transparente e segura, associamos achados de um estudo de campo no tema com dados da consulta realizada pelo CGI.br sobre regulação de plataformas digitais.

Durante uma conferência, foi realizado um minicurso de três horas, estruturado com base nas metodologias de “Aprendizagem Baseada em Problemas” e no conceito de “Design para Não-Designers” (Ibarra, 2016), cujo foco foi explorar o tema design manipulativo numa abordagem teórica e prática. Esse minicurso foi organizado em três etapas principais.

Inicialmente, os participantes foram introduzidos aos conceitos, com tipos e casos reais (exemplos concretos de práticas enganosas de plataformas digitais). Posteriormente, cerca de 15 participantes de diferentes contextos acadêmicos e profissionais (como estudantes, professores e profissionais em exercício) foram divididos em subgrupos de cinco pessoas, seguindo critérios de afinidade. Cada grupo analisou estratégias manipulativas empregadas por empresas de tecnologia em mídias sociais, identificando padrões enganosos conforme a taxonomia disponível no site *Deceptive Patterns* (Brignull; Leiser, 2023). Como ferramenta de análise, foi adotado um formulário criado no Google Forms (Figura 2), em que os grupos registraram os padrões observados nas mídias sociais de sua escolha e sugeriram soluções baseadas em suas experiências, como diretrizes para o desenvolvimento de funcionalidades mais éticas e responsáveis.

Figura 2 – . Formulário para avaliação de padrões antiéticos de design e esboço de soluções.



* Indica uma pergunta obrigatória

Vocês poderiam se identificar? *

Sua resposta

Que tal dar um nome para a sua equipe?



Sua resposta

Quais padrões você conseguiu identificar? *

Está com dúvida sobre o que cada um significa? Basta acessar <https://www.deceptive.design/types>

- Prevenção de comparação
- Confirmshaming
- Anúncios disfarçados
- Falsa escassez
- Prova social falsa
- Urgência falsa
- Ação forçada
- Redação complicada
- Dificuldade para cancelar
- Despesas ocultas
- Assinatura oculta
- Nagging
- Obstrução
- Pré-seleção
- Sneaking
- Interferência visual

Como estes padrões foram implementados? *

Ex.: "identificamos o padrão x na funcionalidade y do Instagram com o uso de...".

Sua resposta

Quais boas práticas ou recomendações você daria para tratar/evitar estes padrões? *

Ex.: "este padrão poderia ser evitado a partir da construção de uma interface com...". "este outro padrão poderia ser contornado com a reflexão pela equipe sobre...".

Sua resposta

Fonte: Elaboração própria.

Na terceira e última etapa, cada grupo apresentou suas análises, destacando os pontos mais relevantes e promovendo uma discussão coletiva. Essa troca de ideias incentivou reflexões críticas sobre impactos do design manipulativo e desenvolvimento de abordagens éticas. Todas as falas desse segundo momento foram registradas, com consentimento das equipes, e posteriormente transcritas e classificadas por dois pesquisadores a partir de um processo de Síntese Temática (Cruzes; Dybå, 2011), técnica que busca analisar dados qualitativos via codificação de textos, criação de temas descritivos (grupo de códigos similares) e geração de descrições dos temas para construir explicações ou derivar hipóteses. Assim, tem-se uma abordagem indutiva para mapear achados em um conjunto dos estudos via softwares, como Microsoft Word.

Esse mesmo processo foi adotado numa oficina do período 2023.2 de curso de Ciência da Computação, com um total de 30 estudantes, cujo processo foi realizado em seis equipes. Assim, ao todo, tivemos 45 participantes no estudo sobre design antiético de plataformas.

Em paralelo, em 2023, o CGI.br realizou uma consulta que recebeu um total de 1.336 contribuições de pessoas e organizações do setor governamental, do terceiro setor e do setor empresarial, bem como da comunidade científica e tecnológica. O resultado desse levantamento foi publicado como um relatório, apresentado de forma acessível e sintética em diversos formatos (*site*, PDF e CSV). Analisamos a ocorrência do conceito de design enganoso e manipulativo nestas extensas contribuições; para isso, empregamos os seguintes passos. Inicialmente, buscamos os termos *deceptive*, *dark pattern* e *nudge* bem

como suas versões e variações em português: *enganoso*, *manipulativo*, *manipulação* e *obscuros* (expressões mais amplamente adotadas na academia para discutir essa problemática), segundo relatório liberado pelo CGI.br.

Ao todo, foram extraídos 21 trechos em que esses termos são abordados, todos selecionados e migrados para um arquivo de texto, em que foram classificados em temas. Com o processo de rotulagem dos dados da consulta, identificamos cinco temas: Manipulação Comportamental, Legislação e Boas Práticas de Design, Privacidade e Proteção de Dados, Publicidade e Letramento Digital. Utilizamos esses eixos para extrair insumos correspondentes na pesquisa de campo realizada (por exemplo, minicurso e oficina). Dessa forma, desenvolvemos uma tabela³ combinando os dois conjuntos de dados, posteriormente utilizada para síntese dos dados – explicação da percepção da academia e dos participantes da consulta sobre regulação de plataformas.

Neste artigo, exploramos o conceito de design antiético a partir da combinação de estudos em campo com resultados de consultas promovidas pelo NIC.br: uma metodologia rica por alinhar a visão de profissionais que atuam na área de design e Ciência da Computação com grupos que representam diversos setores da sociedade na busca por uma Internet mais justa e segura. Ademais, temos a incorporação da literatura neste texto, o que confere a nossos resultados (conforme seção a seguir) uma triangulação e um rigor necessários para avanços no tema. Desses elementos, foram derivadas recomendações e boas práticas ligadas aos princípios de Design Justo (*Fair / Bright Patterns*), propostas na discussão final deste artigo. A principal intenção dessa abordagem é fomentar o bem-estar e a autonomia dos usuários, em oposição à lógica manipulativa que predomina em muitas plataformas digitais.

Resultados

Os estudos de campo realizados trouxeram uma percepção sobre quais padrões enganosos e manipulativos os participantes percebiam nas plataformas de mídia social. As Figuras 3 e 4 trazem as opções identificadas pelos grupos. Como cada grupo preencheu o resultado uma vez, cinco respostas em uma figura simbolizam a análise de cinco equipes formadas por cinco pessoas cada. Além disso, apesar da liberdade de escolha (em termos de mídias sociais), os times concentraram-se na avaliação do Instagram, TikTok e Youtube (diante de funcionalidades similares às demais plataformas, como “seguir perfis”, “postar comentários”, “postar conteúdo” e “reagir”).

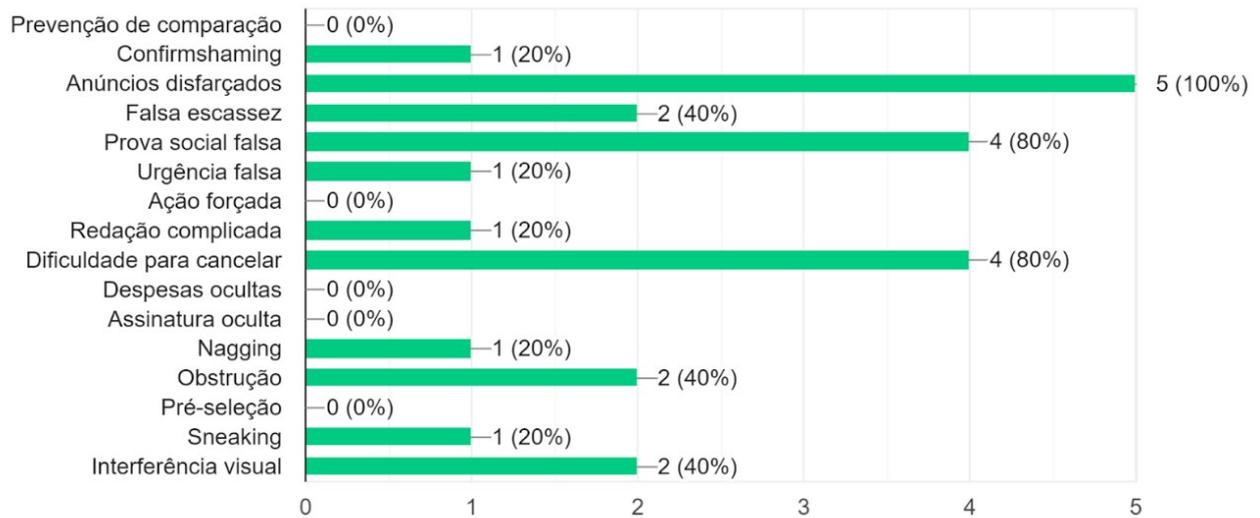
Nos dois estudos, houve convergência quanto aos tipos de design enganoso e manipulativo mais frequentemente encontrados em mídias sociais, como Instagram e TikTok: *Anúncios Disfarçados*, *Prova Social Falsa* e *Dificuldade para Cancelar*. Todas as equipes perceberam a publicidade velada, geralmente acompanhada por depoimentos inverídicos (em vídeo ou em artes geradas, inclusive, por inteligência artificial [IA]). Além disso, reconheceram que o cancelamento do serviço em si, como o encerramento de uma conta, é forçosamente custoso em termos de número de passos, identificação de menus ou opções na interface. Algumas divergências foram percebidas no levantamento feito com estudantes: padrões como *Confirmshaming*, *Redação Complicada*, *Nagging*, *Obstrução* e *Pré-seleção* surgiram com maior frequência nas respostas desse grupo.

Figura 3 – . Presença dos padrões antiéticos de design na visão dos participantes da conferência.

³ A tabela com os dados classificados está disponível em: <https://tinyurl.com/2vkb4vj>. Acesso em: 5 maio 2025.

Quais padrões você conseguiu identificar?

5 respostas

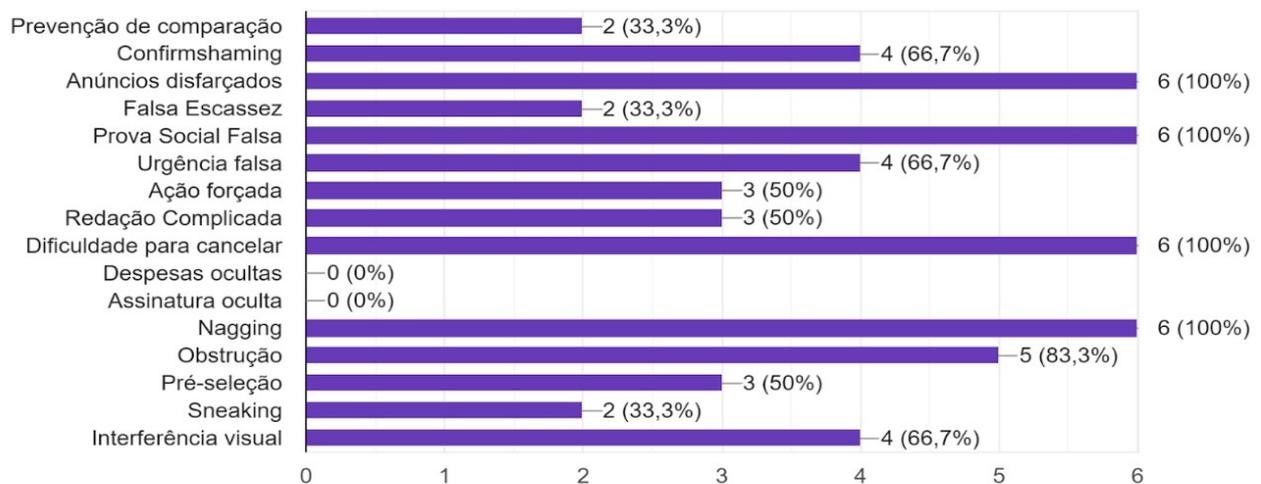


Fonte: Elaboração própria.

Figura 4 – . Presença dos padrões antiéticos de design na visão de estudantes universitários.

Quais padrões você conseguiu identificar? Está com dúvida sobre o que cada um significa? Basta acessar: <https://www.deceptive.design/types>

6 respostas



Fonte: Elaboração própria.

Como esses padrões são, em geral, associados ou mais percebidos em funcionalidades de suporte ou de configuração do perfil nas plataformas, como ajustes de privacidade, entendimento de termos de uso ou desativação da conta, uma hipótese seria que esses participantes buscam, com maior frequência, acessar essas funcionalidades para estabelecer um ambiente mais seguro online, o que gera maior conhecimento sobre as barreiras intencionais para utilizá-los. Para além de verificações mais quantitativas como

frequência de uso ou adoção pelas pessoas participantes, foi possível levantar percepções mais subjetivas. As transcrições e as classificações dos dados dos estudos de campo são apresentadas à luz dos cinco temas mapeados nas contribuições feitas à consulta do NIC.br no que diz respeito a design manipulativo. Assim, nas seções a seguir, os dados do mapeamento empírico são entrelaçados com os da consulta sobre regulação de plataformas.

Manipulação Comportamental

As contribuições oferecidas à consulta do CGI.br reforçam que o design persuasivo, com técnicas de *nudge*⁴, é um meio para manipulação comportamental: uma estratégia adotada por grandes empresas de tecnologia com elevado poder de mercado para manter a lógica dos seus modelos de negócio. Ao desenvolverem plataformas de software que expõem usuários a conteúdos que manipulem o seu comportamento e garantam engajamento, elas aumentam o tempo de tela e sustentam a geração de valor, com alta rentabilidade para seus negócios.

Também foi destacado que a manipulação é possível graças ao uso de dados (de navegação, pessoais etc.) para direcionamento de conteúdos e *profiling*. Com isso, diversos grupos, como crianças e adolescentes, tornam-se alvo de exploração comercial, ou seja, técnicas de design manipulativo ou enganoso conduzem o subconsciente desses e de outros grupos de usuários a se comportar de determinada forma (por exemplo, para que adquiram um dado produto). Como crianças são mais suscetíveis a essas técnicas, a manipulação via design impacta negativamente o desenvolvimento, com possíveis perdas de oportunidades e estímulos a um crescimento livre e pleno desses indivíduos.

Já na visão dos participantes dos estudos de campo, a origem dos padrões antiéticos de design também está no modelo de negócio das plataformas de mídia social. Foi reforçado que esses ambientes são baseados na monetização de conteúdo, isto é, um *influencer* ou usuário regular só pode promover seguindo regras da plataforma, por exemplo. Portanto, para que seu conteúdo seja impulsionado, eles respeitam métricas sociais, como curtidas e compartilhamentos, ou buscam mais seguidores. Essa relação de poder-dependência, por meio da qual as *Big Techs* usam seu respeito e conhecimento para estabelecerem regras em detrimento do bem-estar dos usuários, faz com que os usuários alimentem padrões enganosos devido à falta de consciência sobre a dinâmica da plataforma ou à necessidade de se beneficiar da plataforma financeira e socialmente. Uma das pessoas participantes indicou que “é um ciclo (em que o) nosso comportamento alimenta essas estruturas e essas empresas mudam o nosso comportamento. No nosso dia a dia, o Tiktok influencia a forma com que eu me relaciono com as pessoas [e isso] oferece dados para que o Tiktok continue influenciando”. Um exemplo preciso de um padrão manipulativo de design foi descrito por outra participante para ilustrar esse cenário: “[a]rolagem infinita sempre traz algo para você ver e te manter usando [a plataforma]”.

Por fim, percebemos empiricamente que usuários destacam um dos efeitos mais nefastos na mudança de comportamento: buscar retribuições, como exemplificado no seguinte relato:

[...] é muito real essa questão do processo de recompensa. Nosso cérebro é condicionado a buscar recompensas pela sensação de prazer. A gente está vendo a aplicação disso frequentemente no Tiktok com a implementação de múltiplos vídeos durante um *story*. Chegamos a um nível em que o cérebro está exigindo cada vez mais, em um processo viciante. Isso (es)tá sendo usado a todo momento, a cada notificação.

⁴ Conceito da economia comportamental que se refere a uma técnica de persuasão para influenciar as decisões das pessoas (Fischer, 2023).

Legislação e Boas Práticas de Design

As contribuições da consulta sobre regulação de plataformas também abordaram aspectos como regras e orientações ligadas ao design antiético. Do ponto de vista internacional, houve destaque para o direcionamento da Organização das Nações Unidas (ONU) no Comentário Geral n. 25 (UN, 2021), que incentiva os Estados-membros a introduzirem ou atualizarem suas regulações de proteção de dados, bem como seus padrões de design, para que identifiquem, definam e proíbam práticas que manipulem ou interfiram no direito das crianças à liberdade de pensamento e à crença no ambiente digital (por exemplo, evitar inferências ou análises emocionais). Assim, foi reforçada a importância da experiência de navegação de crianças e adolescentes, a qual deve restringir padrões enganosos e priorizar um design ético de interfaces e interação.

A recente Lei de Serviços Digitais (*Digital Services Act* – DSA) foi citada como mecanismo legal para atenuar a manipulação e a desinformação (European Commission, 2022). Em conformidade com essa legislação da União Europeia, plataformas devem diminuir a exposição de usuários a conteúdos ilegais e proteger direitos fundamentais. Além disso, devem promover maior transparência – ponto considerado um dos problemas mais importantes no contexto de plataformas, segundo os participantes da consulta.

Em nossos estudos de campo, também percebemos a demanda por definição de diretrizes gerais com as quais plataformas precisem garantir conformidade: “[é preciso] criar leis que regularizem essas empresas”. Essa regulamentação, segundo as equipes, poderia funcionar não somente para plataformas de mídias sociais, mas também “para a Internet como um todo”, por exemplo, esclarecer ainda mais as regras bem como os valores a serem obtidos por influenciadores logo no início da sessão, ao iniciarem a produção de conteúdo. Por fim, uma das participantes ilustrou a falta de diligência e suas consequências, quando plataformas não cumprem regras estabelecidas por elas próprias: “por mais que exista a regra universal, que não permite crianças, quem nunca teve um perfil com 12 anos? O acesso muito prematuro da criança [às plataformas] afeta a questão social e de autoconfiança”.

Privacidade e Proteção de Dados

Se o design enganoso e manipulativo floresce a partir da coleta massiva de dados, este também deve ser o foco das preocupações no contexto de regulação e design ético. As contribuições ao CGI.br reforçaram posicionamentos de instituições, como Instituto Alana e Data Privacy Brasil, que indicam que o tratamento de dados de crianças e adolescentes permite expor, armazenar e utilizar seus extensos e preciosos rastros digitais no presente e no futuro. Logo, o uso indiscriminado de dados ligados aos usuários (facilitado por padrões antiéticos de design, como *Interferência Visual*, *Pré-seleção* e *Obstrução*) para segmentação de anúncios é uma violação de privacidade.

Ainda segundo os resultados da consulta, *Big Techs* como Meta (responsável pelo Whatsapp, Facebook e Instagram), Alphabet (responsável pelo Youtube e Youtube Kids) e Bytedance (responsável pelo TikTok) desenvolvem funcionalidades persuasivas a partir de dados multidimensionais, que vão muito além de categorias demográficas e pessoais. A esse universo estão associadas também plataformas de parceiros. Por exemplo, se ampliarmos o cenário, abarcamos o contexto educacional, em que temos ferramentas ou extensões terceiras associadas ao Google Classroom, que realizam perfilação de dados de estudantes e professores.

Em campo, coletamos variados exemplos alinhados à perspectiva indicada: coleta de dados para fins de manipulação. O padrão antiético de design *Fofura* (*Cuteness*), que cria interfaces excessivamente lúdicas ou aparentemente inofensivas para gerar uma resposta afetiva no usuário com o propósito de coletar dados (fenômeno conhecido por “miopia de dados”) (Lacey; Caudwell, 2019), foi percebido por uma participante do estudo de campo conduzido em uma conferência: “essa fofura [permite à plataforma] saber qual tipo de

planta é mais adequado para aquele espaço, por exemplo, e (ela) conseguiria vender mais para aquele público”. Ou seja, o vínculo afetivo com uma determinada plataforma mais lúdica, como o Youtube Kids ou mesmo um robô social, reduz o grau de agência de crianças. De forma equivalente, para entender interesses, o Instagram exhibe postagens de perfis que não são seguidos pelo usuário e pergunta se é de seu interesse, uma maneira sutil de extrair mais dados de uso e preferência via interface manipulativa apontada pelos participantes do estudo.

Em casos mais específicos, identificamos nas falas os padrões *Interferência Visual e Obstrução*. No primeiro caso, no contexto de gerenciamento de *cookies*, um participante indicou: “você está em uma página e ela não te deixa seguir de jeito nenhum; você não tem ali a opção [de rejeitar]; aceita ou sai da página; você acaba forçadamente aceitando o *cookie*”. Nesse caso, uma informação necessária da página é omitida para facilitar a coleta de dados. Por fim, uma participante com deficiência visual fez o relato a seguir, demonstrando que o número excessivo de etapas para gestão dos dados numa plataforma é, acima de tudo, símbolo de um design não-inclusivo: “primeiro, fui nas configurações; digitei ‘dados’, mas não encontrei. Então, fui para outra opção... Dei uma volta com o leitor de tela tentando descobrir”.

Publicidade

Em inúmeros pontos, a consulta sobre regulação de plataformas trouxe contribuições que cruzavam os contextos do design manipulativo e publicidade direcionada. As propostas buscam vedação estrita da utilização de *nudges* e *dark patterns* pelas plataformas para fins de publicidade comercial para crianças diante dos riscos trazidos. Esses anúncios foram apontados como resultado de processos de perfilamento, sendo responsáveis pelo aumento da captura de dados pessoais e manipulação desses usuários.

A partir de padrões enganosos, como *Anúncios Disfarçados*, o usuário acredita que está clicando em um elemento da interface ou de conteúdo nativo, mas, na verdade, está consumindo um anúncio velado. Isso foi apontado em uma contribuição a partir da menção de propagandas com design similar ao conteúdo serem um padrão antiético de design que não explicita as consequências de ações dos usuários. Em outro ponto, esse tipo de design foi definido como “publicidade opaca ou enganosa”, uma estratégia altamente persuasiva semelhante às de jogos de azar. Como soluções, foi indicado i) o impulsionamento de conteúdos que busquem orientar proteger a privacidade dos usuários, informando-os sobre como evitar práticas manipulativas; e ii) uma maior transparência na publicidade digital para evitar práticas fraudulentas.

Os participantes dos estudos de campo também identificaram uma presença recorrente de padrões de anúncios disfarçados em postagens e *stories* que não podem ser imediatamente identificados como conteúdo patrocinado. Além disso, há conteúdos promocionais em plataformas, como o Instagram, que não trazem uma *tag* de anúncio. Um participante elaborou: “em um *story*, uma jovem fala sobre um produto para o rosto, maquiagem... Muitas vezes, você não percebe que é um anúncio; [mas] é pago”.

Também verificamos nos dados duas estratégias empregadas para ocultar anúncios: i) utilizar influenciadores para promover produtos ou serviços, e ii) integrar perfeitamente esse conteúdo ao *feed*, imitando a aparência de postagens regulares. Segundo os participantes, esses anúncios disfarçados afetam a navegação, criando interrupções a cada dois ou três *stories* no Instagram. Um participante mencionou que, enquanto visualiza *stories* e postagens de amigos, de repente era apresentado a anúncios, levando a uma “confusão na navegação do perfil e dos stories”.

Letramento digital

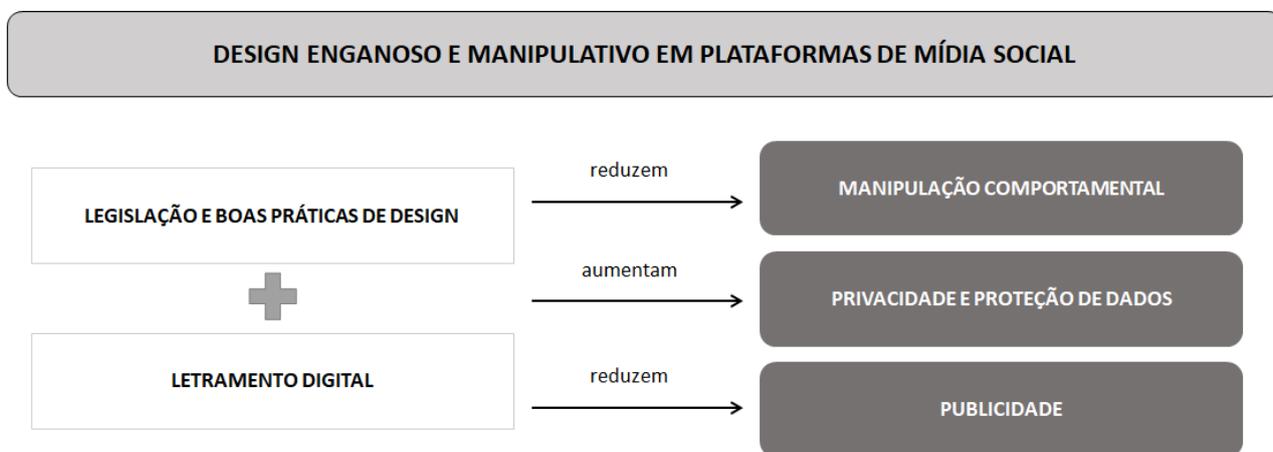
Uma contribuição feita à consulta sobre regulação de plataformas apontou ser preciso promover a alfabetização midiática e digital, ou seja, estruturar programas de educação e

conscientização para capacitar usuários a identificarem conteúdos desinformativos e extremismos, bem como discursos de ódio, por exemplo. Dessa forma, usuários seriam capazes de avaliar e discernir sobre a confiabilidade dessas informações (algo que envolve os padrões antiéticos *Prova Social Falsa* e *Urgência Falsa*), a fim de compreender estratégias de manipulação e verificar fontes e credibilidade do conteúdo geral. Em campo, observamos algo semelhante, com propostas de iniciativas para orientação do usuário, ou seja, esclarecer para o usuário quais informações serão usadas e como uma ferramenta funciona, um princípio de transparência fundamental no contexto de regulação. Isso evitaria escolhas definidas previamente na criação do perfil, como o padrão de manipulação *Pré-seleção*, que estabelece muitas vezes uma opção padrão – *default* – que não protege dados pessoais do usuário, por exemplo.

Discussão e Conclusão

Nesta seção final, destacamos convergências entre os estudos de campo e a consulta realizada pelo CGI.br no contexto de regulação de plataformas, de forma a extrairmos direcionamentos. As cinco dimensões de análise extraídas dos dados empíricos e adotadas para classificar as contribuições ligadas ao design enganoso e manipulativo são aqui inter-relacionadas (Figura 5). Derivamos um modelo que descreve esse cenário e nos permite refletir e traçar estratégias para promover melhores decisões e priorização dos objetivos para o bem-estar dos usuários. Assim, podemos ter um caminho para regulações mais eficazes, com elementos que podem ser pautados em projetos de lei sobre o tema.

Figura 5 – . Interrelação entre as dimensões de análise identificadas nos dados de campo e consulta.



Fonte: Elaboração própria.

As iniciativas de **legislação** e promoção de **boas práticas de design** no contexto de plataformas digitais são fundamentais para reduzir a **manipulação comportamental**. Instituições envolvidas com pesquisas, tanto na academia quanto no terceiro setor, precisam evitar que padrões de manipulação sejam vistos como algo amorfo ou subconceituado. Estudos sobre tais padrões devem se fundamentar em perspectivas normativas, como bem-estar individual, autonomia e tomada de decisão informada, a fim de se conferir maior apoio a responsáveis por formular políticas em ações regulatórias.

Como antítese dos padrões antiéticos, tem-se padrões justos de design, que se referem à funcionalidade persuasiva do design da interface do usuário que prioriza os objetivos dos usuários acima dos objetivos de plataformas digitais. A ideia é que as pessoas estejam no controle por meio de tecnologias projetadas em cima de padrões éticos que visam

promover transparência, acessibilidade e respeito pelas escolhas dos usuários, como comunicação clara de opções, disponibilização de controles de configuração de conta fáceis de usar e suporte à autonomia do usuário em decisões online (Potel-Saville; Rocha, 2023). Dessa forma, promove-se privacidade e proteção de dados, uma vez que esses padrões éticos complementam a imprecisão de leis de privacidade.

Além disso, descrições do porquê de padrões enganosos e outros aspectos do design serem preocupantes e produzirem efeitos negativos sobre indivíduos (como promoção de publicidade e valores como consumismo) trazem uma visão mais rica do que críticas subjetivas ao design de soluções. As interpretações sobre práticas injustas de design precisam ser feitas por pesquisas transdisciplinares, como pesquisas em interface humano-computador combinadas a relatórios de órgãos reguladores para apoiar novas legislações de proteção aos usuários e sanções legais contra empresas de tecnologia que afetam a autonomia de crianças via design antiético.

Por outro lado, iniciativas como debates, publicações ou oficinas abordando o design antiético e injusto (tipos, impactos etc.) garantem letramento digital. Em combinação com a regulação, esse tipo de esforço por diferentes instituições (por exemplo, painéis, como os do “Fórum da Internet do Brasil”, realizado anualmente pelo NIC.br, ou o “Seminário Educação, Governança de Dados e Inteligência Artificial”, conduzido pelo MEC, em 2024, em torno de temas como privacidade e uso responsável de dados) podem informar adultos e crianças sobre como plataformas podem enganar e coagi-los a ter atitudes que, de outra forma, não teriam, expondo-os a riscos. Assim, aceleramos a mudança de abordagem: caracterizar o design antiético para aplicar resultados de pesquisa no mundo real, influenciando práticas e regulações – como o *Digital Services Act* da União Europeia, que estabelece proibições legais contra o design enganoso ou o recente desenvolvimento de diretrizes éticas por órgãos como o European Data Protection Board. Por meio da fiscalização, da possibilidade de sanções e da discussão sobre padrões justos de design, ocorrerá a dissuasão de plataformas e conscientização de profissionais, como designers e programadores, que poderão evitar manipulação e criar interfaces e funcionalidades mais transparentes e centradas, de fato, no usuário.

Uma das limitações desta pesquisa foi contar com uma visão restrita do campo: a partir da percepção de um grupo predominantemente acadêmico, foi possível obter relatos sobre design antiético em plataformas de mídia social. Além disso, há um julgamento subjetivo pelos participantes desses estudos, já que os padrões estudados não têm métricas ou características específicas para identificação e análise. No entanto, contornamos essa limitação com uma iniciativa de letramento com os grupos, explicando cada tipo da taxonomia de Harry Brignull. Em estudos futuros, pretendemos ampliar o levantamento de dados, envolvendo profissionais da indústria e instituições do terceiro setor, presentes no levantamento realizado pelo CGI.br.

Referências

AEPD – Agencia Española de Protección de Datos. Patrones adictivos en el tratamiento de datos personales: Implicaciones para la protección de datos, 2024.

CHANG, Weichen Joe; SEABORN, Katie; ADAMS, Andrew A. Theorizing Deception: A Scoping Review of Theory in Research on Dark Patterns and Deceptive Design. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. 2024. p. 1-7.

CREPAX, Tommaso; MÜHLBERG, Jan Tobias. Upgrading the protection of children from manipulative and addictive strategies in online games: Legal and technical solutions beyond privacy regulation. arXiv preprint arXiv:2207.09928, 2022.

CRUZES, Daniela S.; DYBA, Tore. Recommended steps for thematic synthesis in software engineering. In: 2011 International Symposium on Empirical Software Engineering and Measurement. p. 275-284.

FISCHER, Márcio Santos. Economia comportamental e marketing: uma revisão introdutória. 2023.

IBARRA, Maria Cristina. (2016) "O design e suas possíveis interações com práticas criativas desenvolvidas por não-designers." In: Revista ARCOS design., p. 165. UERJ.

JAFARI, Elaheh; VASSILEVA, Julita. Designing Effective Warnings for Manipulative Designs in Mobile Applications. In: Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization. 2024. p. 12-17.

LACEY, Cherie; CAUDWELL, Catherine. Cuteness as a 'dark pattern' in home robots. In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2019. p. 374-381.

LEISER, Mark; SANTOS, Cristiana. Dark Patterns, Enforcement, and the emerging Digital Design Acquis: Manipulation beneath the Interface. 2023.

MILDNER, Thomas. Mitigating dark patterns through responsible design-ethical design considerations for user-centred technologies. 2024. Tese de Doutorado. Universität Bremen.

MONGE ROFFARELLO, Alberto; DE RUSSIS, Luigi. (2022) "Towards understanding the dark patterns that steal our attention". In: CHI Conference on Human Factors in Computing Systems Extended Abstracts. p. 1-7.

OBI, Ikechukwu et al. Let's Talk About Socio-Technical Angst: Tracing the History and Evolution of Dark Patterns on Twitter from 2010-2021. arXiv preprint arXiv:2207.10563, 2022.

O GLOBO. "Número de usuários do Instagram ultrapassa 2 bilhões e se aproxima do Facebook". 2022. Disponível em: <https://oglobo.globo.com/economia/tecnologia/noticia/2022/10/numero-de-usuarios-do-instagram-ultrapassa-2-bilhoes-e-se-aproxima-do-facebook.ghtml>.

POTEL-SAVILLE, Marie; DA ROCHA, Mathilde. From Dark Patterns to Fair Patterns? Usable Taxonomy to Contribute Solving the Issue with Countermeasures. In: Annual Privacy Forum, 2023. p. 145-165.

SANCHEZ CHAMORRO, Lorena; LALLEMAND, Carine; GRAY, Colin M. "My Mother Told Me These Things are Always Fake" - Understanding Teenagers' Experiences with Manipulative Designs. In: Proceedings of the 2024 ACM Designing Interactive Systems Conference. 2024. p. 1469-1482.

SOUSA, Carla; OLIVEIRA, Ana. The Dark Side of Fun: Understanding Dark Patterns and Literacy Needs in Early Childhood Mobile Gaming. In: European Conference on Games Based Learning. 2023. p. 599-610.

TRZASKOWSKI, Jan. Manipulation by design. Electronic Markets, v. 34, n. 1, p. 14, 2024.

Significados do discurso de ódio: desafios regulatórios e disputas discursivas na visão de diferentes setores sociais

Eliza Bacheга Casadei, Gabriela Agostinho Pereira

Resumo

O presente artigo analisa a complexidade das dinâmicas que envolvem a regulação de discursos de ódio em plataformas digitais. A pesquisa enfatiza que as soluções regulatórias demandam mais do que a mera identificação dos conteúdos ofensivos, mas envolvem também as formas pelas quais esse fenômeno é nomeado e interpretado. Com base na Análise de Discurso de Linha Francesa, são exploradas as respostas dos três setores sociais consultados na Consulta Pública sobre Regulação de Plataformas Digitais feita pelo CGI.br em 2023. A análise destaca as distinções entre os setores acerca do entendimento do discurso de ódio e, conseqüentemente, as soluções regulatórias, revelando as tensões que permeiam o campo da regulação digital e os desafios para estabelecer consensos em torno do enfrentamento do discurso de ódio nas plataformas digitais.

Palavras-chave: discurso de ódio; plataformas digitais; regulação; análise do discurso.

Introdução

A complexidade do enfrentamento às manifestações do discurso de ódio em plataformas digitais reside no fato de que as soluções regulatórias demandam mais do que a mera identificação dos conteúdos ofensivos: elas envolvem as formas pelas quais esse fenômeno é nomeado e interpretado no espaço social. Nesse sentido, as formas de tratar o problema dependem de uma partilha dos nomes - posto que o ato de nomear traz, em si, uma perspectiva normativa de como as coisas devem ser (Gomes, 2003) – e de uma prática discursiva que constrói sentidos e posiciona atores sociais de maneiras diferenciais no espaço discursivo. A identificação dos agentes, a construção dos enunciados e a circulação sobre o que constitui o discurso odioso em plataformas digitais revelam um campo de disputas semânticas e ideológicas, que materializam tensões entre as diferentes interpretações sobre o fenômeno que circulam no tecido social.

A Consulta Pública sobre Regulação de Plataformas Digitais feita pelo Grupo de Trabalho Regulação de Plataformas do Comitê Gestor da Internet do Brasil (CGI.br) em 2023 (Núcleo de Informação e Coordenação do Ponto BR, 2023) traz contribuições importantes para o enfrentamento ao discurso de ódio, ao mapear extensamente como diferentes setores sociais identificam e interpretam os agentes e soluções regulatórias possíveis.

A partir dos pressupostos teóricos da Análise de Discurso de Linha Francesa (Maingueneau, 2005), o objetivo do presente artigo é analisar as contribuições à Consulta do CGI.br sobre regulação de plataformas digitais para observar como os diferentes grupos sociais entrevistados articulam a construção discursiva em torno da definição dos discursos de ódio e quais implicações isso tem para as propostas de regulação de plataformas. A partir dessa perspectiva geral, discutiremos como os participantes entendem e definem o discurso de ódio; quais grupos são identificados como alvos ou propagadores desses discursos pelos entrevistados; quais posições de sujeito são reforçados nos discursos sobre regulação; e como os participantes percebem o papel das plataformas e do Estado em lidar com discursos de ódio e extremismos.

Os discursos, em seu conjunto, carregam “uma compreensão de como o mundo ‘funciona’ e, portanto, uma noção de como sociedades e indivíduos devem reagir a esse funcionamento” (Gomes, 2019, p. 273). Isso significa que os diferentes entendimentos sobre o que é discurso de ódio implicam em variações na legitimidade das soluções regulatórias propostas entre os grupos sociais.

A Análise do Discurso de linha francesa parte do pressuposto de que os sujeitos enunciadorees não são fonte original de seus próprios discursos, mas sim, atores que medeiam discursos sociais circulantes mais amplos que são materializados nas dispersões e/ou regularidades dos enunciados que, por sua vez, revelam diferentes formações discursivas sobre a construção dos objetos (Maingueneau, 2005). Assim, os discursos apresentados nas respostas à consulta pública do CGI.br revelam as distintas matrizes ideológicas e os sentidos em disputa na definição do problema e nas soluções propostas pela Comunidade Científica e Tecnológica, pelos representantes do Terceiro Setor e do Setor Empresarial. A análise evidencia que as falas dos participantes não apenas expressam opiniões individuais, mas reproduzem e disputam lugares discursivos associados a posições de sujeito, revelando as tensões que permeiam o campo da regulação digital e os desafios para estabelecer consensos em torno do enfrentamento do discurso de ódio nas plataformas digitais.

Entendimentos sobre o discurso de ódio

As palavras atuam a partir de uma “operação de demarcação”, ou seja, de “recorte a partir do qual um segmento se abre a construções imaginárias, a produções sociais, mecanismo de produção, sobretudo à possibilidade de disciplina e controle” (Gomes, 2003, p. 95). E é nesse sentido que os termos não são meramente um mecanismo de representação de mundo, mas sim, de apresentação de uma realidade “e de um pensar que se modela pelas palavras” (Gomes, 2003, p. 95).

O termo “discurso de ódio”, sob essa perspectiva, é um termo de difícil definição posto que materializa tensões de sentido em torno de processos de controle e disciplina. Não há uma definição fechada sobre o termo, posto que se trata de um conceito em disputa e em processo. Para Leal da Silva et al. (2011, p.447), o discurso de ódio “é uma manifestação segregacionista, baseada na dicotomia superior (emissor) e inferior (atingido) e, como manifestação que é, passa a existir quando é dada a conhecer por outrem [...]”. Waldron (2012), por sua vez, afirma que o termo deveria ser substituído por “difamação coletiva”, pois trata-se da difamação de grupos inteiros que geram consequências sociais. Já Norbert Elias e John L. Scotson (2000), mostram que a intolerância pode ter como base a luta pelo poder.

Nas respostas dos entrevistados da Consulta do CGI.br, é possível observar que, de uma maneira geral, há uma percepção de que o discurso de ódio é um risco significativo para a sociedade, intimamente relacionado ao poder das grandes plataformas digitais. Ao todo, é possível encontrar, na consulta, 74 respostas com menções ao discurso de ódio, com concentração no eixo “2.4. Grupo de Riscos - Riscos relacionados a ameaças à democracia e aos direitos humanos”, com 54 respostas.

O conjunto dessas respostas fornece elementos para analisar como diferentes grupos sociais entendem e definem o discurso de ódio, a partir de suas posições como sujeitos do discurso. Posto que “uma cultura é sustentada por diversas formações discursivas” que “concorrem para o conjunto que compõe o pensamento de uma comunidade” (Gomes, 2019, p. 276), as distintas definições de discurso de ódio mostram que “para cada discurso em particular, e para o todo das formações, aplica-se a regra da construção de um entendimento de mundo e das ações correlatas” (Gomes, 2019, p. 277). As interpretações de discurso de ódio, portanto, não descrevem um fenômeno, mas sim, atuam como expressões das formas particulares com que cada respondente percebe o mundo, seus perigos e os limites da liberdade. Esse conjunto de visões constrói uma teia de sentidos que sinaliza como os diferentes grupos sociais esperam que seja regulado o espaço público no contexto das plataformas digitais.

É necessário esclarecer que, ao longo do questionário, não é solicitado aos participantes que definam discursos de ódio, embora esse seja um termo bastante citado nas respostas.

Para que possamos delimitar os sentidos mobilizados para o termo, portanto, foi necessário identificar, em um primeiro momento, as recorrências discursivas que atrelam o discurso de ódio a outros discursos no interior das respostas. Isso é relevante porque o discurso se forma a partir de sua relação com outros discursos dentro de um campo discursivo. Assim, os discursos não existem de forma isolada, mas se constroem em relação (seja de aliança ou de conflito) com outros discursos. O interdiscurso estabelece “um sistema no qual a definição da rede semântica que delimita a especificidade de um discurso coincide com a definição de suas relações com o Outro” (Maingueneau, 2005, p. 35).

Após uma leitura extensa das respostas, foi possível delimitar as seguintes categorias, referentes às temáticas que se repetem em sua correlação com o discurso de ódio.

- Ameaça à democracia: os participantes frequentemente relacionam o discurso de ódio a uma ameaça aos valores democráticos, ao fomentar narrativas antidemocráticas e desinformativas. A propagação do discurso de ódio é vista como um risco, capaz de incitar ações extremistas que desafiam os direitos humanos e as instituições democráticas.
- Risco à liberdade de expressão: as falas sobre regulação do discurso de ódio muitas vezes são acompanhadas com a preocupação em equilibrar a necessidade de proteger a sociedade contra discursos odiosos, mas sem comprometer a liberdade de expressão. Algumas respostas destacam o desafio de evitar que as políticas para mitigar o discurso de ódio acabem limitando a capacidade dos cidadãos de expressarem suas opiniões e exercitarem direitos fundamentais.
- Modelo de negócio das plataformas: o modelo de engajamento das plataformas, que busca maximizar a interação dos usuários, é descrito como um fator central para a disseminação de conteúdos extremistas e de ódio. Esse modelo incentiva o engajamento por meio da exposição a conteúdos polarizadores, manipulando o comportamento dos usuários em troca de lucro.
- Radicalização e violência social: o discurso de ódio é frequentemente associado à radicalização e à promoção da violência. Participantes observam que as redes sociais têm um papel importante na amplificação de discursos que incentivam a discriminação, a violência e o extremismo.
- Moderação de conteúdos: muitos participantes sugerem que as plataformas precisam implementar políticas de moderação. Há a defesa do uso de algoritmos e de regras claras para identificar e coibir discursos odiosos. A criação de diretrizes específicas para a remoção de conteúdos violentos também é frequentemente mencionada.

A análise frequencial desses temas nas respostas revela que: o risco à liberdade de expressão está presente em 57 respostas (77% das respostas que mencionam discurso de ódio); Ameaça à democracia: 37 respostas (50%); Moderação de conteúdos: 30 respostas (40,5%); Modelo de negócio das plataformas: 22 respostas (29,7%); Radicalização e violência social: 21 respostas (28,4%).

Posto que uma mesma resposta pode ser enquadrada em mais de uma categoria, temos, como combinações mais frequentes:

- Ameaça à democracia e risco à liberdade de expressão: 11 respostas
- Risco à liberdade de expressão e moderação de conteúdos: 7 respostas
- Ameaça à democracia, risco à liberdade de expressão, radicalização e violência social: 7 respostas
- Ameaça à democracia, risco à liberdade de expressão, moderação de conteúdos: 5 respostas

Esses resultados revelam que, para os respondentes, a construção do sentido de “discurso de ódio” é atravessada por uma tensão discursiva entre a salvaguarda dos direitos individuais, especialmente a “liberdade de expressão”, e a necessidade de proteger o espaço público contra conteúdos que minem a democracia. O termo “liberdade de expressão” emerge como tema central e como ponto de disputa simbólica, frequentemente associado a questões de “ameaça à democracia” e “moderação de conteúdos”, refletindo posições ideológicas sobre os limites do aceitável no espaço digital.

Ainda que o discurso de ódio seja um tema transversal, cada setor traz à tona aspectos específicos, refletindo diferentes preocupações e abordagens. A partir dos dados analisados, observamos recorrências discursivas distintas entre os setores entrevistados, revelando preocupações e prioridades variadas de acordo com a área de atuação.

A tabela abaixo descreve os temas de maior preocupação em relação ao discurso de ódio para cada setor:

Categoria	Terceiro Setor (%)	Setor Empresarial (%)	Comunidade Científica e Tecnológica (%)
Ameaça à democracia	55,9	53,8	40,7
Modelo de negócio das plataformas	38,2	30,8	18,5
Moderação de conteúdos	38,2	15,4	55,6
Radicalização e violência social	26,5	7,7	40,7
Risco à liberdade de expressão	76,5	92,3	70,4

A tabela evidencia que o setor empresarial apresenta uma preocupação destacada com a liberdade de expressão (92,3%), frequentemente associando-a a uma regulação que não comprometa o ambiente econômico. O Terceiro Setor também prioriza a liberdade de expressão (76,5%), mas destaca fortemente o impacto do discurso de ódio na democracia (55,9%) e nos direitos sociais. Já a Comunidade Científica e Tecnológica mostra-se particularmente atenta à moderação de conteúdos (55,6%), indicando a regulação técnica como uma prioridade para enfrentar o problema.

De uma maneira geral, a Comunidade Científica e Tecnológica aborda o discurso de ódio principalmente como parte de uma questão regulatória mais ampla, onde definições técnicas e a responsabilidade das plataformas são temas centrais. Para os respondentes deste setor, o discurso de ódio é frequentemente mencionado em conjunto com a necessidade de moderação de conteúdo e governança estruturada, refletindo um foco na elaboração de critérios claros para a regulamentação.

Para exemplificar esse aspecto, podemos citar a resposta do Centro de Tecnologia e Sociedade da FGV Direito Rio à pergunta 30 (“Caso haja medidas de mitigação para os riscos associados a infodemias como desinformação, extremismos, discurso de ódio e discurso terrorista que não tenham sido mencionadas, descreva a seguir”), em que se afirma que:

Iniciativas em âmbito público e privado, em nível global, reconhecem a necessidade de obrigações de transparência mais rigorosas para grandes plataformas, desfrutando atualmente de um bom nível de consenso entre os diferentes setores da sociedade. O desafio imposto às medidas de mitigação relacionadas à falta de transparência é definir princípios, diretrizes e critérios claros para orientar decisões sobre o que se deve publicizar, para quem,

em que contexto e de que forma, além de detalhar medidas e obrigações de transparência específicas para plataformas digitais, respeitando os diversos contextos de atuação.

Também em relação à mesma pergunta, Alex Camacho Castilho traz a preocupação em abordar critérios objetivos para mitigação dos efeitos de discurso de ódio como “Promoção de alfabetização midiática e digital”, “Parcerias com verificadores de fatos”, “Promoção de padrões de moderação mais rigorosos”, entre outros.

Conforme evidenciado nas respostas, a Comunidade Científica e Tecnológica tende a enfatizar a importância de uma abordagem regulatória técnica e pautada em critérios objetivos, contemplando a transparência nos processos de moderação e a criação de diretrizes claras para o tratamento do discurso de ódio.

Em relação ao Setor Empresarial, essa preocupação também é evidente e recorrente nas respostas. Há, contudo, algumas particularidades. Um outro discurso emoldura as respostas e propõe uma cena discursiva distinta. No Setor Empresarial, o discurso de ódio é também abordado sob a ótica dos riscos econômicos e do impacto no poder de mercado. Os participantes deste setor relacionam, com maior frequência, o discurso de ódio a práticas comerciais e ao impacto econômico, destacando a responsabilidade regulatória como uma ferramenta para proteger o ambiente de negócios e a confiança dos consumidores. O discurso de ódio é, portanto, integrado a uma visão mais ampla de proteção de mercado, onde a regulação ajuda a estabelecer um ambiente digital confiável e competitivo.

Para Sérgio Garcia Alves, por exemplo, “Empresas de tecnologia têm desafiado os limites da ciência da computação para promover a identificação e remoção de conteúdo problemático em escala. Tais avanços somente são possíveis em contextos legais que ofereçam flexibilidade para a inovação, ou seja, que não criem soluções rígidas que não se adaptem ao desenvolvimento tecnológico”. Para Thais Covolato, “um excesso na responsabilização das plataformas por conteúdos sejam eles monetizados ou não, pode gerar o chamado chilling effect ou efeito inibidor” para o ambiente de negócios.

Além disso, trata-se do setor que demonstra maior preocupação com a noção de que os mecanismos regulatórios contra o discurso de ódio não atrapalhem a liberdade de expressão. Para a Câmara Brasileira da Economia Digital, “é essencial dispor de uma legislação resiliente sobre direitos de liberdade de expressão que sopesa a sua importância ao mesmo tempo que leva em consideração a proteção aos direitos humanos (discurso de ódio, terrorismo, incitamento a tais comportamentos)”. Para a Brasscom, é necessário considerar que “ainda que usuários compartilhem conteúdos potencialmente ilícitos, a criação de regimes de responsabilidade de intermediário (...) potencialmente representam restrições à liberdade de expressão dos usuários”.

O Setor Empresarial, portanto, propõe um entendimento em que a regulação do discurso de ódio se entrelaça com a necessidade de manter um ambiente de negócios propício à inovação e à competitividade. Isso é suportado, em grande medida, por um discurso que defende a liberdade de expressão e modelos mais flexíveis de regulação.

O Terceiro Setor coloca o discurso de ódio no enquadramento discursivo dos direitos humanos e do impacto social, enfatizando a proteção das minorias e a preservação democrática. Este setor é o que mais defende uma regulação que impeça a disseminação de preconceitos e discursos autoritários. Os participantes do Terceiro Setor frequentemente contextualizam o discurso de ódio dentro de debates mais amplos sobre equidade, enfatizando a importância de políticas regulatórias que impeçam a perpetuação de discriminações e de atitudes que enfraqueçam a coesão social e os valores democráticos.

O GT-RI ISOC Brasil, por exemplo, destaca que “verifica-se uma tendência preocupante de governos em usar tecnologia e controle da Internet para suprimir a dissidência e fortalecer seu poder. Isso pode tomar várias formas, desde a censura e o bloqueio de sites até a vigilância massiva e a coleta de dados pessoais”. Para Rhiana Valois, “a defesa dos direitos

humanos estabelece ferramentas na forma de direitos que podem ser utilizados para mitigar esses problemas”. Para Blogueiras Negras, precisamos “considerar a moderação de conteúdo feita por pessoas humanas que representem a pluralidade de raça, gênero, identidade de gênero e marcadores afins, que sejam contratadas proporcionalmente de acordo com o número de pessoas que usam cada plataforma”.

O enquadramento discursivo adotado pelo Terceiro Setor associa os discursos de ódio de forma mais contundente à defesa de direitos humanos e da justiça social, bem como à defesa dos valores democráticos. A moderação de conteúdo, a partir desse viés discursivo, deve fortalecer e sustentar os valores democráticos e proteger grupos sociais mais vulneráveis contra as violências em plataformas digitais.

A análise por setores sociais revela disputas discursivas significativas em torno da construção do sentido do conceito de discurso de ódio. Embora haja preocupações transversais, cada setor mobiliza o termo de acordo com posições ideológicas, projetando diferentes demandas e pressões sobre as práticas regulatórias. A partir da análise, é possível observar como o processo interdiscursivo se manifesta em torno das distintas visões sobre o discurso de ódio. Enquanto compartilham a preocupação com o impacto do discurso odioso, os setores o abordam a partir de lentes próprias, trazendo diferentes expectativas para a regulação. O processo interdiscursivo, assim, surge como um campo de tensões e negociações discursivas, onde cada setor dialoga com os outros, ao mesmo tempo em que reforça suas próprias posições a partir de enquadramentos discursivos distintos.

Constituição de sujeitos discursivos

As disputas discursivas em torno do conceito de discurso de ódio também são materializadas, nas respostas, a partir da constituição de posições-sujeito construídas *no e pelo* discurso. Essas posições-sujeito não são meramente papéis assumidos, mas práticas discursivas que emolduram o conceito de discurso de ódio, produzindo efeitos de sentido que legitimam ou deslegitimam ações e exigências regulatórias. Elas também atuam como instâncias de produção de sentido que delineiam o que deve ser considerado discurso de ódio e quais políticas são necessárias para seu enfrentamento.

Em relação à constituição dos sujeitos discursivos, serão observadas, nas respostas: (1) Quem são os sujeitos descritos como responsáveis pela disseminação dos discursos de ódio; e (2) Quem são os alvos/vítimas desses discursos odiosos.

Em relação às posições descritas como responsáveis pela disseminação dos discursos de ódio, nas respostas analisadas, as plataformas digitais são mencionadas como grandes contribuidoras para a disseminação de conteúdos potencialmente prejudiciais em todos os setores. Elas são descritas como facilitadoras devido à sua estrutura de engajamento, o que contribui para a disseminação de discursos polarizantes. Além disso, há uma expectativa, nas respostas, de que as plataformas sejam reguladas e responsabilizadas, com sugestões de que deveriam agir ativamente para proteger os usuários e moderar o conteúdo. Os enquadramentos em relação a esse tema seguem a estrutura discursiva delineada no item anterior, onde a Comunidade Científica e Tecnológica parece preocupada com o delineamento de diretrizes claras e objetivas para o enfrentamento do problema; o Setor Empresarial reflete uma preocupação com os riscos da regulação para o mercado e com a liberdade de expressão; e o Terceiro Setor com os riscos sociais e para a democracia que o discurso de ódio pode trazer.

Há, ainda, menções a grupos que disseminam “pensamentos autoritários” (como no comentário de Raul Luis, representante do terceiro setor: “As grandes corporações através de suas plataformas digitais estão desestabilizando democracias, interferindo na cultura do povo, disseminando a cultura do ódio, disseminando preconceitos, permitindo a proliferação de pensamentos autoritários”), indicando uma percepção de que certos grupos

ideológicos são propagadores desses discursos. As referências a extremismo e autoritarismo aparecem geralmente como preocupações amplas sobre o conteúdo disseminado online e os riscos para a sociedade, sem apontar diretamente para grupos definidos como disseminadores desses pensamentos.

Especialmente nos comentários do terceiro setor, há a preocupação de que o Estado atue de forma autoritária na regulação (como nos comentários “o Brasil se aproxima de realidades políticas autoritárias ao redor do mundo que vêm controlando o acesso à Internet de maneira arbitrária” e “É preciso, portanto, ter cuidado com esse tipo de determinação, uma vez que conceitos indeterminados ou vagos podem abrir brechas para interferências no direito de liberdade de expressão e participação das pessoas, principalmente em países com histórico autoritário”). Os dados mostram que o Terceiro Setor é o mais preocupado com questões relacionadas ao autoritarismo, com 19 menções. Em seguida, a Comunidade Científica e Tecnológica apresenta 11 menções e o Setor Empresarial conta com 7 menções nas respostas analisadas.

A análise das respostas evidencia que as plataformas digitais são majoritariamente descritas como agentes centrais na disseminação dos discursos de ódio, posicionando-as como sujeitos que, por sua arquitetura, contribuem para a proliferação de ideologias extremistas, projetando, portanto, uma responsabilidade implícita na propagação de discursos nocivos. A definição de outros responsáveis, contudo, carrega uma ambiguidade discursiva onde, de um lado, há referências a grupos autoritários que promovem discursos de ódio (embora esses grupos nem sempre sejam claramente identificados); de outro, o próprio Estado é objeto de uma desconfiança latente, uma vez que sua intervenção regulatória é vista com cautela, devido ao risco de que essas ações possam assumir contornos autoritários. Esse jogo discursivo evidencia uma tensão entre a dificuldade de responsabilizar agentes sociais específicos e a preocupação com a possibilidade de o Estado, sob o pretexto de combate ao discurso de ódio, exercer um controle excessivo sobre as vozes e os conteúdos na esfera pública.

Em relação aos alvos/vítimas preferenciais do discurso de ódio, as respostas de cada setor variam consideravelmente. O Terceiro Setor apresenta 17,6% de seus comentários mencionando diretamente vítimas específicas como jovens, mulheres e outros grupos vulneráveis. Essa proporção é significativamente maior do que as de outros setores, reforçando a preocupação do Terceiro Setor com a proteção de determinados grupos. Em contrapartida, a Comunidade Científica e Tecnológica e o Setor Empresarial abordam a questão do discurso de ódio de forma mais ampla. Apenas 7,4% dos comentários da Comunidade Científica e 7,7% dos comentários do Setor Empresarial fazem referência a grupos específicos. Estes setores tendem a focar mais em riscos e responsabilidades gerais, sem detalhar quem são os alvos específicos do discurso de ódio, indicando uma abordagem mais estruturante e menos voltada a minorias identificáveis.

Isso pode ser observado em comentários como o de Bruno Ribeiro, representante da Comunidade Científica e Tecnológica, que coloca que “é fundamental que a regulamentação das plataformas digitais seja cuidadosamente elaborada, levando em consideração a complexidade do ambiente digital, bem como os direitos e interesses de todos os envolvidos, incluindo usuários, empresas e a sociedade como um todo”. Assim, a construção discursiva em torno dos grupos vulneráveis ou vítimas de discursos de ódio ocorre de maneira genérica, sem nomeação explícita de minorias ou grupos específicos.

Isso é relevante posto que a constituição de sujeitos discursivos no discurso odioso, para Butler (2021, p. 11), é particularmente importante posto que “se ser chamado é ser interpelado, a denominação ofensiva tem o risco de introduzir no discurso um sujeito que utilizará a linguagem para rebater a denominação”. O discurso articula uma existência social uma vez que “o chamamento constitui um ser no interior do circuito possível de reconhecimento e, conseqüentemente, fora dele, na abjeção” (Butler, 2021, p. 13).

Ao evitar mencionar identidades específicas, o discurso se desvia de confrontar diretamente as desigualdades que amplificam a vulnerabilidade de determinados grupos em relação a outros. Essa neutralidade implícita constrói uma cena discursiva que universaliza a preocupação com o discurso de ódio e, ao mesmo tempo, silencia a necessidade de medidas específicas de proteção para grupos mais afetados.

Trata-se de um enquadramento distinto daquele adotado pelo Terceiro Setor. Mulheres Negras, por exemplo, apontam que “É preciso levar em conta que no contexto local as mulheres negras são um dos grandes alvos dos discursos e práticas de ódio que assolam nossa sociedade”; Victor Durigan aponta que é “importante mencionar riscos relacionados à radicalização social, violência contra grupos sociais vulneráveis ou marginalizados, como é o caso da violência política contra mulheres, em especial contra candidatas em períodos eleitorais ou jornalistas”; Bia Barbosa aponta que “Não se pode abrir margem para uma proteção jurídica a discursos violentos contra determinados grupos sociais vulnerabilizados, como a população LGBTQIAP+, pessoas negras, mulheres, pessoas com deficiência etc.”.

Nessas respostas, ao mencionar diretamente grupos específicos como mulheres, jovens, pessoas negras e a população LGBTQIA+, destaca-se uma abordagem convocatória na constituição dos sujeitos discursivos alvos de discursos de ódio. O Terceiro Setor articula uma preocupação explícita com a vulnerabilidade desses grupos e, conseqüentemente, enquadra-os discursivamente como sujeitos que devem ser protegidos nas medidas de regulação. Ao nomear esses sujeitos, o Terceiro Setor traz a necessidade de regulação específica que reconheça a especificidade das violências enfrentadas por essas populações. Dessa forma, o discurso convocatório do Terceiro Setor constrói uma cena discursiva de apelo político, onde a proteção desses sujeitos vulneráveis é descrita como uma responsabilidade coletiva no enfrentamento ao discurso de ódio.

Efeitos de sentido reforçados nos discursos sobre regulação

Conforme evidenciado no item anterior, a análise revela como os discursos dos participantes constroem as plataformas como “sujeitos responsáveis” por mitigar os discursos de ódio e extremismos. As plataformas digitais são frequentemente descritas como detentoras de um poder econômico excessivo, o que gera preocupações sobre a sua influência na circulação de informações e no discurso público. Participantes mencionam que o poder de mercado das plataformas afeta a segurança, privacidade e liberdade dos usuários, e que uma regulação mais robusta é necessária para mitigar esses efeitos.

A escolha entre autorregulação e regulação estatal, contudo, mostra diferentes atribuições de responsabilidade. A Análise de Discurso ajuda a interpretar esses posicionamentos ideológicos subjacentes: cada proposta de intervenção carrega valores e ideologias específicas. Defender a autorregulação implica em uma ideologia de liberdade de mercado e confiança nas corporações, enquanto a regulação estatal pode sugerir uma postura de proteção social e vigilância sobre o poder econômico.

Em uma perspectiva geral, 2,7% dos comentários mencionam a autorregulação, sugerindo que há pouca confiança nas plataformas para se autorregular sem intervenção externa. 12,2% dos comentários fazem referência à regulação estatal, proteção social ou controle, indicando uma preferência maior por uma intervenção que limite o poder das plataformas e proteja os usuários.

A análise dos comentários por setor revela diferenças nas abordagens em relação à regulação das plataformas digitais. No Setor Empresarial, 23,1% dos comentários abordam a necessidade de regulação estatal. Trata-se, ainda, do setor que se mostra menos favorável à autorregulação das plataformas. O Setor Empresarial é o mais cauteloso em relação à regulamentação, mencionando preocupações sobre possíveis restrições à liberdade de

expressão e os efeitos de uma regulação que poderia ser onerosa para as empresas. A Comunidade Científica e Tecnológica, por outro lado, possui um total de 27 comentários sobre esse assunto, dos quais 7,4% mencionam a necessidade de regulação estatal, mas nenhuma referência a autorregulação. O Terceiro Setor se destaca por apresentar uma combinação de perspectivas: 14,7% de seus comentários são favoráveis à regulação estatal, enquanto 5,9% mencionam autorregulação.

A exceção de poucas respostas – como por exemplo, da ABRANET, que sugere que as “práticas de autorregulação” devem ter “sua importância aqui reconhecida e preservada” – há, em todos os setores, uma postura cética em relação à capacidade das plataformas de se autorregular em no enfrentamento ao discurso de ódio, em um posicionamento ideológico que media a ideia de que as plataformas não devem ser as únicas responsáveis pelo controle de conteúdos potencialmente prejudiciais. Isso fica evidente em comentários como o do Instituto de Pesquisa em Direito e Tecnologia do Recife (“Por muito tempo, as plataformas adotaram um modelo de autorregulação que, como se pôde notar, não foi suficiente para barrar problemas como desinformação e discurso de ódio na Internet”) ou do Intervozes (“Não podemos defender a autorregulação a cargo apenas das plataformas”).

Defanti (2018, p. 149) afirma que no Brasil, o estudo da regulação é frequentemente associado ao papel do Estado, especialmente às agências reguladoras – vínculo este decorrente da tradição brasileira de concentrar a responsabilidade pela regulação em órgãos estatais específicos, como as agências, que têm o papel de supervisionar, normatizar e fiscalizar setores estratégicos. Esse cenário contribui para que o conceito de regulação no Brasil seja amplamente associado a uma intervenção direta do Estado, ao invés de processos de autorregulação corporativa (Maradei Júnior, 2023; Pacheco, 2006). Sob a perspectiva da Análise de Discurso, a cultura exerce uma influência profunda sobre a construção de sentidos no discurso, deixando marcas que refletem valores, práticas e crenças socialmente compartilhadas.

Nas soluções propostas para o enfrentamento do discurso de ódio, isso se manifesta em práticas discursivas enraizada em uma cultura de valorização da intervenção pública. Trata-se de um traço que estrutura a forma como os indivíduos interpretam e respondem às tensões das dinâmicas sociais.

Considerações finais

A análise das respostas à consulta pública realizada pelo CGI.br permitiu identificar o enfrentamento ao discurso de ódio em plataformas digitais como um espaço de disputas discursivas, que, por sua vez, implicam nos juízos de valor que são dados às soluções regulatórias pelos diferentes grupos sociais. Essas disputas revelam não apenas as complexidades envolvidas na regulação de conteúdos violentos, mas também o caráter ideológico presente em cada setor do levantamento.

A divisão de setores sociais permitiu observar que, ainda que temas como democracia, liberdade de expressão, modelo de negócios das plataformas, radicalização, violência e moderação de conteúdo estejam presentes nas respostas dos representantes de todos os setores (Comunidade Científica e Tecnológica, Terceiro Setor e Setor Empresarial), o foco atribuído a cada um deles é distinto.

A construção de sentido acerca do conceito de discurso de ódio possui certa transversalidade, mas, a partir dos dados analisados, observamos diferentes prioridades de acordo com o setor e visões distintas sobre o conceito. Além disso, cada setor projeta diferentes expectativas regulatórias que se alinham com suas posições ideológicas particulares. Para o Setor Empresarial, a prioridade é garantir a proteção de uma liberdade de expressão que não afete o ambiente econômico. O Terceiro Setor, por sua vez, além de

priorizar a liberdade de expressão, também se preocupa com os efeitos que o discurso de ódio pode exercer na democracia e zela pela proteção de grupos potencialmente vulneráveis. Por fim, a Comunidade Científica e Tecnológica está preocupada, de maneira geral, com a moderação de conteúdos e a regulação técnica das plataformas.

Essas distinções de enfoque estão presentes não apenas em relação ao conceito de discurso de ódio, mas também ao observarmos nas respostas dos entrevistados quais grupos são considerados alvos ou propagadores desses discursos e o papel das plataformas digitais e do Estado no enfrentamento do fenômeno. As diferenças nas respostas dos setores revelam as interpretações ideológicas de cada grupo e as disputas de significado que permeiam o campo da regulação no Brasil.

É preciso destacar que em todos os setores, as plataformas digitais são percebidas como centrais na circulação e disseminação de conteúdos odiosos, violentos e polarizadores, evidenciando a necessidade de serem responsabilizadas de alguma forma pelo conteúdo presente em seus espaços. Também foi possível observar que de maneira geral, as propostas de enfrentamento e regulação estão ancoradas em uma valorização da intervenção pública, uma vez que há desconfiança em todos os setores na capacidade de autorregulação das plataformas.

Por fim, consideramos que soluções regulatórias eficientes devem ser pensadas visando o equilíbrio entre as preocupações de cada setor social. A complexidade do discurso de ódio plataformizado faz emergir a necessidade de uma regulação que considere as especificidades desse contexto e as disputas discursivas e ideológicas entre os atores envolvidos.

Referências

BUTLER, Judith. *Discurso de ódio: uma política do performativo*. São Paulo: Editora Unesp, 2021.

DEFANTI, Francisco. “Um ensaio sobre a autorregulação: características, classificações e exemplos práticos”. *Revista de Direito Público de Economia*, v. 16, n. 63, 2018, p. 149-181.

ELIAS, Norbert; SCOTSON, John L. *Os estabelecidos e os outsiders: sociologia das relações de poder a partir de uma pequena comunidade*. Rio de Janeiro: Jorge Zahar, 2000.

GOMES, Mayra Rodrigues. “Palavra de ordem / dispositivo disciplinar”. *Galáxia*, n. 5, v. 1, 2003, p. 91-108.

GOMES, Mayra Rodrigues. “As materialidades e seus discursos”. *Comunicação, Mídia e Consumo*, v. 16, n. 46, 2019, p. 271-290.

LEAL DA SILVA, Rosane et al. “Discursos de ódio em redes sociais: jurisprudência brasileira”. *Revista Direito - GV*, v. 7, n. 2, p. 445-468, jul./dez. 2011.

MAINGUENEAU, Dominique. *Gênese dos Discursos*. Curitiba: Criar Edições, 2005.

MARADEI JÚNIOR, João Carlos. “O papel das agências reguladoras na tutela dos direitos do consumidor”. *Revista Digital de Direito Administrativo*, v. 10, n. 1, 2023.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR. *Sistematização das contribuições à consulta sobre regulação de plataformas digitais*. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2023.

PACHECO, Regina. “Regulação no Brasil: desenho das agências e formas de controle”. *Revista de Administração Pública*, v. 40, n. 4, 2006.

WALDRON, Jeremy. *The Harm in Hate Speech*. Harvard, versão e-book, Kodo, 2012.

Regulação da desinformação nas plataformas digitais e o contexto eleitoral brasileiro

Marina Giovanetti Lili Lucena, Leandro Léo Rebelo, Giovanna Milanese Tavares

Resumo: O presente trabalho investiga a regulação das plataformas digitais no contexto eleitoral. A pesquisa é documental e jurisprudencial, e utiliza a literatura especializada, as opiniões na consulta pública do Comitê Gestor da Internet no Brasil (CGI.br), as decisões do Tribunal Superior Eleitoral (TSE) e as normas eleitorais, notadamente as Resoluções publicadas pelo TSE. Será analisado o conceito de desinformação e os termos relacionados utilizados nas resoluções eleitorais. A partir dessa análise conceitual, busca-se compreender se as medidas de mitigação propostas pela consulta pública no contexto das eleições são inovadoras ou se já estão previstas na legislação eleitoral brasileira.

Palavras-chave: Desinformação. Regulação. Plataformas digitais. Eleições. TSE.

Introdução

O tema “desinformação” ganhou novos contornos nos últimos anos devido ao crescimento das plataformas de redes sociais, nas quais conteúdos dotados de algum tipo de falsidade são facilmente reproduzidos e compartilhados, atingindo público relevante em poucos minutos. Assim, o presente artigo analisará a regulação das plataformas no contexto eleitoral.

Em períodos de eleições, há a realização mais contundente de manifestações por parte dos candidatos(as) e partidos políticos, principalmente por meio da propaganda eleitoral, as quais manifestações ocorrem nos mais diversos meios (presencialmente, na rádio, na televisão e na Internet), todos eles regulamentados pela legislação eleitoral. No contexto eleitoral, a desinformação é ainda mais preocupante, já que pode influenciar a forma como os eleitores recebem informações e, conseqüentemente, pode alterar seu voto. Trata-se de temática essencial para o futuro da Internet no Brasil, visto que os períodos eleitorais são recorrentes e, dada a importância do ambiente digital e das plataformas digitais no contexto brasileiro, é razoável esperar que a incidência de propagandas e manifestações políticas nesses ambientes seja cada vez mais forte no futuro¹.

Em virtude da relevância da temática sobre regulação das plataformas no contexto brasileiro, o Comitê Gestor da Internet no Brasil (CGI.br) realizou consulta pública sobre regulação de plataformas digitais em 2023, com 1.336 contribuições de 140 pessoas físicas e representantes dos quatro setores que compõem o CGI.br (governamental, terceiro setor, setor empresarial e comunidade científica e tecnológica) (NIC.br, 2023).

A pesquisa realizada será documental e de jurisprudência. Na primeira, é feita análise da consulta do CGI.br sobre regulação de plataformas digitais, em específico o tema das democracias e processos eleitorais. Assim, é feito um recorte para estudo

¹ Um exemplo é o da candidatura de Pablo Marçal para a prefeitura de São Paulo. Embora o candidato do Partido Renovador Trabalhista Brasileiro (PRTB) não tenha se apresentado no horário eleitoral na TV e no rádio durante o primeiro turno, em razão de sua forte presença nas redes sociais, ele obteve votação expressiva no primeiro turno, com 28,14% dos votos.

das manifestações que abordam o problema da desinformação nas plataformas digitais durante o período eleitoral, incluindo seu conceito. Na jurisprudência, são exploradas as decisões do Tribunal Superior Eleitoral (TSE) para melhor compreensão dos termos relacionados à desinformação no contexto eleitoral, especificamente “fatos notoriamente inverídicos” e “fatos gravemente descontextualizados”, mencionados em diversos dispositivos da Resolução n. 23.610 (Brasil, 2019) do TSE, que estabelece regras para a propaganda eleitoral².

O objetivo geral deste artigo é compreender em que medida as contribuições à consulta pública do CGI.br refletem as iniciativas existentes de regulação de plataformas, especialmente as do TSE. Será também analisado se as medidas de mitigação propostas ao contexto eleitoral têm sido implementadas, a fim de compreender se a consulta pública traz inovações nessa temática.

No primeiro item, o artigo aborda o conceito de desinformação. As contribuições concordaram que riscos, como a desinformação, devem ser considerados para a regulação de plataformas; não existe, contudo, um consenso sobre como especificamente regular e remover a desinformação. Por outro lado, na esfera eleitoral, o limite entre “fatos” (passíveis de verificação) e “opiniões” (que não o são) é especialmente tênue. Assim, a seção discutirá os desafios em estabelecer essa definição a partir de normativas internacionais, julgamentos de cortes estrangeiras e insumos trazidos na consulta pública.

A segunda parte analisa o conceito de desinformação na visão do TSE. Esse tópico aborda como o tribunal tem conceituado desinformação no contexto eleitoral, especialmente à luz de sua jurisprudência mais recente, apresentando uma análise de precedentes relacionados ao que ele considera fatos inverídicos e descontextualizados, conceitos centrais na regulação de propaganda eleitoral que aparecem com destaque na Resolução n. 23.610 (Brasil, 2019).

Por fim, a terceira parte estuda as medidas de mitigação da desinformação eleitoral sugeridas na consulta pública e as medidas implementadas pelo TSE. Nesse item, são analisadas as contribuições feitas na consulta pública sobre as formas de mitigação da desinformação, essencialmente no contexto eleitoral, visando compreender se há inovações e avanços nas medidas sugeridas, ou seja, se essas sugestões estão incorporadas no ordenamento jurídico brasileiro.

A partir dessas análises, busca-se concluir se a consulta pública do CGI.br traz avanços e desenvolvimentos em duas esferas principais: i) na definição da desinformação eleitoral e ii) na sugestão de medidas que visam combater a propagação de conteúdo desinformativo eleitoral.

1 Conceito de desinformação: o verdadeiro e o falso

A pergunta colocada pelo item 2.4 da consulta pública é se as “infodemias, como desinformação, extremismos, discurso de ódio”, entre outros, constituem um risco à Democracia e aos Direitos Humanos. Das 20 respostas recebidas, 17 concordam que sim. Contudo, o consenso em *regular* plataformas digitais é mais fácil de atingir do que o consenso sobre *como fazê-lo* (Scott, 2023). De fato, diante de perguntas específicas sobre a implementação de medidas, as contribuições dividem-se.

² Os artigos que mencionam estes termos são: arts. 9º-C, *caput*; 9º-D, *caput* e § 1º; 9º-E, inc. II; 9º-F, *caput*; 9º-G, *caput* e § 7º; e 38, § 8º-A da Resolução n. 23.610 (Brasil, 2019).

Nesse contexto, as contribuições evitam comentar a respeito de como identificar os conteúdos que apresentam riscos. A respeito da desinformação, sob o item 27.i, o IP.rec traz as definições sugeridas pelo Relatório de 2017 ao Conselho da Europa (Wardle; Derakshan, 2017), enquanto Irineu Barreto afirma, de forma mais sucinta, que *fake news* são parte de uma “sofisticada estratégia de comunicação”³. As demais contribuições a este item não estão preocupadas de forma central em definir a desinformação.

O objetivo deste artigo não é chegar à definição *correta* da desinformação, tampouco descrever as garantias procedimentais adequadas para sua aplicação prática. De fato, não parece ser possível, ou desejável, chegar a definições que não lancem mão de termos abstratos e não sejam flexíveis diante de cada caso⁴. A profusão de definições acadêmicas de “desinformação” e seus termos correlatos (Tandoc Júnior; Lim; Ling, 2018), contudo, indica que a definição adotada tem consequências práticas. Não por outra razão, inúmeras contribuições destacam a necessidade de definições “precisas”: a Coalizão Direitos na Rede, no item 29.i, fala de não incluir a ideia de “incitação” entre os ilícitos, e o IAB Brasil, no item 27.i, indica a necessidade de evitar “subjetividade”.

A complexidade da definição adotada deve considerar quem a aplicará. Quanto mais complexos forem os ilícitos, mais minuciosa deve ser a avaliação legal do conteúdo a ser removido. Essa dinâmica reflete-se nas contribuições à consulta pública, que sugerem evitar a atribuição de análises complexas às plataformas. A contribuição de *Electronic Frontier Foundation* e *Access Now* fez referência a declarações da Relatoria Especial para Liberdade de Expressão da Comissão Interamericana de Direitos Humanos, as quais afirmam que as plataformas digitais não são capazes de fazer análises legais precisas dos conteúdos. No mesmo sentido, a Câmara Brasileira da Economia Digital aponta, no item 29, que mesmo os órgãos jurisdicionais “têm dificuldade” na análise da licitude de conteúdos, devido à “falta de clareza” dos ilícitos. O ITS Rio, por sua vez, ressalta, no item 29.i, que “não se pode esperar que plataformas digitais façam uma análise detida dos conteúdos”.

A mesma abordagem pode ser encontrada nas decisões da Corte de Justiça da União Europeia, que pode determinar judicialmente a ilegalidade de certas publicações, ficando os provedores de plataformas obrigados a remover conteúdo “idêntico”. A legislação brasileira apresenta uma disposição similar: o TSE realiza uma análise extensa da legalidade de conteúdos, e os demais atores ficam vinculados a essa decisão (conforme art. 9º-F da Resolução 23.610 [Brasil, 2019]). Nos casos *Eva Glawischmig-Piesczek* e *C-401/19* (Zingales *et al.*, 2024b), julgou-se que não devem ser criadas obrigações às plataformas digitais que requeiram uma “avaliação independente” da legalidade dos conteúdos discutidos. Isso é especialmente relevante quando a ilegalidade se baseia em elementos contextuais, por exemplo; como nota o IAB Brasil, no item 27.i, diferentes ilícitos contam com diferentes padrões de prova.

É amplamente conhecido que as plataformas digitais processam suas ações de moderação em uma escala vastamente superior à de mecanismos regulatórios estatais (Douek, 2019). Ainda que seja possível esperar que o parecer legal do TSE seja mais bem-fundamentado que o das plataformas, é razoável supor que as plataformas examinarão um volume de conteúdo muito superior ao do Judiciário.

³ Esta contribuição, embora destaque aspectos da desinformação relevantes a seu autor, não tenta definir desinformação para fins de moderação de conteúdo.

⁴ Para uma análise mais específica de definições de desinformação, consulte Zingales *et al.* (2024a, p. 2-5).

Portanto, pode ser problemático que as determinações sobre a permissibilidade de conteúdo sejam majoritariamente feitas por uma entidade apontada como ausente de capacidade institucional (ou de legitimidade⁵) de fazê-lo adequadamente. Se a grande maioria dos participantes da consulta pública concorda que certos tipos de conteúdo conformam *riscos*, deixá-los no ar até a decisão definitiva de sua legalidade pode apenas agravar os danos eventualmente causados. Como aponta Douglas Laycock (1990), há situações em que os danos gerados pela expressão são de difícil ou impossível reparação, o que desafia propostas que dependem de uma análise pormenorizada do conteúdo.

Por fim, vale discutir uma questão específica presente na definição legal de desinformação. De forma geral, as definições de desinformação tendem a enfatizar o aspecto da *veracidade* da informação veiculada (Cavaliere, 2022). Não há dúvidas de que as características específicas das redes sociais apresentam novos desafios para a regulação, ainda que, em muitas instâncias, seja inteiramente possível aplicar normativas pré-existentes (Lessig, 1999). Aparentemente no sentido contrário, os sistemas regionais de Direitos Humanos estabeleceram, em casos anteriores à Internet, uma contundente jurisprudência *limitando* a possibilidade de sancionar expressões por não serem verídicas.

No Sistema Interamericano, o maior exemplo é o caso *Tristán Donoso vs. Panamá*. Donoso fez acusações ao Procurador Geral do Panamá e foi condenado por calúnia, não tendo conseguido prová-las em juízo. A própria Corte não afirma que há provas que corroborem as acusações. Contudo, dado o evidente interesse público das afirmações e a existência de razões para que ele acreditasse no que dizia, a Corte entendeu que as acusações eram permissíveis. O Sistema Europeu de Direitos Humanos igualmente não busca determinar se as informações no centro das disputas que analisa são ou não verdadeiras. O Tribunal investiga, por exemplo, se houve boa-fé, se as informações foram apuradas de acordo com a ética jornalística e se foram usados métodos adequados para a obtenção das informações (Cavaliere, 2022).

Assim, colocar a avaliação da veracidade das informações como crucial para julgar sua licitude apresenta desafios práticos significativos que precisam de definição substancial antes de que qualquer ação seja tomada. Mesmo sem uma única resposta possível, essa questão precisa ser diretamente enfrentada pelas normativas.

2 Desinformação no contexto eleitoral: fatos inverídicos ou descontextualizados

A desinformação no contexto eleitoral apresenta desafios substanciais à democracia brasileira. Como abordado no item anterior, é complexo delimitar previamente o que é “desinformação”. No entanto, tratando-se de um passo essencial para regular a questão, o TSE tem assumido esse papel por meio de suas resoluções e entendimentos jurisprudenciais que tratam diretamente da temática.

O TSE é o órgão superior da Justiça Eleitoral brasileira (art. 118 [Brasil, 1988]) e tem competência para regulamentar questões necessárias para o efetivo cumprimento das leis eleitorais (art. 23, IX do Código Eleitoral [Brasil, 1965]⁶ e art. 57-J e 105 da

⁵ Ambos os argumentos aparecem nas contribuições ao item 29.i. A Coalizão Direitos na Rede e o Instituto Vero parecem se concentrar na ideia de legitimidade, enquanto o ITS Rio enfatiza a ideia de capacidade institucional.

⁶ Esse artigo prevê que expedir instruções que julgar convenientes para a execução do Código Eleitoral é competência do TSE.

Lei das Eleições [Brasil, 1997]⁷). Nos últimos anos, os desafios trazidos pelo problema da desinformação fizeram com que a atuação do TSE fosse ainda mais contundente, principalmente através de suas resoluções.

Considerando a relevância dos posicionamentos do TSE, foi realizada pesquisa de jurisprudência no mês de outubro de 2024 utilizando os termos-chave “fatos (notoriamente) inverídicos” e “fatos (gravemente) descontextualizados”⁸. Foram selecionadas as decisões mais recentes, apresentadas nas primeiras páginas dos resultados da pesquisa, a fim de garantir a análise das tendências jurisprudenciais mais atuais. A pesquisa foi complementada por informações extraídas do Sistema de Alertas de Desinformação Eleitoral (SIADÉ, s.d.).

A partir da análise da jurisprudência do TSE, é possível evidenciar uma preocupação do Tribunal com os efeitos da desinformação em diversas frentes, como a credibilidade das urnas eletrônicas, a confiança na Justiça Eleitoral e a honra de candidatos. A Representação n. 0601365-65.2022 (Brasil, 2024b), por exemplo, abordou a disseminação de boatos sobre uma suposta manipulação das urnas eletrônicas, acompanhados de alegações infundadas envolvendo o Tribunal Regional Eleitoral (TRE) de São Paulo. Mesmo após esclarecimentos oficiais, as informações descontextualizadas continuaram a circular, configurando um caso clássico de manipulação intencional para induzir o público ao erro. A decisão destacou que tal prática compromete a legitimidade do processo eleitoral e ilustra como a desinformação é frequentemente utilizada para atacar instituições democráticas.

Além disso, o TSE tem sido enfático ao tratar da manipulação de conteúdos audiovisuais como uma estratégia recorrente de desinformação. No caso da Representação n. 0601200-18.2022 (Brasil, 2022c), o Tribunal analisou vídeos que distorciam falas de candidatos ao retirar declarações de seu contexto original, criando, assim, uma narrativa artificial. Esse tipo de desinformação prejudica os candidatos diretamente envolvidos e subverte o debate público. O TSE concluiu no caso que, por alterarem o sentido original das mensagens, esses conteúdos violavam os princípios fundamentais da ética eleitoral e do processo democrático.

A conexão entre “desinformação” e “intenção deliberada de engano” é outro aspecto central. A Representação n. 0601556-13.2022 (Brasil, 2024c) tratou da circulação de informações que sugeriam vínculos inexistentes entre partidos políticos e organizações criminosas. O Tribunal considerou que as publicações tinham o claro propósito de enganar o eleitorado, utilizando elementos descontextualizados para criar uma narrativa falsa.

O conceito de desinformação como fenômeno multifacetado também aparece na Representação n. 0600920-47.2022 (Brasil, 2024a), que analisou a disseminação de informações desmentidas previamente por agências de checagem de fatos. O Tribunal destacou que a insistência em veicular conteúdos falsos, mesmo após sua refutação, configura uma tentativa deliberada de manipular o eleitorado. Logo, essa decisão enfatiza a importância de parcerias entre a Justiça Eleitoral e as agências de checagem para identificar rapidamente quais conteúdos são desinformativos e mitigar seus efeitos antes que eles atinjam proporções irreversíveis.

⁷ Segundo o art. 57-J, o TSE pode regulamentar a questão da propaganda eleitoral de acordo com o cenário e as ferramentas tecnológicas existentes em cada momento eleitoral, de modo a promover regras de boas práticas para as campanhas eleitorais na Internet. Por sua vez, o art. 105 trata da possibilidade que o TSE tem de expedir instruções para a fiel execução da Lei das Eleições (Brasil, 1997).

⁸ As pesquisas foram realizadas em: <https://jurisprudencia.tse.jus.br/#/jurisprudencia/pesquisa>. Acesso em 12 maio 2025.

A análise da jurisprudência também evidenciou uma tentativa de o TSE delimitar os critérios que orientam a identificação da desinformação. O uso de elementos verdadeiros apresentados de forma gravemente descontextualizada é um dos pontos mais complexos desse processo. Um exemplo claro está na Representação n. 0601185-49.2022 (Brasil, 2022b), na qual foram analisadas afirmações que sugeriam, de forma descontextualizada, que Ciro Gomes estaria apoiando a candidatura de Jair Bolsonaro. A decisão sublinhou que a descontextualização de fatos ou declarações pode induzir o eleitor ao erro, configurando uma grave ameaça à legitimidade do pleito.

Outro aspecto relevante na abordagem do TSE é a proteção das instituições eleitorais contra ataques desinformativos. A Representação n. 0600956-89.2022 (Brasil, 2022a), por exemplo, mostra como conteúdos falsos ou manipulados podem ser utilizados para desacreditar a atuação do TSE e de seus integrantes. Nesse caso, o TSE analisou publicações que afirmavam, de forma inverídica, que o candidato Jair Bolsonaro e seu partido provocaram a suspensão da lei que instituiu o piso salarial nacional para profissionais de enfermagem, apesar de essa suspensão ter decorrido de decisão cautelar do Superior Tribunal Federal (STF). Diante disso, o TSE reconheceu que tal prática viola a legislação eleitoral e compromete a confiança pública nas instituições democráticas.

O TSE também tem buscado estabelecer parâmetros claros para lidar com conteúdos que utilizam elementos factuais para criar narrativas enganosas. Um exemplo emblemático é a Representação n. 0601752-80.2022 (Brasil, 2023), na qual o Tribunal analisou postagens que insinuavam, sem provas, que um candidato defendia posições políticas polêmicas. Assim, o TSE destacou que a manipulação de informações verdadeiras para fins desinformativos é prática especialmente perigosa, pois dificulta a identificação da falsidade pelo público.

Além disso, o TSE tem enfatizado, em suas decisões, duas categorias principais de desinformação: “fatos notoriamente inverídicos” e “fatos gravemente descontextualizados”. Os “fatos notoriamente inverídicos” referem-se a informações manifestamente falsas e objetivamente verificáveis, como alegações infundadas de fraudes eleitorais ou manipulação de resultados, situações consideradas particularmente graves quando têm o potencial de comprometer a confiança no sistema eleitoral ou induzir o eleitorado ao erro. Já os “fatos gravemente descontextualizados”, embora contenham elementos de verdade, são apresentados de maneira a distorcer seu significado original e frequentemente envolvem o uso de edições, cortes ou manipulações que alteram o contexto factual ou temporal das informações, criando uma narrativa falsa. Portanto, as decisões analisadas indicam que o TSE tem buscado enfrentar os desafios impostos pela disseminação da desinformação em ambiente digital com base em alguns critérios: i) a intencionalidade do conteúdo, ii) o impacto potencial sobre o eleitorado e iii) a verificação por agências especializadas para subsidiar suas decisões.

Essa intencionalidade é geralmente evidenciada pela análise contextual, ou seja, a forma como os conteúdos são disseminados, frequentemente com o objetivo deliberado de enganar ou influenciar o eleitorado. O Tribunal reconhece, portanto, que a desinformação, em sua essência, busca minar a confiança do público no processo eleitoral e nas instituições democráticas, o que se reflete nas decisões sobre boatos envolvendo as urnas eletrônicas ou a Justiça Eleitoral.

Sobre o impacto da desinformação no equilíbrio do pleito, o TSE tem considerado não apenas o conteúdo das mensagens, mas também sua capacidade de afetar a

percepção pública e influenciar o resultado eleitoral. A gravidade dos casos intensifica-se quando as informações desinformativas são amplamente disseminadas, especialmente num ambiente digital caracterizado pela velocidade e alcance das redes sociais.

Em suma, nota-se que a análise da jurisprudência do TSE revela um esforço contínuo para lidar com os desafios impostos pela desinformação no contexto eleitoral digital. O Tribunal tem consolidado uma abordagem que combina rigor jurídico, colaboração interinstitucional e foco na preservação da legitimidade democrática, permitindo ao Tribunal avançar na construção de uma jurisprudência que busca coibir práticas nocivas sem comprometer direitos fundamentais, como a liberdade de expressão.

No entanto, os casos analisados também apontam para a necessidade de aprimoramento contínuo, especialmente no que diz respeito à uniformidade dos critérios aplicados. Assim, a subjetividade inerente a conceitos como “fatos gravemente descontextualizados” ainda gera desafios, como apontado em decisões críticas que destacam a dificuldade de aplicar esses critérios de forma uniforme.

3 Medidas de mitigação do risco de desinformação eleitoral: análise da consulta pública e das medidas adotadas pelo TSE

A análise realizada até o momento demonstra a importância de compreender melhor a definição e os contornos dos conceitos relacionados à desinformação. Agora, passa-se a analisar as contribuições feitas na consulta pública sobre os riscos (incluindo a desinformação) e suas formas de mitigação, essencialmente no contexto eleitoral. Objetiva-se compreender se há inovações e avanços nas medidas sugeridas ou se essas sugestões já estão incorporadas no ordenamento jurídico brasileiro e na atuação do TSE.

Como observado, há nas manifestações analisadas um consenso sobre a desinformação ser um risco a ser abordado na regulação das plataformas digitais, assim como relevante concordância sobre a utilidade de obrigações mais abrangentes para as plataformas em períodos eleitorais (NIC.br, 2023). A consulta apresenta perguntas que sugerem formas de mitigação dos riscos associados a infodemias, como a desinformação, ou seja, sugestões e respostas que são analisadas (perguntas 29, 30, 33 e 34).

A primeira sugestão, na pergunta 29.i, é a de definir categorias ou tipos de conteúdos que devem ser objeto de remoção obrigatória pelas plataformas digitais por simples notificação, além dos casos previstos no Marco Civil da Internet (MCI) (Brasil, 2014). Sobre a responsabilidade por danos decorrentes de conteúdo gerado por terceiros, o art. 19 do MCI traz a regra de que o provedor somente será responsabilizado civilmente pelos danos quando não tomar as providências para tornar o conteúdo indisponível após a ordem judicial específica. As exceções estão no art. 21 (imagens de nudez ou ato sexual) e art. 19, § 2º (direito autoral).

Há uma preocupação nítida nas manifestações em manter o que está estabelecido no MCI, com a remoção obrigatória apenas após a notificação judicial: caso seja realizada pela mera notificação é considerada desproporcional e ineficiente (Instituto Vero, ITS Rio e CEPI FGV Direito SP). Assim, a remoção obrigatória pelas plataformas de alguns tipos de conteúdos deve estar vinculada a danos graves e imediatos a direitos individuais ou coletivos (Diracom). Além disso, tais categorias deveriam ser precisas e detalhadas (ALAI).

Considera-se que essa sugestão já está implementada pela Resolução n. 23.610 do TSE (Brasil, 2019), que traz, em seu art. 9º-E, casos expressos de indisponibilização imediata de conteúdos, sob pena de responsabilidade solidária. São os casos de riscos considerados mais graves, quais sejam: atos antidemocráticos, fatos notoriamente inverídicos ou gravemente descontextualizados que atinjam a integridade do processo eleitoral, grave ameaça aos membros ou ao Poder Judiciário, e discurso de ódio e conteúdo sintético fabricado ou manipulado em desacordo com as normas.

A segunda sugestão (29.ii) é estabelecer responsabilidade das plataformas digitais pelos conteúdos impulsionados e monetizados. Nesse caso, percebeu-se uma maior aceitação das opiniões manifestadas sobre a adequação dessa medida para a mitigação dos riscos. Segundo a Coalizão Direitos na Rede, as plataformas já realizaram análise sobre os conteúdos e uma escolha em lucrar a partir deles.

Algumas participações mencionam que essa medida já está prevista, em alguma medida, no ordenamento jurídico brasileiro. Para o Idec, o Código de Defesa do Consumidor garante essa responsabilização objetiva, ao passo que para o IP.rec (Instituto de Pesquisa em Direito e Tecnologia do Recife), a base jurídica é o regime de responsabilidade subjetiva de plataformas do MCI. Há também algumas manifestações contrárias a essa sugestão, a exemplo do ITS Rio e da ALAI.

O TSE traz inúmeros requisitos para o impulsionamento de conteúdo político-eleitoral, seja na Lei das Eleições (Brasil, 1997) ou na Resolução 23.610 (Brasil, 2019). Segundo o art. 29, § 4º da Resolução, poderá haver responsabilização do provedor, nos moldes previstos no art. 19 do MCI (Brasil, 2014). Ademais, a priorização paga de conteúdos em aplicações de busca na Internet também possui vários requisitos (art. 28, § 7º-B [Brasil, 2019]).

A terceira medida, na pergunta 29.iii, é a definição de critérios e mecanismos para melhorar a exposição de usuários à diversidade de conteúdos, incluindo posições políticas e abordagens científicas; a partir disso, seria possível aprimorar a relação dos usuários com a diferença e a diversidade. O CTS FGV Direito Rio menciona que seria válido dar ao usuário uma opção de escolha, para que outro ponto de vista político, filosófico ou religioso seja exibido a ele. O Intervezes e o Instituto Vero também se manifestam sobre a questão da transparência nos Termos de Uso e nos parâmetros utilizados na recomendação de conteúdo, a fim de permitir ao usuário uma exposição livre dos conteúdos.

Apesar de o TSE não ter uma normativa específica sobre multiplicidade de opiniões e conteúdos, pode-se considerar que a atuação da Justiça Eleitoral brasileira, de modo geral, busca a exposição múltipla e igualitária dos candidatos durante o período eleitoral. Justamente por isso são impostas medidas contra o abuso de poder econômico e normas sobre o que é proibido ou permitido nas propagandas eleitorais na Internet. Consideramos que, embora não haja um arcabouço normativo robusto em relação a essa medida no Brasil, ela está de acordo com os objetivos da regulação eleitoral. Contudo, seria interessante que a regulação sobre regulação de plataformas analisasse o oferecimento de conteúdo múltiplo aos usuários, inclusive pelas recomendações algorítmicas.

A pergunta 33 menciona sobre as medidas de mitigação dos riscos associados a ameaças a processos eleitorais. O Instituto Vero manifesta-se sobre a necessidade de a entidade reguladora atuar em cooperação com o Tribunal Superior Eleitoral e as regras eleitorais. São medidas citadas: “i) plataformas devem apresentar, em tempo hábil, as regras e políticas que serão aplicadas em períodos eleitorais; ii) plataformas devem implementar sistema que possibilite a prestação de contas eleitorais, como

gastos com publicidade, autoria etc.; iii) biblioteca de anúncios específica para esses períodos; iv) proibição de microdirecionamento ou utilização de critérios de direcionamento que inibam o acesso à informação por todos”. Tais medidas são englobadas por uma ideia de maior transparência na relação entre plataformas e usuários no período eleitoral; no entanto, sobre a terceira, menciona-se que o TSE conta com um repositório de decisões sobre o enfrentamento à desinformação eleitoral, de acordo com o art. 9º-G da Resolução 23.610 (Brasil, 2019). Além disso, o art. 27-A da resolução exige um repositório de anúncios de impulsionamento político-eleitoral.

Na pergunta 33, a consulta sugere também cinco medidas de mitigação sobre as ameaças aos processos eleitorais. A primeira é sobre o estabelecimento de obrigações de transparência mais abrangentes em períodos eleitorais, a qual se harmoniza com as obrigações exigidas das plataformas sob o MCI (Brasil, 2014) e o direito do consumidor. O TSE cumpre essa ideia, ao menos parcialmente, em suas Resoluções. Sobre propaganda eleitoral, a Resolução 23.610 (Brasil, 2019) exige que haja identificação clara do impulsionamento de conteúdos na Internet (art. 29) e da proibição de disparo em massa (art. 31, §1º-A), dentre outras medidas.

A segunda sugestão é sobre a definição de responsabilidades mais amplas para moderação de conteúdos em períodos eleitorais; em geral, também há concordância sobre essa medida. A Coalizão Direitos na Rede menciona a necessidade de cooperação e diálogo com a Justiça Eleitoral. Essa sugestão também é cumprida pelo TSE, em grande parte, na Resolução 23.610 (Brasil, 2019) e na Resolução 23.714 (Brasil, 2022), que trazem previsões sobre exclusão de conteúdo indevido no período eleitoral, para preservar a integridade das eleições.

A terceira sugestão (pergunta 33.iii) é a definição de critérios sobre o limite de gastos em publicidade nas plataformas digitais durante os períodos eleitorais. A Lei das Eleições (Brasil, 1997) traz inúmeras previsões sobre os recursos nas campanhas eleitorais a partir de seu art. 17, enquanto o art. 10 da Resolução 23.610 (Brasil, 2019) também traz requisitos específicos para a publicidade eleitoral. Além disso, há uma necessidade de transparência do provedor sobre esse impulsionamento, segundo o art. 27-A da mesma Resolução.

A quarta medida é a limitação do perfilhamento para uso de propaganda eleitoral, com a qual as manifestações também são de concordância. Apesar de o TSE não vedar, traz requisitos de publicidade e acesso público às informações de perfilhamento eleitoral (art. 27-A, I; 33-B, I; 33-D, §1º, II).

Por fim, a última medida (pergunta 33.v) é o impedimento de impulsionamento de conteúdos durante a campanha eleitoral. No geral, as manifestações concordam com essa sugestão; entretanto, alguns consideram que esse impedimento pode ser uma medida excessivamente restritiva. Como afirmado, apesar de possível, o impulsionamento político-eleitoral é restrito pelas regras do TSE.

A perguntas 30 e 34 permitem que os participantes respondam com outras medidas de mitigação ainda não mencionadas. Em seguida, analisamos em conjunto algumas dessas sugestões. Para o IRIS (Instituto de Referência em Internet e Sociedade), são necessárias, além da moderação de conteúdo, curadoria para aqueles de caráter indefinido, mas não recomendáveis, ou seja, quando há dúvida sobre sua veracidade. Além disso, são sugeridos canais de comunicação eficientes para que os usuários façam denúncias.

Existem, atualmente, algumas iniciativas do TSE que implementam, ainda que parcialmente, essas sugestões. Sobre a exclusão de conteúdos ilegais, menciona-se as inúmeras previsões do TSE nesse sentido, a exemplo da Resolução 23.610 (art. 7º, *caput* e §1º; art. 9º-B, §4º; art. 9º-E; art. 9º-F, §3º; art. 38, §§4º e 5º [Brasil, 2019]) e Resolução 23.714 (art. 2º, §1º e art. 4º [Brasil, 2022]). Sobre os canais de denúncia, menciona-se o SIADÉ criado pelo TSE⁹, o disque-denúncia com o número 1491¹⁰ e o aplicativo Parda¹¹.

Para Alex Camacho e CTS FGV Direito Rio, outras medidas são importantes, como a educação cívica, a promoção de alfabetização midiática e digital (com habilidades críticas para que os indivíduos consigam identificar as informações e fontes confiáveis) e as parcerias com verificadores de fatos (para identificação de informações falsas e desinformativas com base em evidências) e veículos de mídia tradicionais. Ademais, o Tribunal Superior Eleitoral também tem atuado em iniciativas dessa natureza. Um exemplo é a campanha informativa “Jornalismo é confiável, fala nossa língua, protege da desinformação e fortalece a democracia”¹², que busca aumentar a confiança na mídia durante as eleições. Além disso, o Programa de Enfrentamento à Desinformação¹³ tem parcerias com empresas de checagem de fatos.

Nesse sentido, a partir da investigação realizada no presente item, observa-se que muitas sugestões mencionadas na consulta pública já estão, no contexto eleitoral, endereçadas pelo TSE, medidas apresentadas principalmente na Resolução 23.610 (Brasil, 2019), atualizada em 2024 pela Resolução 23.732 (Brasil, 2024X), que traz previsões relevantes sobre a desinformação eleitoral e a atuação das plataformas nesses casos.

Conclusão

O problema da desinformação é uma realidade imposta nos últimos anos. Compreender o conceito de desinformação, especialmente no âmbito eleitoral, embora seja um desafio bastante complexo, é essencial para lidar melhor com esse problema. Existe amplo consenso sobre a necessidade de combater a desinformação devido aos riscos que apresenta à Democracia, consenso que não parece se traduzir na existência de um plano de ação coeso e operacionalizável. Nessa discussão, a veracidade, um fator essencial, tem uma série de problemas práticos em sua avaliação.

A jurisprudência do TSE evidencia a relevância do papel da corte na aplicação prática de conceitos relevantes para o enfrentamento da desinformação no contexto eleitoral. Casos paradigmáticos, como aqueles envolvendo manipulação de vídeos ou disseminação de boatos sobre a Justiça Eleitoral, refletem uma preocupação do Tribunal com os efeitos da desinformação na formação de uma opinião pública informada e no fortalecimento da democracia. As decisões analisadas apontam que o

⁹ Disponível em: <https://www.tse.jus.br/eleicoes/sistema-de-alertas>.

¹⁰ Disponível em: <https://www.tse.jus.br/comunicacao/noticias/2024/Agosto/sos-voto-cidadao-pode-denunciar-mentiras-sobre-o-processo-eleitoral-pelo-1491-1>.

¹¹ Disponível em: <https://www.tse.jus.br/eleicoes/eleicoes-2018/aplicativos-da-justica-eleitoral/pardal>.

¹² Disponível em: <https://www.tse.jus.br/comunicacao/noticias/2024/Agosto/tse-e-aner-lancam-campanha-informativa-para-auxiliar-no-enfrentamento-das-mentiras-nas-eleicoes>.

¹³ Disponível em: <https://www.justicaeleitoral.jus.br/desinformacao/>.

TSE busca equilibrar a proteção ao debate público e a salvaguarda dos direitos fundamentais, com a necessidade de garantir a lisura do processo eleitoral. Contudo, também evidenciam os desafios inerentes à aplicação uniforme desses conceitos e ao combate das novas formas de desinformação em um ambiente digital dinâmico e em constante mudança.

Muitas sugestões apresentadas na consulta pública para lidar com a desinformação, como a imposição de responsabilidades mais amplas para as plataformas digitais durante o período eleitoral e a disponibilização de informações confiáveis aos eleitores, têm sido endereçadas pelo TSE. Assim, as contribuições da consulta pública trouxeram contribuições relevantes para enriquecer a regulação já existente, principalmente a partir de um maior nível de detalhamento a algumas questões.

No entanto, é necessário avançar com medidas mais amplas e efetivas não só por parte do TSE, mas da regulação de plataformas de modo geral, como o aperfeiçoamento dos conceitos e das medidas específicas impostas para as plataformas no período eleitoral, incluindo sua responsabilização. Somente com maior transparência e controle público da atuação das plataformas poderá haver avanço no tema da desinformação, sem que sejam comprometidos outros direitos igualmente relevantes em nosso ordenamento jurídico, como a liberdade de expressão e manifestação.

Referências:

BRASIL. Tribunal Superior Eleitoral. *Pardal*. Disponível em:

<https://www.tse.jus.br/eleicoes/eleicoes-2018/aplicativos-da-justica-eleitoral/pardal>.

Acesso em: 28 nov. 2024.

BRASIL. Tribunal Superior Eleitoral. *Programa Permanente de Enfrentamento à*

Desinformação. Disponível em: <https://www.justicaeleitoral.jus.br/desinformacao/>. Acesso

em: 28 nov. 2024.

BRASIL. Tribunal Superior Eleitoral. *SOS Voto: cidadão pode denunciar mentiras sobre o*

processo eleitoral pelo 1491. 28 ago. 2024. Disponível em:

<https://www.tse.jus.br/comunicacao/noticias/2024/Agosto/sos-voto-cidadao-pode-denunciar-mentiras-sobre-o-processo-eleitoral-pelo-1491-1>. Acesso em: 28 nov. 2024.

BRASIL. Tribunal Superior Eleitoral. *TSE e Aner lançam campanha informativa para auxiliar*

no enfrentamento das mentiras nas eleições. 06 ago. 2024. Disponível em:

<https://www.tse.jus.br/comunicacao/noticias/2024/Agosto/tse-e-aner-lancam-campanha-informativa-para-auxiliar-no-enfrentamento-das-mentiras-nas-eleicoes>. Acesso em: 28 nov. 2024.

BRASIL. Tribunal Superior Eleitoral. *Recurso na Representação nº 0600920-*

47.2022.6.00.0000. Acórdão Brasília/DF. Relatora Ministra Cármen Lúcia. Julgamento:

02/04/2024, 2024a.

BRASIL. Tribunal Superior Eleitoral. *Representação nº 0601365-65.2022.6.00.0000*. Acórdão

Brasília/DF. Relator Min. Benedito Gonçalves. Julgamento: 25/05/2024, 2024b.

BRASIL. Tribunal Superior Eleitoral. *Representação nº 0601556-13.2022.6.00.0000*. Acórdão

Brasília/DF. Relator Min. Benedito Gonçalves. Julgamento: 25/05/2024, 2024c.

BRASIL. Tribunal Superior Eleitoral. *Resolução nº 23.714, de 20 de outubro de 2022*. Disponível em: <https://www.tse.jus.br/legislacao/compilada/res/2022/resolucao-no-23-714-de-20-de-outubro-de-2022>. Acesso em: 18 nov. 2024.

BRASIL. Tribunal Superior Eleitoral. *Referendo na Representação nº 0600956-89.2022.6.00.0000*. Acórdão Brasília/DF. Relatora Min. Paulo de Tarso Sanseverino. Julgamento: 22/09/2022, 2022a.

BRASIL. Tribunal Superior Eleitoral. *Referendo na Representação nº 0601185-49.2022.6.00.0000*. Acórdão Brasília/DF. Relatora Min. Maria Cláudia Bucchianeri. Julgamento: 30/09/2022, 2022b.

BRASIL. Tribunal Superior Eleitoral. *Referendo na Representação nº 0601200-18.2022.6.00.0000*. Acórdão Brasília/DF. Relator Min. Paulo de Tarso Sanseverino. Julgamento: 03/09/2022, 2022c.

BRASIL. Tribunal Superior Eleitoral. *Representação nº 0601752-80.2022.6.00.0000*. Acórdão Brasília/DF. Relator Min. Benedito Gonçalves. Julgamento: 09/11/2023.

BRASIL. Tribunal Superior Eleitoral. *Resolução nº 23.610, de 18 de dezembro de 2019*. Disponível em: <https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019>. Acesso em: 18 nov. 2024.

BRASIL. Tribunal Superior Eleitoral. *Sistema de Alertas de Desinformação Eleitoral (SIADE)*. Disponível em: <https://www.tse.jus.br/eleicoes/sistema-de-alertas>. Acesso em: 28 nov. 2024.

CAVALIERE, Paolo. *The Truth in Fake News: How Disinformation Laws Are Reframing the Concepts of Truth and Accuracy on Digital Platforms*. Edinburgh School of Law Research Paper n. 2022/12. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4151908. Acesso em: 20 out. 2024.

CGI.br. *Consulta sobre Regulação de Plataformas Digitais: contribuições recebidas*. 2023. Disponível em: <https://dialogos.cgi.br/documentos/debate/consulta-plataformas/>. Acesso em: 15 out. 2024.

CORTE IDH. *Caso Tristán Donoso vs. Panamá*. Sentença de 27 de janeiro de 2009 (Exceção Preliminar, Mérito, Reparações e Custas), série C, n. 193.

DOUEK, Evelyn. *Verified Accountability: Self-Regulation of Content Moderation as an Answer to the Special Problems of Speech Regulation*. Aegis: Security Policy in Depth, set. 2019, Series Paper No. 1903.

LAYCOCK, Douglas. *The death of the irreparable injury rule*. Harvard Law Review, vol. 103, n. 3, 1990, p. 687-771. Disponível em: <https://www.jstor.org/stable/1341345>. Acesso em: 20 out. 2024.

LESSIG, Lawrence. *The Law of the Horse: What Cyberlaw Might Teach*. Harvard Law Review, vol. 113, 1999, p. 501-546. Disponível em: <https://cyber.harvard.edu/works/lessig/finalhls.pdf>. Acesso em: 20 out. 2024.

NIC.br. *Sistematização das contribuições à consulta sobre regulação de plataformas digitais*. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2023. Disponível em: https://cgi.br/media/docs/publicacoes/1/20240227162808/sistematizacao_consulta_regulacao_plataformas.pdf. Acesso em: 15 out. 2024.

SCOTT, Richard Mackenzie-Gray. *Managing Misinformation on Social Media: Targeted Newsfeed Interventions and Freedom of Thought*. *Northwestern Journal of Human Rights*, v. 21, n. 2, 2023.

TANDOC JÚNIOR, Edson C.; LIM, Z. W.; LING, R. *Defining "Fake News": A typology of scholarly definitions*. *Digital Journalism*, vol. 6, n. 2, 2018, p. 137-153. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/21670811.2017.1360143>. Acesso em: 20 out. 2024.

WARDLE, Claire; DERAKHSHAN, Hossein. *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Conselho da Europa, relatório DGI(2017)09. Disponível em: <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>. Acesso em: 20 out. 2024.

ZINGALES, Nicolo *et al.* *Codes of Conduct: a promising tool to counter online disinformation*. 2024a. Disponível em: <https://diretorio.fgv.br/sites/default/files/arquivos/policy-brief-cts-fgv-codigos-de-conduta.pdf>. Acesso em: 22 out. 2024.

ZINGALES, Nicolo *et al.* *TSE e Desinformação: Comentários sobre as Resoluções do TSE*. 2024b. Disponível em: <https://diretorio.fgv.br/sites/default/files/arquivos/cartilha-orientativa-tse-e-desinformacao-volume-2.pdf>. Acesso em: 20 out. 2024.

Atenção, dados e poder: uma análise crítica da infodemia e da desinformação

Luis Gonçalves, Leonardo Castro, Leandro Módolo

Resumo

A partir da leitura crítica da Sistematização das Contribuições à Consulta sobre Regulação de Plataforma (CGI.br, 2023), neste artigo argumentamos que o debate sobre infodemia e desinformação deve ser ampliado para além de suas implicações imediatas. Partimos da problematização dos conceitos de infodemia e desinformação, e noções associadas, tais como os sistemas de crenças, para reenquadrá-los em um contexto sócio-histórico mais amplo, em diálogo com os conceitos de economia da atenção e capitalismo de plataformas. Argumentamos que a infodemia e a desinformação são aspectos determinantes da economia digital contemporânea, e não simplesmente falhas de mercado. Assim, o texto explora como a infodemia e a desinformação se entrelaçam com o desenvolvimento de inteligência artificial (IA) e a produção massiva de dados digitais, configurando "rotas de suprimento" de dados para o treinamento de IA. Conclui-se que o enfrentamento a esses problemas deve considerar sua inserção em estruturas econômicas mais amplas do capitalismo de plataforma, sugerindo que abordagens limitadas às dimensões regulatórias ou técnicas podem ser insuficientes.

Palavras-chave: Infodemia. Desinformação. Economia da atenção. Capitalismo de plataforma. Inteligência artificial.

Introdução

Frequentemente, as discussões sobre infodemia e desinformação são delimitadas por seus efeitos sobre a saúde mental e coletiva, a política eleitoral e a liberdade de expressão, focalizando a necessidade de dispositivos regulatórios entre outros (CGI.br, 2020; 2023). Em que pese o conhecimento acumulado e as capacidades críticas possibilitadas por essas análises, muitas vezes esses fenômenos são abordados em sua *factualidade imediata*, destacados de processos e dinâmicas mais amplos dos quais eles participam, expressam e medeiam.

As dificuldades e os obstáculos a uma compreensão mais adequada dos reais impactos desses fenômenos por vezes nos convidam a atribuir sua causação às plataformas, aplicativos e algoritmos. Com isso, deixamos de considerar como o próprio funcionamento dessas tecnologias é limitado e atravessado estruturalmente por sistemas de crenças e outras formas de consciência, bem como por dispositivos de governamentalidade, o que torna ainda mais complexo seu entendimento.

Para contribuirmos com a compreensão dessas complexidades, combinamos uma leitura crítica da Consulta do CGI.br sobre regulação de plataformas digitais com uma breve revisão interdisciplinar das literaturas sobre desinformação, infodemia e economia política em busca de lacunas, abordagens e temas que possam ser ainda mais desenvolvidos. Assim, na primeira seção do artigo, iniciamos argumentando que, apesar do acúmulo sobre os conceitos de infodemia e desinformação, segue sendo desafiador um quadro consensual de suas características, causas e formas de enfrentamento, de modo que a ênfase nessas dificuldades possa nos ajudar a evitar abordagens simplistas, determinísticas e de viabilidade incerta.

Por outro lado, apesar de seu caráter multideterminado, na segunda seção, iniciamos a análise de que certos agentes e práticas infodêmicas e desinformativas refletem e organizam necessidades e estruturas capitalistas mais amplas que buscam se realizar por meio da infraestrutura desregulada das plataformas e aplicativos. De certa forma, tais

fenômenos existem no interior da chamada Economia da Atenção (Simon, 1971; Goldhaber, 2006), dependendo e valendo-se de suas leis para causarem seus efeitos.

Na terceira e na quarta seções, descrevemos como a economia da atenção, a infodemia e a desinformação inserem os internautas na cadeia industrial e global de produção e consumo de dados e de inteligência artificial (IA). Ao estimularem a produção de conteúdos e o engajamento, os ambientes informacionais infodêmicos e desinformativos atuam, também, como “rotas de suprimento” de dados (Zuboff, 2021) necessárias a essa nova economia.

Por fim, concluímos ressaltando que as discussões regulatórias sobre a infodemia e a desinformação podem tornar-se limitadas e adaptativas a essa realidade político-econômica mais ampla – apenas “enxugando gelo” – se não as reconhecermos como participantes de tais contextos. Longe desse aumento de complexidade se refletir em impotência, essa abordagem, na realidade, reconhece e conecta esses enfrentamentos aos conflitos e embates mais profundos do capitalismo de plataformas.

Problematizando a infodemia e a desinformação

Descrevemos introdutoriamente a “infodemia” como um neologismo que costuma ser associado a uma superabundância de informações mais ou menos verdadeiras, seja no ambiente informacional geral ou em um tema em específico, como no caso de eleições ou a exemplo da pandemia do coronavírus SARS-CoV-2 (CGI.br, 2023; Gonçalves *et al.*, 2024). Em todos os casos, infodemia denota um contexto atravessado e potencializado pelas tecnologias de informação e comunicação (TIC), no qual há mais informações sendo emitidas do que capacidades cognitivas e sociais dos receptores de absorvê-las e processá-las (Wardle, 2017; Carr, 2020). Implícito nisso está o problema da relevância e da credibilidade dessas informações pois, dada a superabundância de conteúdos, produtores, investidores, mediadores e consumidores, ainda não há soluções sociotécnicas ou regulatórias capazes de promover um ambiente comunicacional satisfatoriamente seguro com relação a esses aspectos (CGI.br, 2023).

Definir a desinformação também pode ser desafiador pois, em certos casos, o “conteúdo falso e enganoso é um conceito amorfo sem definições ou limites claramente estabelecidos, dificultando a comparação entre estudos” (OECD, 2024, p. 8), o que inclui a falta de consenso sobre os motivos para sua disseminação por parte dos internautas (Gonçalves *et al.*, 2024). Por outro lado, é suficiente destacarmos alguns pontos: i) a desinformação como uma informação verificável não factual com o potencial de causar danos individuais e coletivos (Avazz, 2020); ii) ela pode estar apenas atualizando na Internet os atores e os “padrões de manipulação da informação já verificados na imprensa, como a ocultação e a fragmentação de fatos” (Barreto, 2024, p. 340); iii) em certos casos, essa atualização pode sugerir uma disputa entre antigas e novas comunidades epistêmicas pelo poder de determinar a verdade (Oliveira, 2020); iv) nesses casos, a desinformação pode surgir como “uma estratégia racional que se desdobra em busca de objetivos políticos” (Lewandowsky *et al.*, 2017), v) por meio de um “aumento contínuo da complexidade e do tamanho de cadeias produtivas e redes de atores que surgiram, estimuladas por *altos investimentos financeiros* destinados a essas atividades” (CGI.br, 2023, p. 23, grifos nossos).

Entretanto, tal “dilúvio informacional” também emerge nutrido por outros fatores objetivos e subjetivos interconectados, por exemplo os desdobramentos da virada linguística e a radicalização, as crises e as disputas da acumulação capitalista em sua fase neoliberal (Anderson, 2022). No caso do neoliberalismo, ele (des)organiza nossa necessidade de depositar e inspirar confiança para mantermos e ampliarmos nossa conexão social (Boltanski; Chiapello, 2009), dentre outras formas, por meio da chamada *cultura do perigo* (Foucault, 2008), em que “os indivíduos são abordados como agentes morais que devem lidar com os riscos sociais e inseguranças” (Lemke, 2017, p. 7), justamente num momento histórico de crises e disputas epistêmicas (Cesarino, 2022).

Juntos, esses e outros fatores podem evocar a sensação de vulnerabilidade e desamparo como mediações para o governo e a condução de populações (Najar; Castro, 2021).

Observamos outro aspecto desses conflitos de governamentalidade nas dinâmicas dos sistemas de crenças. Por um lado, por serem estruturas psicossociais mais profundas, resilientes e estáveis (Eagleton, 2019), os sistemas de crenças podem mais orientar o comportamento do internauta em contextos desinformativos e infodêmicos do que o contrário. A literatura sugere que os indivíduos têm maior probabilidade de procurar fontes que se alinhem com sua visão de mundo, em vez de buscar informações que possam desafiá-las (Agle; Xiao, 2021; Bruns *et al.*, 2022; Walter; Tukachinsky, 2020), ainda que a exposição à desinformação num ambiente infodêmico possa criar certos “efeitos de influência contínua” que, ao longo do tempo, podem interferir nas crenças (Ecker *et al.*, 2022).

Por outro lado, é possível que a “indústria da desinformação” (CGI.br, 2023) não precise alterar os sistemas de crenças se puder instrumentalizá-los, por exemplo, por meio de mimeses inversas e suas bifurcações (Cesarino, 2022). Nesse caso, há uma dinâmica em que significantes são utilizados de maneira estratégica para criar novos significados, muitas vezes distorcendo ou subvertendo o sentido original. Assim, para certas pautas culturais, partidárias e político-econômicas, “vacina” pode virar “*vachina*” e a regulação das redes sociais pode se tornar uma ataque à liberdade de expressão e até à democracia. Ao lado da isso, ontologias políticas podem deslizar para “uma realidade duplamente distribuída: uma realidade oficial, mas superficial, e seu oposto, uma realidade mais profunda, oculta, ameaçadora, que era não oficial, mas muito mais real” (Boltanski, 2014).

Toda essa multiplicidade de fatores coloca em dúvida as ideias de algoritmos como implacavelmente persuasivos; logo, a falta de discernimento do público ou de informações “verdadeiras” seriam vetores suficientemente explicativos para a desinformação e a infodemia, e, ao mesmo tempo, chama a atenção para certas dimensões subjetivas do neoliberalismo. Ao se desdobrar numa “superprodução de elites” (Turchin, 2023) com um crescente poder político independente dos demais segmentos e de instituições em busca de sua autorrealização, o neoliberalismo pode estar atualizando seus diagramas de poder e governamentalidade para além dos paradigmas da democracia liberal. Essas tendências, por sua vez, poderiam convergir com as atuais dinâmicas infodêmicas e desinformativas, liberando, com isso, “formas epistêmicas ‘não modernas’ baseadas em temporalidades messiânicas e causalidades ocultas”, que seriam, “por assim dizer, a ‘superestrutura’ adequada ao capitalismo neoliberal contemporâneo” (Cesarino, 2020, p. xx).

Tais aspectos político-econômicos e psicossociais nos ajudam a considerar a multideterminabilidade dos contextos infodêmicos e desinformativos, evitando, assim, abordagens factuais, simples e exclusivas. Mas essas considerações não encerram nossa discussão; na verdade, esses aspectos coexistem com certas estruturas sociotécnicas e mercadológicas bastante objetivas, apontadas pela Consulta (CGI.br, 2023) que, como informado, embora não sejam determinísticas, buscam sistematicamente realizar suas necessidades e finalidades. Por isso, a seguir, discutiremos alguns desses elementos e suas implicações, o que podem contribuir com o enfrentamento à infodemia e à desinformação.

Economia da atenção, da infodemia e da desinformação

Parte dessas estruturas sociotécnicas e mercadológicas que podem contribuir com o engajamento de usuários, abordadas na Consulta, por exemplo, em suas relações com as recomendações algorítmicas, os efeitos de rede e a publicidade direcionada (CGI.br, 2023), podem ser aprofundadas a partir de um conceito construído por diferentes ângulos e objetivos: a *economia da atenção*.

Na psicologia e na neurociência, a atenção humana é uma capacidade orgânica e social de

foco, interferida pelas emoções e articulada durante suas atividades, sejam elas mais sensíveis ou mais teleológicas e introspectivas (Vygotsky, 2003; Damásio, 2012). Portanto, a atenção não pode ser uma *coisa externa* (tal qual alguns de seus estímulos), pois é uma capacidade do *ser*. Entretanto, considerando ser um problema de coordenação da cooperação social, a atenção tem sido ressignificada pela economia comportamental como um *recurso escasso* em relação à sua contraface e à informação superabundante (Simon, 1971; Goldhaber, 2006). Antes mesmo disso, com o surgimento do financiamento de veículos de comunicação por publicidade paga, no final do século XIX, a atenção humana já era alvo de pesquisas e disputas (Wu, 2017).

Desde então, essas abordagens têm estimulado esforços para a quantificação, a análise, a previsão e a predição – ou produção – desse "recurso" em escalas grandes e dirigidas (Smythe, 1977; Zuboff, 2021). Isso atribui a tal recurso uma crescente importância para as dinâmicas de poder no âmbito do Estado ampliado (Moraes, 2010) (caso da (des)informação) e da acumulação capitalista (caso da audiência publicitária) (O'Reilly; Strauss; Mazzucato, 2024). Chamamos a atenção para o fato de que tais esforços e interesses se medeiam sob uma *estrutura de mercado*, o que implica agentes econômicos relativamente independentes¹ concorrendo entre si pela produção e usufruto desse recurso como condição para a remuneração de seu trabalho ou de seu capital (CGI.br, 2023). Nesses termos – e em analogia com Polanyi (2013) –, a atenção pode assumir a materialidade social de uma *mercadoria fictícia*.

Atualmente, *Big techs* como Alphabet, Meta e ByteDance – e os conglomerados financeiros que as controlam (Ström, 2022) – têm concentrado e dirigido a economia da atenção, principalmente após a introdução das lógicas e capitais desse mercado na World Wide Web; destacadamente, primeiro com a exibição de publicidade paga no mecanismo de busca do Google em 2000 (Srnicsek, 2017) e, desde 2002, com as teorias e as tecnologias de leilão de segundo lance², que regulam automaticamente essa veiculação (Varian, 2007). Esses e outros fatores concorreram para que essas e outras empresas coproduzissem para si certos *efeitos de rede* (CGI.br, 2023), em que, por exemplo, 59,4% da população mundial usa suas mídias sociais, conectada cerca de seis horas por dia (Kemp, 2023) a um mercado de ofertas atencionais em que apenas algumas das principais empresas³ valiam juntas, em 2023, US\$ 10,6 trilhões (Fortune, 2024).

Na prática, isso significa que um agente social público – "empreendedor" (Antunes, 2018), ministro do Superior Tribunal Federal (STF), Brasil Paralelo, Ministério da Saúde ou Galinha Pintadina –, a fim de se tornar conhecido e relevante para exercer partes de suas potencialidades, precisa submeter-se e valer-se cada vez mais das obscuras leis da economia da atenção, expressas em métricas, como impressão em tela, *likes*, cliques, taxas de conversão etc. (Google, 2024).

Infodemia, desinformação e IA

Se as métricas descrevem os resultados das lutas pela atenção, geralmente é sob os *algoritmos de recomendação* que tal concorrência acontece (CGI.br, 2023). Devido à "escassez de atenção", essas IA são oferecidas para selecionar com relativa autonomia o que veremos a seguir em nossas telas, baseando-se, alegadamente, em nossas preferências

¹ Principalmente, trabalhadores e empresas de comunicação e tecnologia, *influencers*, fabricantes, provedores de aplicações e conteúdos (*Content and Application Providers – CAP*), provedores de serviços de Internet (*Internet Service Provider – ISP*) e outras mídias e veículos, anunciantes e investidores dentre outros (CGI.br, 2023), como abordaremos adiante.

² O leilão de segundo preço é aquele em que o vencedor paga um valor equivalente ao lance do concorrente imediatamente abaixo dele. Segundo seus teóricos, isso incentiva os anunciantes a fazerem lances mais próximos do valor real que atribuem à posição (Varian, 2007).

³ Apple, Alphabet, Meta, Amazon, Jingdong e Alibaba (Fortune, 2024).

(Chaney; Stewart; Engelhardt, 2018).

Este é o ponto em que a economia da atenção, a infodemia e a desinformação revelam tanto sua natureza entrelaçada, como também os outros agentes que constituem essa teia. Muito simplificadamente, a capacidade de essas IA selecionarem conteúdos em nosso lugar torna-se possível a partir de uma série de fatores, dos quais destacamos: i) a *extração de preferências* do usuário a partir da observação de seu comportamento e a introdução em sua programação (por parte de seus desenvolvedores⁴), ii) de *metas* (função-objetivo, funções de recompensas e de perda etc.⁵) e iii) de *meios específicos* para alcançá-las (aprendizado profundo ou por reforço etc.⁶) (Silva, 2009; Russell, 2022; Prince, 2023).

Entretanto, nestes casos, geralmente não é possível entendermos qual foi a *função de política*⁷ adotada pelas IA para uma tomada de decisão (Russell, 2022) (por que uma dancinha e não um gato ou uma notícia política (verificável ou não) apareceu na *timeline* de um usuário do Instagram?). Embora haja esforços corporativos para que esses e outros processos de automação prossigam obscurecidos sob inúmeras *camadas de abstração*⁸ (Luitse, 2024), hoje, sabemos que i) metas e recompensas de IA condensam tanto os objetivos dos usuários quanto os de desenvolvedores e seus outros clientes (Gonçalves, 2024) e ii) os desenvolvedores dão às IA autonomia para que elas desenvolvam políticas altamente abstratas e pragmaticamente independentes de valores sociais, que podem buscar influenciar as preferências do usuário de modo a facilitar sua satisfação, apenas por ser a condição para a maximização das recompensas do próprio agente computacional (Albanie; Shakespeare; Gunter, 2017; Russell, 2022).

Como o próprio campo da IA adverte, no limite, o objetivo do usuário final pode tornar-se um meio ou condição necessária, uma variável a ser regulada, para o alcance das metas imputadas no sistema pelos desenvolvedores, e não o contrário (Russell, 2022.; Viljoen *et al.*, 2021). Isso poderia ser realizado, por exemplo, se um algoritmo do TikTok recomendar o consumo de vídeos que, por suas diversas características, sejam capazes de influenciar a regulação da atividade de certas sub-regiões do cérebro⁹ de modo a desestimular o autocontrole e a noção de tempo, dentre outros (Su *et al.*, 2021).

⁴ Neste texto, utilizamos o termo "desenvolvedores" para nos referirmos principalmente a matemáticos, psicólogos, engenheiros, *designers*, analistas de marketing, economistas, administradores, advogados, comerciantes e investidores que, efetivamente, desenvolvem a maioria das ofertas da economia da atenção em seus vários e combinados aspectos.

⁵ Na computação, a função-objetivo é a formalização matemática de um objetivo social que orienta as decisões das IA. Geralmente, esse objetivo é maximizar ou minimizar alguma coisa, como lucros e custos. Sem o alcance (ainda que aproximado) da função-objetivo, a IA não obtém suas recompensas. De diferentes formas, essas IA são programadas para a obtenção de suas recompensas (pontuações numéricas) em primeiro lugar, o que, entretanto, está condicionado a entrega de objetivos humanossociais inseridos por seus desenvolvedores (Russell, 2022).

⁶ No aprendizado profundo (ou neuronal), as IA realizam seus cálculos dividindo-os e correlacionando-os em diferentes camadas que, posteriormente, se convergem numa única camada de saída – de um modo que os desenvolvedores ainda não sabem explicar satisfatoriamente. No aprendizado por reforço, as IA testam continuamente várias formas de alcançar um objetivo de forma a eventualmente conseguir maximizar suas recompensas ao longo do tempo (Russell, 2022).

⁷ Em IA, *função de política* é a estratégia ou o mapeamento do espaço de ações possíveis em um ambiente dinâmico e incerto para orientar as IA e condicionar suas ações *de modo que* seus efeitos tendam a maximizar sua recompensa no horizonte de longo prazo (Russell, 2022).

⁸ Camadas de abstração referem-se aos empacotamentos de "operações subjacentes complicadas em comandos únicos, 'escondendo', assim, muitas complexidades técnicas relacionadas à construção, operação e gerenciamento de sistemas computacionais", de modo que seus usuários não possam conhecer, "modificar ou reprogramar os elementos selecionados de acordo com suas preferências específicas" (Luitse, 2024).

⁹ Nesse caso, são os subcomponentes da rede de modo padrão (*Default Mode Network* – DMN), a área tegmentar ventral e as regiões discretas, incluindo pré-frontal lateral, tálamo anterior e cerebelo, e seu acoplamento aprimorado às vias visuais e auditivas (Su *et al.*, 2021).

Infodemia, desinformação e dados digitais

O advento de tecnologias, como as telecomunicações, as IA e o *big data*, é, também, consequência e mediação atual da permanente luta entre agentes econômicos, políticos e culturais por conhecimentos inéditos sobre o mundo natural e o social de forma que, a partir dessa vantagem, possam realizar, reproduzir ou substituir hegemonias, relações e modos de produção (Pinto, 2005; Faustino; Lippold, 2023). Desde os anos 1980, tal processo produz e é coproduzido por um crescente mercado com atores de diversos segmentos, portes e aportes que, em conjunto e concorrencialmente, vêm criando IAs cada vez mais poderosas, complexas e, para isso, mais dependentes de quantidades massivas de dados digitais (Cassiolato; Dantas; Lastres, 2024; Blackwell, 2024; Sun *et al.*, 2017).

Todavia, o ainda pouco compreendido *dado digital* não pré-existe no mundo como o petróleo, o minério ou a soja, que, para serem usados, precisam antes ser extraídos, minerados ou coletados: *o dado digital precisa antes ser produzido* em escala industrial (Srnicek, 2017; Sadowski, 2019). Entretanto, o fato de, após serem produzidos, os dados serem armazenados em repositórios mais ou menos acessíveis (sites, redes sociais etc.), de modo mais ou menos legal pelos desenvolvedores (CGI.br, 2023), costuma inspirá-los a uso de metáforas extrativistas que elidem a origem social desses produtos, facilitando sua apropriação e seu reuso (Zuboff, 2021). Grosso modo, as IA só podem ser desenvolvidas se puderem acessar esses dados estocados, ainda que essa etapa produtiva (o treinamento de modelos) seja apenas o início cíclico de todo o processo. Lançados no mundo, esses modelos precisam também de dados inéditos, isto é, de novas "entrada[s] significativa[s] do mundo real (uma frase, um arquivo de som, uma imagem etc.)" (Prince, 2023) para seguirem se ajustando (a extração de preferências) e, principalmente, produzirem seus resultados – padrões, inferências, predições etc. Justamente para que se ampliem suas “rotas de abastecimento” de dados inéditos (Zuboff, 2021), a Meta inseriu, sem opção de rejeição por parte dos usuários, sua Meta AI no aplicativo WhatsApp (WhatsApp, 2024; Kaspersky, 2024).

Sejam inéditos ou arquivados, geralmente essas entradas significativas do mundo real só podem ser resultados de algum tipo de interação humano-computador (IHC) (Russell, 2022) – seja um engenheiro mapeando um território ou uma célula, um policial procurando suspeitos periféricos usando visão computacional treinada por *crowdworkers* periféricos, ou bilhões de pessoas utilizando seus dispositivos para conversar, trabalhar, se divertir, enfim, para viver em relações sociais mediadas por plataformas (Srnicek, 2022).

Nesse sentido, se decisões rivais precisam ser tomadas em um ambiente incerto baseando-se em análises de IA que dependem de quantidades massivas de dados que, em muitos casos, só podem ser produzidos mediante a atividade humana, a própria atividade humana deve ser provocada. Em outras palavras, para toda essa cadeia de agentes e mercados, i) é imperioso que as pessoas sejam conduzidas e engajadas a interagir com certos dispositivos digitais, a fim de ii) produzirem dados sobre si e sobre essa interação, que iii) coproduzirão IA e, por sua vez, iv) farão inferências sobre saúde, comportamento, subjetividade, renda, crédito etc. de indivíduos e populações na forma de novos dados digitais (Conte *et al.*, 2012). Como resultado, essa *cadeia de produção e consumo de dados* acumulou, apenas em 2024, cerca de 174 zettabytes¹⁰ (Statista, 2024).

Desse modo, tal imperativo aponta, dentre outras fontes, para certas instâncias de produção abundante de dados digitais controladas pela economia da atenção, principalmente as redes sociais, plataformas de *streaming*, mecanismos de busca e aplicativos similares (Russell, 2013; Russell, 2022). Assim, se os algoritmos de curadoria das redes sociais fizerem as recomendações certas, os usuários podem interagir mais a partir de comentários polêmicos e emocionalmente engajadores (Horwitz, 2023) que

¹⁰ O equivalente a 147.000.000.000.000.000GB.

confirmem (ou sugiram) crenças (Altay; Araújo; Mercier, 2022), eventualmente conduzindo alguns deles para grupos radicalizados (Nascimento *et al.*, 2022) ou aumentando sua exposição pessoal em troca de *likes* no Instagram (Bentes, 2021), por exemplo, produzindo, dessa forma, parte dos dados necessitados pelos mercados de IA.

Conclusão

No estado atual do debate sobre a economia da atenção – à semelhança das IA baseadas em recompensas – seus diversos agentes econômicos recebem muitos incentivos e pouca regulação para produzirem audiência e interação, independente da relevância, da credibilidade e dos efeitos sociais de suas ofertas (Wu, 2017; CGI.br, 2023). Nesses termos, à primeira vista, a infodemia e a desinformação surgem como falhas de mercado (Arrow, 1963) que poderiam ser corrigidas e reguladas, por exemplo, a partir de instrumentos antitruste e da responsabilização *ex post* (por exemplo, remoções de conteúdos).

Entretanto, tais “falhas” perfazem oportunidades e estruturas essenciais para outros mercados e agentes político-culturais com os quais a economia da atenção se conecta.

Nesse sentido, uma Internet sem infodemia e desinformação poderia ser um ambiente informacional mais saudável, confiável e socialmente mais justo e próspero; porém, sua contenção ou eliminação levaria a uma certa redução de espaço para a circulação e a acumulação capitalista em nível global e igualmente a uma limitação crucial dos ciclos comerciais de desenvolvimento das IA realmente existentes, que precisamos compreender melhor. Aqui, referimo-nos, além do citado trilionário mercado de empresas líderes na economia da atenção, ao mercado de anunciantes que as sustentou em 2023 com investimentos de US\$ 546 bilhões (Statista, 2023), do ascendente mercado de IA avaliado em US\$ 184 bilhões (Statista, 2024) e do seletivo grupo de fundos de investimento que controla a maioria desses mercados, avaliado em US\$ 27 trilhões (Ström, 2022). Assim, nos termos desta discussão, o problema da infodemia e da desinformação é, em parte, um problema da economia política das TIC altamente financeirizada, de modo que a mitigação de um deles depende e reflete o grau de controle social sobre o outro. Isso, combinado ao contexto geopolítico de ascensão do bloco de poder representado por Donald Trump, ou às características da maioria parlamentar no Brasil, pode comprometer e limitar as já frágeis capacidades regulatórias governamentais (CGI.br, 2023), bem os discutíveis conceitos ambientais, sociais e de governança (*Environmental, Social and Governance* – ESG) aplicáveis nesses mercados (Introna, 2016; Viljoen, 2021). Sem essas dimensões de fundo, tais embates podem ser limitados a enquadramentos factualmente jurídicos (legal/ilegal), morais (maldade/altruísmo), sociotécnicos (checagem/correção humana ou algorítmica, remoção etc.) ou comportamentais (literacias) que, dessa forma, tenderão a ser adaptativas (e não disruptivas) a certas causas materialmente estruturais da infodemia e da desinformação.

Considerando as implicações sociais de tais ângulos da questão, abordagens relativas à soberania tecnológica nacional e à economia-política mais ampla emergem como outros elementos centrais a serem considerados nos embates regulatórios. Nesse sentido, a ausência de políticas para infraestruturas críticas reforça a vulnerabilidade dos países periféricos aos interesses geopolíticos e comerciais das empresas de plataformas e seus aliados (Rikap *et al.*, 2024a; Lastres; Cassiolato; Dantas, 2025). Portanto, como discutido na Consulta, em resposta, essas políticas podem envolver o estímulo para a inovação tecnológica nacional, a criação e o uso de *softwares* livres, protocolos abertos e infraestruturas públicas, o armazenamento e o tratamento local de dados, o uso estratégico das compras públicas, políticas claras para transferência internacional de dados, limites a aquisições de concorrentes por grandes plataformas, a desconcentração econômica, dentre outros (CGI.br, 2023). Não menos importante, esses e outros aspectos de políticas de soberania digital precisam também considerar suas implicações participativas, ambientais

e sociais (CGI.br, 2023; Rikap, 2024b; MTST, 2023; Sadowski; Viljoen; Whittaker, 2021; Abílio; Amorim; Grohmann, 2021).

Referências

- ABÍLIO, Ludmila Costhek; AMORIM, Henrique; GROHMANN, Rafael. Uberização e plataformação do trabalho no Brasil: conceitos, processos e formas. *Sociologias*, v. 23, n. 57, p. 26-56, 2021.
- ALBANIE, Samuel; SHAKESPEARE, Hillary; GUNTER, Tom. Unknowable manipulators: Social network curator algorithms. *arXiv preprint arXiv:1701.04895*, 2017.
- ALTAY, Sacha; DE ARAUJO, Emma; MERCIER, Hugo. "If this account is true, it is most enormously wonderful": Interestingness-if-true and the sharing of true and false news. *Digital Journalism*, v. 10, n. 3, p. 373-394, 2022.
- ANDERSON, Perry. *The H-word: The peripeteia of hegemony*. Verso Books, 2022.
- ANTUNES, Ricardo. *O Privilégio da servidão: o novo proletariado de serviços na era digital*. São Paulo: Boitempo Editorial, 2018.
- AVAZZ. O Brasil está sofrendo uma infodemia de Covid-19, 2020. Disponível em: https://secure.avaaz.org/campaign/po/brasil_infodemia_coronavirus/. Acesso em 21 set. 2024.
- ARROW, K. Uncertainty and the welfare economics of medical care. Em: K. Arrow, *Essays in Theory of Risk-bearing*. Amsterdam/Londres: North Holland, 1962.
- BAKIR, Vian. Psychological operations in digital political campaigns: Assessing Cambridge Analytica's psychographic profiling and targeting. *Frontiers in Communication*, v. 5, p. 67, 2020.
- BARRETO, Helena Martins do Rego. A desinformação em meio à crise do capitalismo e à configuração de uma nova estrutura de mediação social. *Revista Eco-Pós*, v. 27, n. 1, p. 330-352, 2024.
- BENTES, Anna. *Quase um tique: economia da atenção, vigilância e espetáculo em uma rede social*. Rio de Janeiro: Editora UFRJ, 2021.
- BLACKWELL, Alan F. *Moral Codes: Designing Alternatives to AI*. MIT Press, 2024.
- BOLTANSKI, Luc. *Mysteries and conspiracies: Detective stories, spy novels and the making of modern societies*. John Wiley & Sons, 2014.
- BOLTANSKI, Luc; CHIAPELLO, Ève. *O novo espírito do capitalismo*. São Paulo: WMF Martins Fontes, 2009.
- CARR, Nicholas. *The shallows: What the Internet is doing to our brains*. WW Norton & Company, 2020.
- CASSIOLATO, José Eduardo; DANTAS, Marcos; LASTRE, Helena M. M. Marco conceitual e analítico da Economia de Dados, 2024. Disponível: <https://www.ie.ufrj.br/images/IE/grupos/redesist/SITE/PROJETOS/22/NT11%20Cassiolato%20Dantas%20e%20Lastres%2011.pdf>. Acesso: 17 abr. 2025.
- CESARINO, Letícia. O fetichismo do QAnon. *Jacobin Brasil*, [s. l.], 2020. Disponível em: <https://jacobin.com.br/2020/11/o-fetichismo-do-qanon>. Acesso em: 21 dez. 2020.
- _____. *O mundo do avesso: verdade e política na era digital*. São Paulo: Ubu, 2022.
- CHANEY, Allison JB; STEWART, Brandon M.; ENGELHARDT, Barbara E. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In: *Proceedings of the 12th ACM conference on recommender systems*. 2018. p. 224-232. <https://doi.org/10.1145/3240323.3240370>.

COALIZÃO SOBERANIA DIGITAL DEMOCRÁTICA E ECOLÓGICA. Recuperando a soberania digital: um roteiro para construir um ecossistema digital para as pessoas e o planeta, 2024. Disponível em: <https://ppl-ai-file-upload.s3.amazonaws.com/web/direct-files/21923488/e356f4be-4341-48b3-ae80-589aaa9c8c2f/recuperando-a-soberania-digital-1.pdf>. Acesso em: 18 abr. 2025.

COMITÊ GESTOR DA Internet (CGI). Sistematização das contribuições à consulta sobre regulação de plataformas digitais. São Paulo: NIC.br|CGI.br, 2023. Disponível em: https://cgi.br/media/docs/publicacoes/1/20231213081034/sistematizacao_consulta_regulacao_plataformas.pdf. Acesso em: 10 mar. 2024.

_____. Relatório Internet, Desinformação e Democracia. São Paulo: NIC.br|CGI.br, 2020. Disponível em: https://www.cgi.br/media/docs/publicacoes/4/20200327181716/relatorio_Internet_desinformacao_e_democracia.pdf. Acesso em 1 out. 2024.

CONTE, Rosaria et al. Manifesto of computational social science. The European Physical Journal Special Topics, v. 214, n. 1, p. 325-346, 2012.

DAMÁSIO, António. O erro de Descartes: emoção, razão e o cérebro humano. Editora Companhia das Letras, 2012.

EAGLETON, Terry. Ideologia: uma introdução. São Paulo: Boitempo Editorial, 2019.

FAUSTINO, D.; LIPPOLD, W. Colonialismo digital: por uma crítica hacker-fanoniana. 1. Ed. São Paulo: Boitempo, 2023.

FORTUNE global 500. Fortune, [s. l.], 2024. Disponível em: <https://fortune.com/global500>. Acesso em: 24 jul. 2024.

FOUCAULT, M. Nascimento da Biopolítica - curso dado no College de France 1978-1979. São Paulo: Livraria Martins Fontes, 2008.

GOLDHABER, M. The value of openness in an attention economy. First Monday, v. 11, n. 6, 5 jun. 2006. Disponível em: <https://firstmonday.org/ojs/index.php/fm/article/download/1334/1254?inline=1>. Acesso em: 1 out. 2024.

GONÇALVES, Luis H. N. O problema dos três usuários – implicações entre psicologia, IA e economia política. In: Inteligência Artificial e Psicologia / Laura Cristina Eiras Coelho Soares (organizadora) – Curitiba : CRV, 2024. ISBN Digital 978-65-251-6819-7. DOI 10.24824/978652516818.0.

GONÇALVES, Luis; CASTRO, Leonardo; RACHID, Raquel; FORNAZIN, Marcelo. As múltiplas faces da Infodemia. Revista Eletrônica de Comunicação, Informação & Inovação em Saúde, [S. l.], v. 18, n. 3, p. 716–735, 2024. DOI: 10.29397/reciis.v18i3.3796. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/3796> . Acesso em: 1 out. 2024.

GOOGLE. Seu guia do Google Ads. Disponível em: <https://support.google.com/google-ads/answer/6146252>. Acesso: 21 nov. 2024.

GROHMANN, Rafael. Os laboratórios do trabalho digital: entrevistas. São Paulo: Boitempo Editorial, 2021.

HORWITZ, Jeff. The Facebook Files: A Wall Street Journal investigation. The Wall Street Journal. Disponível em: <https://www.wsj.com/articles/the-facebook-files-11631713039>. Acesso em: 15 out. 2024.

INTRONA, Lucas D. Algorithms, governance, and governmentality: On governing academic writing. Science, Technology, & Human Values, v. 41, n. 1, p. 17-49, 2016. DOI: <https://doi.org/10.1177/016224391558736>.

KASPERSKY. Meta AI plans to use the personal data of its users to train generative AI. Kaspersky Daily, 15 nov. 2024. Disponível em: <https://www.kaspersky.com/blog/meta-uses-personal-data/51548/>. Acesso em: 15 nov. 2024.

KEMP, Simon. Digital 2023: Global Overview Report. DataReportal, [s. l.], 23 Jan. 2023. Disponível em: <https://datareportal.com/reports/digital-2023-global-overview-report>. Acesso em: 18 set. 2024.

LEMKE, Thomas. Foucault, governamentalidade e crítica. Plural: Revista de Ciências Sociais, v. 24, n. 1, p. 194-213, 2017.

MÉSZÁROS, István. Para Além do Capital. São Paulo: Boitempo, 2002.

MORAES, Dênis de. Comunicação, hegemonia e contra-hegemonia: a contribuição teórica de Gramsci. Revista Debates, v. 4, n. 1, p. 54-54, 2010.

LASTRES, Helena M. M.; CASSIOLATO, José Eduardo; DANTAS, Marcos. Estado atual da conceituação e mensuração da Economia de Dados no Brasil. Nota Técnica 12. Rio de Janeiro: RedeSist/IE/UFRJ, 18 fev. 2024. Disponível em: https://www.researchgate.net/publication/387269688_Estado_atual_da_conceituacao_e_mensuracao_da_Economia_de_Dados_no_Brasil. Acesso em: 18 abr. 2025.

LÉVY, Pierre. Cibercultura. São Paulo: Editora 34, 1999.

LUITSE, Dieuwertje. "Platform power in AI: The evolution of cloud infrastructures in the political economy of artificial intelligence". Internet Policy Review 13.2., 2024. Disponível em: <https://doi.org/10.14763/2024.2.1768>. Acesso em: 5 set. 2024.

LURIA, Alexandre Romanovich. Curso de psicologia geral. Civilização Brasileira, 1979.

MCQUILLAN, Dan. People's councils for ethical machine learning. Social Media+ Society, v. 4, n. 2, 2056305118768303, 2018.

MOVIMENTO DOS TRABALHADORES SEM TETO (MTST). Cartilha MTSTec: tecnologia para a luta popular, 2023. Disponível em: <https://linktr.ee/tecnologia.mtst>. Acesso em: 18 abr. 2025.

OLIVEIRA, Thaiane. Desinformação científica em tempos de crise epistêmica: circulação de teorias da conspiração nas plataformas de mídias sociais. Revista Fronteiras, v. 22, n. 1, 2020. DOI: 10.4013/fem.2020.221.03

O'REILLY, Tim; STRAUSS, Ilan; MAZZUCATO, Mariana. Algorithmic Attention Rents: A theory of digital platform market power. Data & Policy, v. 6, p. e6, 2024. DOI:10.1017/dap.2024.1.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). The OECD truth quest survey: methodology and findings. Disponível em: https://www.oecd.org/en/publications/the-oecd-truth-quest-survey_92a94c0f-en.html. Acesso em: 10 jun. 2024.

POLANYI, Karl. A grande transformação: as origens da nossa época. Elsevier Brasil, 2013.

PRINCE, Simon JD. Understanding deep learning. MIT Press, 2023.

NAJAR, Alberto; CASTRO, Leonardo. Um nada 'admirável mundo novo': medo, risco e vulnerabilidade em tempos de covid-19. Saúde em Debate, Rio de Janeiro, v. 45, n. especial 2, p. 142-155, 2021. DOI: <https://doi.org/10.1590/0103-11042021E210>.

L. NASCIMENTO, L. CESARINO, P. FONSECA, T. BARRETO, V. MUSSA. Públicos refratados: grupos de extrema-direita brasileiros na plataforma Telegram. Internet e Sociedade. V. 3 / N. 1 / p. 31-60, 2022.

RIKAP, Cecilia; DURAND, Cédric; PARANÁ, Edemilson; GERBAUDO, Paolo; MARX, Paris. Reclaiming Digital Sovereignty: A roadmap to build a digital stack for people and the planet. London: University College London, Institute for Innovation and Public Purpose, 2024. Disponível em: <https://www.ucl.ac.uk/bartlett/sites/bartlett/files/reclaiming-digital-sovereignty.pdf>. Acesso em: 18 abr. 2025.

RUSSELL, Stuart. Human-Compatible Artificial Intelligence. Oxford: Oxford University Press, 2022. ISBN 0198862539

RUSSELL, Matthew A. Mining the social web: data mining Facebook, Twitter, LinkedIn, Google+, GitHub, and more. Sebastopol: O'Reilly Media, Inc., 2013. ISBN: 978-1-449-36761-9.

SADOWSKI, Jathan. When data is capital: Datafication, accumulation, and extraction. *Big data & society*, 2019, 6.1: 2053951718820549.

SADOWSKI, Jathan; VILJOEN, Salomé; WHITTAKER, Meredith. Everyone should decide how their digital data are used—not just tech companies. *Nature*, v. 595, n. 7866, p. 169-171, 2021. DOI: <https://doi.org/10.1038/d41586-021-01812-3>

SILVA, Valdinei Freire da. Extração de preferências por meio de avaliações de comportamentos observados. 2009. PhD Thesis. Universidade de São Paulo.

SIMON, Herbert. Designing organizations for an information-rich world. In: GREENBERGER, M. (Ed.). *Computers, communications, and the public interest*. Baltimore, MD: The Johns Hopkins Press, 1971.

SMYTHE, Dallas W. Communications: blindspot of western Marxism. *CTheory*, v. 1, n. 3, p. 1-27, 1977.

SRNICEK, Nick. Platform capitalism. John Wiley & Sons, 2017.

_____. Data, compute, labor. In: GRAHAM, Mark; FERRARI, Fabian (ed.). *Digital Work in the Planetary Market*. Cambridge: The MIT Press, 2022. p. 1-20. DOI: [10.7551/mitpress/13835.001.00012022](https://doi.org/10.7551/mitpress/13835.001.00012022).

STATISTA. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. Disponível em: <https://www.statista.com/statistics/871513/worldwide-data-created/>. Acesso em 15 out. 2024.

STRÖM, Timothy Erik. Capital and Cybernetics. *New Left Review*, 135. May/June, 2022. Disponível em: <https://newleftreview.org/issues/ii135/articles/timothy-erik-strom-capital-and-cybernetics>. Acesso em 2, out. 2024.

SU, Conghui et al. Viewing personalized video clips recommended by TikTok activates default mode network and ventral tegmental area. *NeuroImage*, v. 237, p. 118136, 2021. <https://doi.org/10.1016/j.neuroimage.2021.118136>.

SUN, Chen et al. Revisiting unreasonable effectiveness of data in deep learning era. In: *Proceedings of the IEEE international conference on computer vision*. 2017. p. 843-852.

VARIAN, Hal R. Position auctions. *International Journal of Industrial Organization*, v. 25, n. 6, p. 1163-1178, 2007.

VIEIRA PINTO, Álvaro. O conceito de tecnologia. Volume 1. Rio de Janeiro: Contraponto, 2005.

VIGOTSKI, L.S. *Psicologia Pedagógica* - Porto Alegre: Artmed, 2003.

VILJOEN, Salomé. A relational theory of data governance. *Yale LJ*, v. 131, p. 573, 2021. Disponível em: <http://www.jstor.org/stable/45400961>. Acesso: 18 abr. 2025.

VILJOEN, Salomé; GOLDENFEIN, Jake; MCGUIGAN, Lee. Design choices: Mechanism design and platform capitalism. *Big data & society*, v. 8, n. 2, 2021. DOI: <https://doi.org/10.1177/205395172111034312>.

WARDLE, Claire et al. Fake news. It's complicated. *First draft*, v. 16, p. 1-11, 2017. Disponível em: <https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79>. Acesso em 1/10/2023.

WHATSAPP. About AI experiences on WhatsApp. WhatsApp Help Center, 2024. Disponível em: <https://faq.whatsapp.com/1111412106858632>. Acesso em: 15 nov. 2024.

WU, Tim. *The attention merchants: The epic scramble to get inside our heads*. Vintage, 2017.

ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância*. Rio de Janeiro: Editora Intrínseca, 2021.

A questão negra na regulação das plataformas digitais: uma análise da consulta realizada pelo CGI.br

Marco Antônio Sousa Alves, Fernanda dos Santos Rodrigues Silva, Mariana Alves Araújo Lopes

Resumo

Em 2023, o Comitê Gestor da Internet obteve mais de 1.000 contribuições em consulta pública sobre regulação de plataformas digitais no Brasil. Considerando que um dos grupos mais vulnerabilizado e hostilizado online é justamente o de pessoas negras, especialmente mulheres negras, o presente artigo tem por finalidade analisar como surgiu a questão racial e de que modo foi endereçada no debate. Partiu-se do pressuposto de que o marco regulatório brasileiro sobre plataformas digitais precisa ser explicitamente antirracista para impedir a reprodução do racismo estrutural. O método de abordagem foi o indutivo, partindo do texto da consulta e contribuições dos consultados, e o método de procedimento, de análise de conteúdo. As principais conclusões indicam que, embora tenha ocorrido avanço na interseção entre regulação de plataforma e o debate racial, ainda há dificuldades. O consulente, ao formular as questões, foi genérico, sem focar diretamente o racismo. Nas contribuições, sugeriu-se uma regulação que garanta uma moderação de conteúdo mais inclusiva, auditorias algorítmicas, relatórios sobre discriminação racial e a criação de uma autoridade regulatória com composição diversificada, incluindo representação racial, dentre outros. Apesar disso, o total de menções à questão racial foi baixo em relação ao número total de contribuições.

Palavras-chave: Antirracismo. Consulta CGI.br. Moderação de Conteúdo. Plataformas Digitais. Regulação.

Introdução

A necessidade e a urgência de regular plataformas digitais é algo cada vez mais consensual, sendo, inclusive, pauta do documento aprovado unanimemente por 193 países-membros na última Assembleia Geral da ONU, ocorrida em setembro de 2024, intitulada “Pacto para o futuro” (*Pact for the future*) (UN, 2024). Sistemáticas e graves violações de direitos fundamentais, discursos de ódio, exploração sexual infantojuvenil, racismo, desinformação, ações terroristas e ameaças à democracia são alguns dos problemas encontrados. Embora a identificação da pauta mais relevante e a escolha dos meios de enfrentamento estejam longe de ser algo pacificado, a percepção de não mais ser possível deixar as grandes plataformas digitais simplesmente desreguladas ou apenas autorreguladas tende a prevalecer contemporaneamente.

No Brasil, esse debate também está vivo e tem grande projeção. Os esforços para regulação do espaço digital intensificaram-se desde 2009, quando foi iniciada uma consulta pública para redação do projeto do que, em 2014, se tornou o Marco Civil da Internet (Lei 12.965) (Brasil, 2014). Embora esse marco faça um esforço na atribuição de responsabilidades, dentre outras, aos provedores de conteúdo, ele se mostrou insuficiente para dar conta da complexidade de relações, serviços e impactos das plataformas digitais na sociedade brasileira. Diante dessa lacuna, tem-se como principal iniciativa legislativa o Projeto de Lei (PL) 2.630, de 2020, que institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet e tramita atualmente na Câmara de Deputados (Brasil, 2020).

Procurando ampliar esse debate e contribuir para a construção consensual de um marco regulatório brasileiro sobre plataformas digitais, o Grupo de Trabalho de Regulação de Plataformas do Comitê Gestor da Internet do Brasil (CGI.br) promoveu, dentre outros eventos nos últimos anos, uma consulta pública realizada entre abril e julho de 2023. Com

mais de 1,3 mil contribuições recebidas, a consulta apresenta um amplo mapeamento de questões em jogo no debate brasileiro sobre a questão e oferece propostas relevantes para o endereçamento das medidas regulatórias.

O presente trabalho pretende, a partir da análise dessa consulta, refletir sobre como a questão racial negra aparece e é endereçada nesse recorte do debate brasileiro acerca da regulação das plataformas digitais. Embora o Brasil tenha experimentado mais de três séculos de regime escravocrata, sendo fortemente marcado pelo racismo (Moura, 2019; Nascimento, 2016), especialmente contra pessoas negras¹, é também um país com grande dificuldade de abordar diretamente essa questão e enfrentá-la adequadamente (Gonzalez, 1988). Essa dificuldade aparece também no debate sobre a regulação das plataformas digitais, no qual se insere a consulta analisada, que nem sempre trata a discriminação racial como um dos grandes riscos envolvidos em uma moderação de conteúdo inadequada. Apesar de a regulação de plataformas envolver mais do que apenas normas para moderação, esta é uma das atividades mais relevantes realizadas por intermediários, uma vez que atinge diretamente direitos fundamentais, como é o caso da liberdade de expressão. Nesse contexto, um dos grupos mais vulnerabilizado, hostilizado e prejudicado nos ambientes online é justamente o das pessoas racializadas, especialmente as mulheres negras, que sofrem diversas desvantagens e são vítimas de inúmeras formas de violência. Entendemos que o marco regulatório brasileiro sobre plataformas digitais precisa ser direta e ativamente antirracista, de modo a impedir a reprodução do racismo estrutural; logo, procuraremos, neste trabalho, verificar como essa questão foi tratada na consulta do CGI.br.

Para tanto, a metodologia abordada foi indutiva, partindo do texto da consulta e das contribuições dos consultados, com a utilização do método de análise de conteúdo. Assim, na primeira parte deste artigo (seção 2), apresentaremos algumas especificidades do problema racial no Brasil e analisaremos diferentes expressões do racismo contra pessoas negras nas plataformas digitais e por meio delas. Em seguida, empreenderemos uma análise da consulta realizada pelo CGI.br, a fim de identificarmos como essa temática aparece e é enfrentada. Após alguns esclarecimentos metodológicos (seção 3), este artigo procurará sintetizar seus resultados tanto pela perspectiva das perguntas propostas pelo consulente (seção 4), quanto pelos comentários feitos pelos consultados (seção 5).

O racismo antinegro em plataformas digitais

O racismo é um fenômeno múltiplo e complexo, fortemente contextualizado, com diversas formas de manifestação e ocultamento, passando não só por agressões explícitas e intencionais, mas também por ofensas implícitas, encobertas ou não intencionais. Para compreender a forma com que o racismo opera *na e por meio de* plataformas digitais, convém recuperar brevemente as duas formas básicas pelas quais ele se apresenta, quais sejam, a discriminação racial direta e a indireta. A primeira é a forma mais comumente identificada, pois se manifesta de maneira clara e inequívoca. Trata-se do tipo de discriminação proferida, por exemplo, por meio de ofensas, ou seja, mediante termos e expressões que visam diminuir, menosprezar ou colocar em desvantagem as pessoas em virtude de sua raça.

De acordo com Luiz Valério Trindade (2018), as maiores vítimas dessa forma de discriminação nas redes sociais, em especial no Facebook, são mulheres negras de 20 a 35 anos, de classe média, com Ensino Superior completo. O recorte específico desse público-alvo demonstra que nem

¹ Escolhemos neste estudo, como recorte metodológico, abordar a questão do racismo sobre as pessoas negras em função do histórico escravocrata no país, bem como do contexto de vulnerabilidade dessa população no espaço digital, a ser melhor explicitado na próxima seção. A escolha, no entanto, não reflete desconhecimento ou ignorância da questão racial contra pessoas indígenas online, que pode ser objeto de trabalhos futuros.

mesmo a ascensão social é um fator capaz de impedir a discriminação: pelo contrário. As postagens mais vulneráveis a ofensas racistas são justamente aquelas em que mulheres negras demonstram exercer posição de liderança ou de destaque, por exemplo, conduzindo um programa televisivo, desfrutando de férias, viajando ao exterior ou em profissões de maior prestígio social, como medicina e direito. O estudo conduzido por Trindade (2018) mostrou também que as ofensas aumentavam quando mulheres negras se contrapunham ou discordavam de postagens de cunho negativo contra outras pessoas negras.

A misoginia, em casos como esses, foi reafirmada em pesquisa realizada pela Faculdade Baiana de Direito, JusBrasil e PNUD (2023), que analisou decisões judiciais sobre racismo e injúria racial entre os anos de 2010 e 2022. Identificou-se que a maioria dos insultos destinados a mulheres negras versavam sobre sua higiene, sexualidade ou estética. Por outro lado, as ofensas contra homens negros concentravam-se na inferiorização social. Já a discriminação racial indireta é aquela que se manifesta de forma dissimulada, mediante práticas aparentemente neutras, mas que resultam em atos discriminatórios que colocam as pessoas negras novamente em desvantagem em razão de sua raça (Brasil, 2022). Em plataformas digitais, isso ocorre especialmente por meio da moderação de conteúdo, que neste artigo entendemos como o modo pelo qual essas plataformas tratam da gestão e do gerenciamento sobre o conteúdo publicado por terceiros em seus espaços, tanto através da remoção de postagens ou suspensão de perfis, como pela recomendação de conteúdo (Silva; Gertrudes, 2023). Apesar de não haver um consenso explícito sobre a inclusão de recomendação na mesma classificação que remoção de conteúdo, especialistas consideram que ambas podem ser consideradas ações de moderação, ainda que para finalidades distintas: a primeira como uma curadoria e a segunda como ação de intervenção (Silva; Gertrudes, 2023).

Pessoas negras tendem a ter mais postagens removidas indevidamente, principalmente quando expressam suas identidades, por exemplo, tratando de questões raciais ou denunciando situações racistas. Nesse sentido, nos Estados Unidos, verificou-se que as ferramentas de moderação de conteúdo de grandes plataformas digitais tendem a considerar a variação linguística do inglês afro-americano como altamente tóxica, em virtude de determinadas expressões utilizadas. Diferentes influenciadores digitais negros e negras também já relataram “prejuízos observados nas métricas de alcance, o fenômeno do embranquecimento de rostos em sistemas de reconhecimento [como filtros] e resultados de busca de imagens que reforçam estereótipos racistas” (Lima; Lima, 2023, p. 15). No caso de filtros de imagens que promovem o clareamento do tom de pele como uma forma de “embelezamento”, especialistas negros do país classificaram esses sistemas algorítmicos como tecnologias de risco inaceitável (Silva, 2023).

A noção de “racismo algorítmico” é uma maneira de agrupar e compreender esses diferentes fenômenos. Segundo Tarcízio Silva (2022), o racismo algorítmico pode ser definido como “o modo pelo qual a disposição de tecnologias e imaginários sociotécnicos em um mundo moldado pela supremacia branca realiza a ordenação algorítmica racializada de classificação social, recursos e violência em detrimento de grupos minorizados” (p. 68-69). Trata-se de um fenômeno que “trabalha com a categoria de racismo como estrutura e forma sistemática de discriminação”, sendo os vieses algorítmicos aqueles responsáveis por mobilizar “as categorias de preconceito racial e discriminação racial” (Kremer, 2023, p. 28).

Apesar de mais velada que a discriminação racial direta, é importante considerar que a discriminação indireta é igualmente nociva e prejudicial. Influenciadores negros e negras que utilizam as redes sociais como ferramenta de trabalho podem ser, sobremaneira, prejudicados por um alcance reduzido sem justificativa, ainda mais se optarem por abordar questões raciais em seu conteúdo (Thamara, 2023).

Como destaca Silva (2022), lidar com o racismo algorítmico significa ter de desvelar uma dupla opacidade, presente tanto em nossa relação com a tecnologia como também a raça. Primeiro, aquela que gira em torno da tecnologia, uma vez que algoritmos e sistemas de inteligência artificial (IA) são comumente tidos como objetivos e neutros. Segundo, aquela em torno do próprio racismo, que no Brasil é marcadamente por denegação (Gonzalez, 1988), ou seja, um assunto que se evita abordar diretamente. Assim, segundo Silva (2020), a opacidade dos algoritmos, que regem a visibilidade e a classificação dos conteúdos postados nas plataformas digitais sem qualquer transparência, como verdadeiras caixas pretas, é agravada pela ideologia da “cegueira racial”, típica do Brasil, que tradicionalmente dissimulou seu racismo celebrando a miscigenação e a democracia racial.

Considerando o cenário brasileiro, organizações como o coletivo Juristas Negras (2022) e o AqualtuneLab (Barbosa, 2022), ao contribuírem para a construção do marco teórico sobre IA (Sugestões..., 2022), posicionaram-se pela necessidade de uma regulação que preveja expressamente o antirracismo como um princípio basilar. Segundo o coletivo Juristas Negras (2022), é preciso considerar que “o Brasil ainda não reconheceu que suas instituições são racistas e preconceituosas, contexto no qual o estereótipo de criminoso é sempre imputado às pessoas negras, pobres e faveladas” (p. 6). Dessa forma, concluem que o vazio da norma sobre este tema “contribuirá para que mais pessoas negras sigam sendo afetadas, sem que a legislação sequer enfrente minimamente o problema” (Juristas Negras, 2022, p. 6).

Embora sejam contribuições destinadas ao contexto regulatório de sistemas de IA, esse entendimento também pode ser aplicado à regulação de plataformas digitais. Atualmente, a moderação de conteúdo online é realizada majoritariamente por meio de modelos algorítmicos, podendo ser considerada um verdadeiro produto dessas plataformas (Silva; Gertrudes, 2023). Assim, escolhas corporativas por moderar mais e/ou melhor passam também por uma avaliação acerca dos interesses financeiros das empresas, que definirá as prioridades de investimento.

Em 2020, a filósofa Djamila Ribeiro, junto com coletivos negros e uma equipe dirigida pelo jurista Adilson Moreira, chegou a ingressar com uma representação frente ao Ministério Público Federal contra o antigo Twitter, atual X, sob a alegação de que a plataforma, por meio de sua política de moderação de conteúdo, lucraria por meio da exploração econômica do racismo e da misoginia (Tardelli, 2020). Em entrevista, Moreira sustentou que a empresa teria permitido a circulação de “informações inverídicas que fomentam o ódio contra minorias” (Tardelli, 2020), mesmo tendo controle sobre as regras que se aplicam sobre o que pode circular em seu espaço.

Nesse sentido, a construção de uma regulação destinada a traçar regras e parâmetros para plataformas digitais, em especial com diretrizes mínimas para uma moderação online justa e legítima², poderia servir como instrumento para obrigar as empresas a adotarem uma postura mais rigorosa em relação a conteúdos nocivos e ao funcionamento de seus próprios algoritmos. Considerando o caráter estrutural do racismo e sua capacidade de reprodução facilitada pela ausência de mecanismos proativos antirracistas, apresenta-se como fundamental que uma regulação para plataformas digitais preveja dispositivos específicos de enfrentamento direto. Inclusive, essa é uma recomendação de Tendayi Achime (2020), ex-relatora especial da ONU em formas contemporâneas de racismo e intolerâncias correlatas. Em seu relatório sobre discriminação racial em tecnologias emergentes, ela ressalta que:

[...] eles [os Estados] devem abordar não apenas o racismo e a intolerância explícitos no uso e design das tecnologias digitais emergentes, mas também, e com igual seriedade, **as formas**

² Novamente, destaca-se que a intenção não é realizar uma confusão entre regulação de plataformas digitais e regulação de moderação de conteúdo, pelo contrário: ao reconhecer o impacto desta atividade sobre a existência e identidade de pessoas negras no ambiente digital, o foco se dá sobre ela por sua relevância para a pesquisa.

indiretas e estruturais de discriminação racial que resultam do design e uso dessas tecnologias. [...] Os Estados devem rejeitar uma abordagem "cega às raças" à governança e regulamentação das tecnologias digitais emergentes, uma abordagem que ignora a marginalização específica de minorias raciais e étnicas e conceitua problemas e soluções relacionados a essas tecnologias sem levar em conta seus prováveis efeitos sobre esses grupos. (Achieme, 2020, grifo nosso, tradução nossa)

Com efeito, a tendência de uma regulação que se pretende universal é tratar os desiguais de forma igual, resultando no reforço de exclusões e marginalizações dos mesmos grupos já vulnerabilizados. Daí a necessidade de um olhar específico e atento a essas desigualdades, visando endereçá-las de forma adequada. Analisaremos, na sequência deste artigo, como esse tema foi abordado e enfrentado na consulta sobre regulação de plataformas digitais realizada pelo CGI.br.

Análise da consulta do CGI.br: considerações metodológicas

Entre 25 de abril e 16 de julho de 2023, o CGI.br abriu uma consulta pública sobre regulação de plataformas digitais, visando ir além da discussão regulatória até então pautada pelo Projeto 2.630 (Brasil, 2020) no país, popularmente conhecido como “PL das *Fake News*” (Soares, 2023). A consulta explorou definições e classificações de plataformas, apresentando um mapeamento dos riscos envolvidos em suas atividades, assim como medidas regulatórias capazes de mitigá-los. Uma série de perguntas foram estruturadas em três eixos principais, que versavam sobre “quem regular?”, “o que regular?” e “como regular?”. Foram obtidas mais de 1,3 mil contribuições, cujo resultado foi sistematizado e divulgado pelo próprio CGI.br (Regulação..., 2023).

Considerando o impacto desse documento, que representa uma síntese das principais contribuições de diversos setores da sociedade sobre o tema, pretendemos explorar esse material e analisar em que medida a questão racial despontou como central ou não, tanto no texto da chamada do Comitê, quanto nas respostas dos participantes. Antes de apresentarmos os resultados obtidos, explicaremos a metodologia utilizada. Nosso estudo partiu de uma análise de conteúdo com base nas contribuições sistematizadas e disponibilizadas em formato .CSV³ pelo Comitê em dois momentos: primeiro, com foco no endereçamento da temática pelo consulente; na sequência, nos comentários dos consultados.

O primeiro momento procurou compreender se a discussão sobre raça foi incentivada na consulta e, em caso positivo, como foi, analisando-se, portanto, o endereçamento pelo consulente. Para seleção das colunas a serem analisadas, nosso estudo apoiou-se no dicionário disponibilizado pelo Comitê⁴, nas colunas “capítulo” e “pergunta”. Uma vez que houve contribuições nos três capítulos e/ou eixos temáticos e nas 43 perguntas propostas, optamos por analisar os termos utilizados nas duas colunas. Para o segundo momento, cujo objetivo era compreendermos se a raça foi uma categoria indicada nas contribuições e como ela foi mobilizada, limitamo-nos a fazer a busca na coluna “comentario_texto”, descrito como o único campo no qual foi sistematizado o “conteúdo do comentário”.

Tendo em vista o foco deste trabalho, buscamos pelos seguintes termos no documento citado: “racismo”, “raça”, “negro”, “negra”, “preto”, “preta”, “etnia” e “discriminação racial”⁵. Os resultados considerados em ambos os momentos da análise foram aqueles

³ A tabela completa utilizada para a pesquisa pode ser consultada em: <https://dialogos.cgi.br/documentos/debate/consulta-plataformas/>. Acesso em 11 maio 2025.

⁴ Disponível em: <https://dialogos.cgi.br/documentos/debate/consulta-plataformas/>. Acesso em 11 maio 2025.

⁵ A expressão “preconceito racial” não foi utilizada para busca, uma vez que, diferentemente da discriminação racial, considera-se apenas “o juízo baseado em estereótipos acerca de indivíduos que pertençam a um determinado grupo racializado, e que pode ou não resultar em práticas discriminatórias. [...] é a atribuição de tratamento diferenciado a membros de grupos racialmente identificados” (Almeida, 2018).

indexados com a expressão “termo encontrado” no arquivo do documento. Na sequência, foram excluídos apenas resultados que não correspondiam ao tema investigado; em alguns casos, por exemplo, o termo “preta” foi localizado por estar contido na palavra “interpretação”.

A análise das menções finais foi feita a partir das seguintes perguntas orientadoras: Qual o tema do eixo? Quantas menções foram encontradas e de quais termos? Em quais momentos e para qual finalidade foi feita a menção? O que o número de menções significa? Qual sua posição no texto? Em que medida as menções contribuem para uma regulação de plataformas?

A raça na pergunta do consulente

Por meio da busca dos termos “racismo”, “raça”, “negro”, “negra”, “preto”, “preta”, “etnia” e “discriminação racial” nas colunas “capítulo” e “pergunta”, identificamos que nenhum dos termos foram encontrados em quaisquer dos eixos da consulta. Na leitura geral dos capítulos e das perguntas propostas, observamos que os direcionamentos dados pelo CGI.br buscaram i) contextualizar o tema/eixo em questão e ii) apresentar as perguntas com exemplos, quando cabível. Assim, apesar de ter aparentemente buscado uma descrição mais genérica, houve uma preocupação em explicar de forma adequada o que era solicitado, exemplificando possíveis tópicos que poderiam ser abordados nas respostas dos contribuintes.

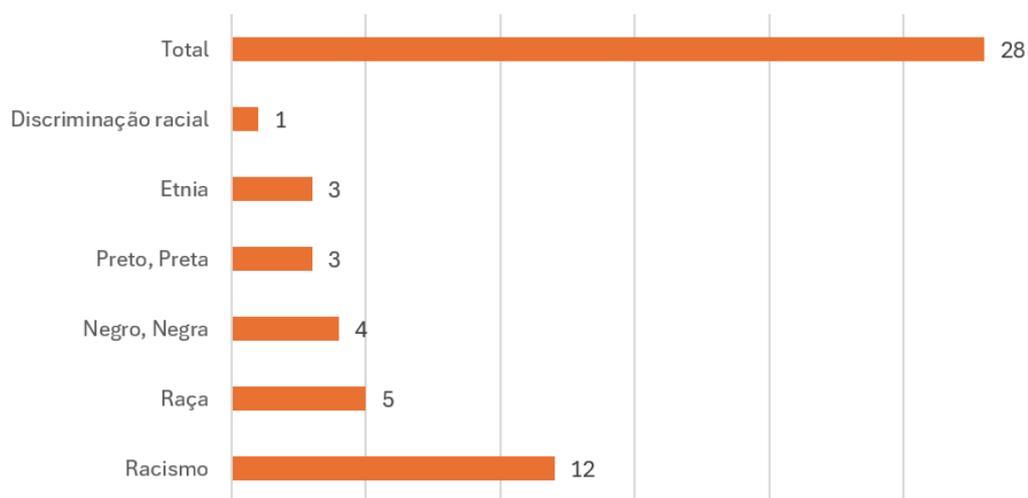
Inferimos, em princípio, que a questão racial está abarcada pelo tema tratado por alguns eixos, como o de riscos à Democracia e aos Direitos Humanos. Na seção “o que regular?”, mencionamos o interesse do consulente em explorar possíveis impactos sobre direitos fundamentais, como o da não discriminação, além dos riscos associados ao discurso de ódio. No entanto, não há um endereçamento específico da questão racial, uma vez que a concepção de Direitos Humanos e não discriminação é ampla, podendo abarcar diversos temas. Em verdade, verificamos que a abordagem mais genérica, sem identificar um grupo em particular, parece ter orientado os textos de todas as perguntas, com exceção daquelas dirigidas especificamente à proteção das “crianças e adolescentes”.

Essa abordagem não afasta, entretanto, a preocupação com relação a uma abordagem regulatória que explicita e proteja as populações racializadas no Brasil, ainda mais pela preocupação expressa com riscos discriminatórios. Em um país estruturalmente racista, abordar a raça como um elemento central, quiçá entre os exemplos mencionados na chamada, poderia contribuir para uma regulação mais apropriada e eficiente, especialmente tendo em vista a contenção de violações aos direitos fundamentais em ambientes online contra grupos racializados.

A raça nos comentários dos consultados

Ao se buscar pelos termos “racismo”, “raça”, “negro”, “negra”, “preto”, “preta”, “etnia” e “discriminação racial” na coluna “comentario_texto” do documento com a sistematização de todas as contribuições, eles foram encontrados, em conjunto ou separadamente, em 22 comentários/contribuições, conforme quadro a seguir:

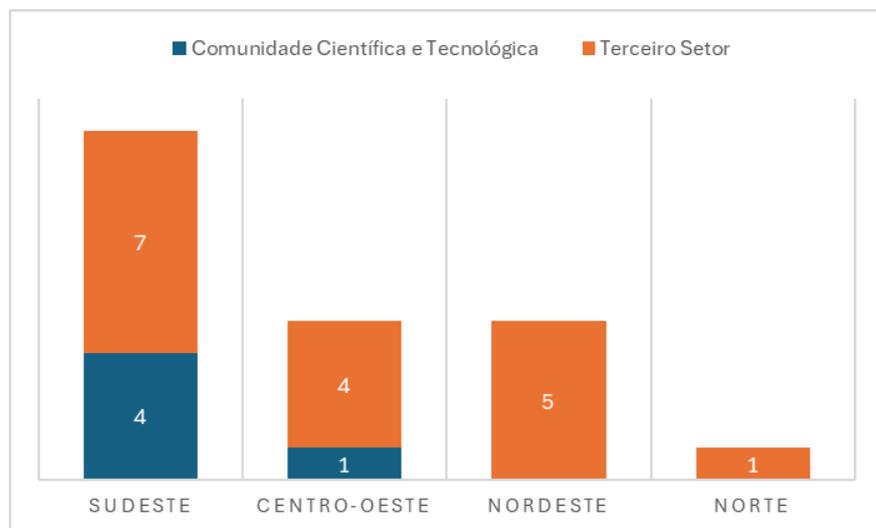
Gráfico 1 – Resultado da busca nas contribuições dos consultados por termos



Fonte: Elaboração própria.

Observamos, segundo os critérios de região e setor do autor das contribuições, maior frequência das menções nas respostas da região Sudeste e do terceiro setor. Notamos que não houve referência aos termos nas contribuições de autores da Região Sul, que respondiam por apenas 4% do total de respostas, nem nas participações do setor empresarial ou governamental, que, juntos, somaram 19,5% do total de contribuições.

Gráfico 2 – Contribuições com menção aos termos por região e setor da autoria



Fonte: Elaboração própria.

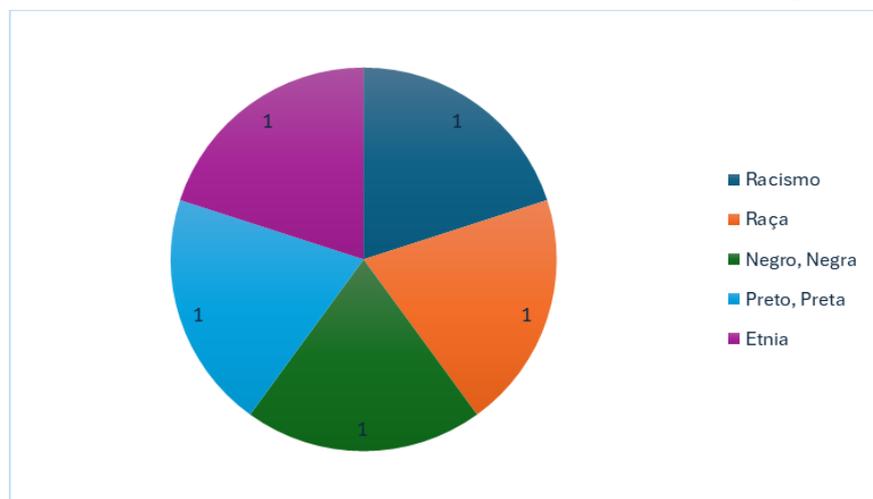
Visando organizar melhor o estudo desse material, o resultado dos termos encontrados serão analisados conforme os tópicos da consulta pública em que apareceram, quais sejam: “Quem regular”, “O que regular” e, por fim, “Como regular”.

5.1 Quem regular?

No tópico “Quem regular?”, o objetivo da consulta foi abordar a abrangência conceitual e o escopo sobre o qual recai a regulação, ou seja, a intenção foi compreender o que se entende por “plataformas digitais”. O capítulo 1 foi o que teve o menor número de perguntas em relação aos outros (quatro ao todo), as quais buscavam por “definições, tipologias e áreas de atuação das plataformas digitais” (NIC.br, 2023, p. 11).

A seção teve 225 contribuições. Os termos pesquisados em nossa análise foram identificados, entretanto, em apenas quatro delas, sempre em perguntas distintas do mesmo capítulo. Foi encontrada uma menção para os termos “racismo”, “raça”, “negro/negra”, “preto/preta” e “etnia”. Não foram identificadas menções ao termo “discriminação racial”.

Gráfico 3 – Resultado da busca nas contribuições dos consultados para o Eixo 1



Fonte: Elaboração própria.

Uma das contribuições, em resposta à melhor definição de “plataformas digitais”, indicou a importância de estimular “a diversidade e a equidade de classe, raça, gênero, orientação sexual e outros”⁶. Outro consultado salientou que a segmentação da sociedade em nichos a partir de padrões, como a etnia, deveria ser considerada fortemente, destacando ser natural às plataformas segmentarem os dados em diversas classes, embora elas criem uma ideia superficial de pluralidade, podendo contribuir para a manutenção de determinados grupos em situação de inferioridade social⁷.

A menção ao racismo ocorreu em comentário que tratou do “controle essencial de acesso (*gatekeeper*)”. O termo foi mencionado em um apelo para que a atuação do *gatekeeper* fosse avaliada à “luz da universalidade de acesso aos cuidados digitais”⁸. Isso é, para além da proteção contra a concentração de mercado, o comentário sugere que, na análise da atuação da plataforma em serviços essenciais, seja considerado como o agente protege corpos e existências de violências, como o racismo.

Por fim, embora tenhamos identificado uma menção nesse capítulo aos termos “pretas” e “negras”, a análise apurou que eles não foram utilizados com sentido racial. O termo “pretas” foi empregado como adjetivo de “caixas pretas”, em referência à comunicação não transparente de empresas que mascaram seu real modelo de negócio⁹, ao passo que o termo “negras” se tratou de uma referência a outro comentário da consulta, das “Blogueiras Negras”¹⁰.

No geral, os termos aparecem em trechos que evidenciam a importância de considerar os impactos das plataformas sobre determinados grupos, apontando para a importância de uma regulação protetiva, sendo a raça ou a etnia indicada como um desses marcadores.

⁶ Comentário_id 1252.

⁷ Comentário_id 892.

⁸ Comentário_id 1170.

⁹ Comentário_id 1168.

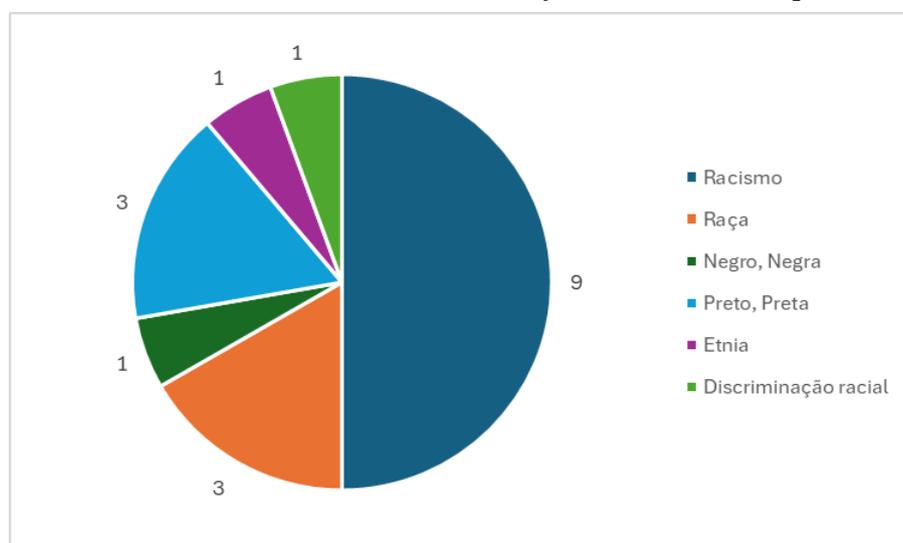
¹⁰ Comentário_id 1168.

5.2 O que regular?

No tópico “O que regular?”, o objetivo da consulta era colher contribuições acerca dos riscos provenientes da atividade de plataformas digitais, assim como possíveis medidas de mitigação. Sendo o eixo que mais recebeu contribuições, contava com 34 perguntas referentes a quatro grandes conjuntos de riscos que versavam sobre: i) concorrência, direito do consumidor, abuso de poder econômico e concentração econômica e de dados; ii) soberania digital, desenvolvimento tecnológico e inovação; iii) trabalho decente; e iv) Democracia e Direitos Humanos (NIC.br, 2023, p. 63).

No total, foram encontradas nove menções ao termo “racismo”, três menções dos termos “raça” e “negro/negra” e uma menção aos termos “preto/preta” e “discriminação racial”. A maioria das menções concentra-se no eixo de riscos à Democracia e aos Direitos Humanos, indicando, como esperado, uma associação maior da questão racial a esses temas.

Gráfico 4 – Resultado da busca nas contribuições dos consultados para o Eixo 2



Fonte: Elaboração própria.

Nesse eixo, os termos apareceram para apontar que o racismo é um problema grave em plataformas digitais, causado tanto pela proliferação de discurso de ódio racista como pelos próprios algoritmos utilizados para moderação de conteúdo que, em alguns casos, reduzem o alcance ou suspendem injustamente a conta de pessoas que se posicionam contra o racismo¹¹. As menções também aparecem relacionadas a medidas de mitigação, como a adoção de uma moderação de conteúdo humana mais plural e representativa, a fim de identificar melhor o discurso de ódio contra pessoas negras¹². Além disso, propõem-se a obrigatoriedade de auditorias algorítmicas externas e avaliação de padrões de códigos discriminatórios¹³, de apresentação periódica de relatórios de transparência e *accountability*, com medidas para enfrentar a discriminação racial, inclusive a indireta, bem como o investimento público para a formação de grupos interdisciplinares e multissetoriais dedicados a desenvolver, analisar e auditar modelos algorítmicos nos termos do Estatuto da Igualdade Racial¹⁴.

¹¹ Comentario_id 151, 247, 1134, 1271 e 541.

¹² Comentario_id 1128.

¹³ Comentario_id 162 e 249.

¹⁴ Comentario_id 1059.

Dentre as ameaças à Democracia e aos Direitos Humanos, identificamos menções que tratavam especificamente dos riscos associados a jogos online, com a questão central nas distorções de imagem proporcionadas por eles: crianças e adolescentes negros possivelmente evitam utilizar avatares que representassem adequadamente a cor de sua pele para não sofrerem racismo¹⁵. Outro problema associado à juventude negra foi encontrado no eixo de riscos ao trabalho decente: uma contribuição apontou que, no caso de aplicativos de entrega de comida, em que o ganho do entregador aumenta com a quantidade e a velocidade de cada entrega, é possível traçar um perfil específico dos adolescentes mais vitimados por esse tipo de trabalho infantil: são meninos de 14 a 17 anos, pretos e de origem familiar vulnerabilizada socioeconomicamente¹⁶. Nesse mesmo eixo, outra menção apontava para a necessidade de respeito às leis trabalhistas locais e à representatividade da população brasileira em gênero, raça e afins para a contratação de trabalhadores por essas plataformas.

Nos eixos dos riscos relacionados a concorrência, consumo, economia, soberania digital e desenvolvimento tecnológico, por sua vez, as menções concentraram-se novamente em propor medidas de mitigação de vieses algorítmicos e, ainda, aplicação de multas e outras sanções. No que tange a essa última, a contribuição foi no sentido da previsão de penalidades quando a qualidade dos serviços oferecidos for comprometida devido à falta de observância a normas locais, como aquelas de combate ao racismo¹⁷. De outra sorte, como medidas de mitigação, foram relacionadas novamente as auditorias algorítmicas, agora ao lado de requisitos de transparência, diversidade, inclusão, salvaguardas contra discriminação¹⁸ e análise de riscos sistêmicos por meio do dever de cuidado¹⁹.

Por fim, houve uma menção ao termo “preto” no eixo sobre concorrência, consumo e economia, mas apenas no título de uma das referências utilizadas: um texto de Paulo Victor Melo que defende a diversidade não somente nas telas, como também por trás delas. A contribuição em si dizia respeito ao impacto negativo da concentração de mercado e do abuso de poder no ecossistema midiático brasileiro²⁰.

De todas essas menções, percebemos que houve uma importante preocupação com os riscos advindos da perpetuação do racismo *através e pelas* plataformas digitais. As sugestões de medidas de mitigação foram variadas e apontam caminhos importantes. Em especial, chama à atenção a sugestão de relatórios de transparência que tratem especificamente das medidas contra a discriminação racial e de times de moderação humana e equipes de desenvolvimento de IA mais diversas, com a participação de pessoas negras. Trata-se de elementos que reconhecem explicitamente a questão racial e a colocam como eixo central para a atuação das plataformas, contribuindo para o desvelamento do problema.

5.3 Como regular?

No capítulo de perguntas “Como regular: arranjos institucionais”, o objetivo foi colher análises e sugestões sobre os arranjos necessários para implementar a regulação das plataformas, considerando a complexidade e a transversalidade de suas atividades. A partir de cinco perguntas, foram exploradas a arquitetura regulatória, a indicação de atores e a definição de seus papéis e responsabilidades (NIC.br, 2023, p. 23).

Essa seção teve o menor número de contribuições: 135 no total. Dentre elas, os termos pesquisados foram identificados em apenas duas contribuições. Foram encontradas duas

¹⁵ Comentario_id 795.

¹⁶ Comentario_id 769.

¹⁷ Comentario_id 749.

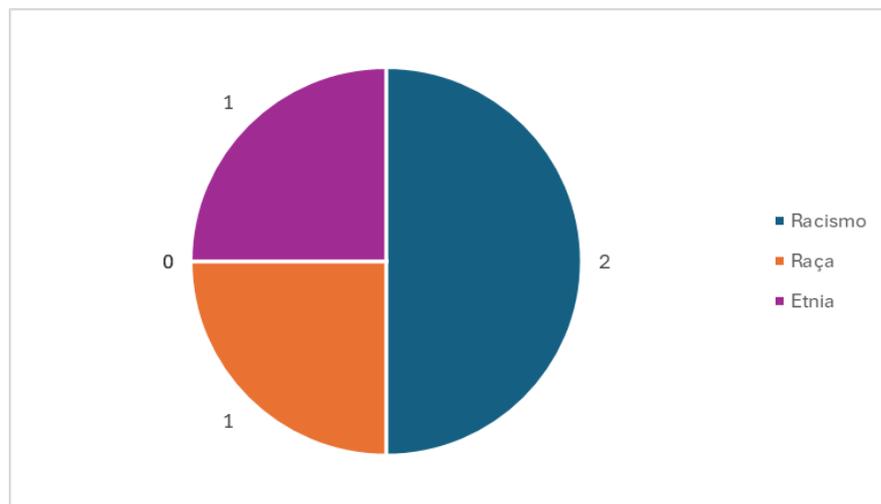
¹⁸ Comentario_id 359.

¹⁹ Comentario_id 1065.

²⁰ Comentario_id 561.

menções ao termo “racismo” e uma a “raça” e “etnia”, e não foram identificadas menções aos termos “negro/negra”, “preto/preta” e “discriminação racial”.

Gráfico 5 – Resultado da busca nas contribuições dos consultados para o Eixo 3



Fonte: Elaboração própria.

A primeira menção ao termo “racismo” foi feita com o intuito de citar a participação de Organizações não Governamentais (ONG) e Institutos de Pesquisa no combate ao racismo, indicados como outros atores que devem ser envolvidos na regulação da plataforma²¹. Incide-se “outros atores” porque a pergunta que dirige a resposta foi feita para mapear *outros* atores para além daqueles envolvidos *diretamente* com a implementação. Convém destacar que não identificamos, no levantamento ou na leitura direta das consultas, menção expressa da importância de representações institucionais comprometidas com a equidade racial estarem diretamente envolvidas na regulação. Os atores, na única citação encontrada, exerceriam um papel apenas secundário.

A segunda menção ao “racismo” e a menção aos termos “raça” e “etnia” ocorreram em um mesmo comentário²², que respondia à pergunta sobre a necessidade de criação de novas instituições para a implementação da regulação de plataformas digitais. Propondo a criação de um novo conselho representativo e multissetorial, com foco nas pessoas, a sugestão cita os problemas estruturais relacionados a classe, misoginia e racismo no Brasil, argumentando que eles demandariam uma atuação especial do órgão. É defendida, portanto, uma autoridade composta com “diversidade regional, de gênero, de etnia, de raça, de deficiência”, afirmando que não há hoje no país órgão com tal desenho e configuração. A proposição é pertinente por estabelecer que a raça seja explicitamente considerada no desenho da autoridade proposta; desse modo, norteia um compromisso de combate ao racismo também na observância de representações comprometidas com a igualdade racial.

Conclusão

Embora a análise da consulta sobre regulação de plataformas digitais realizada pelo CGI.br aponte para um avanço do debate no Brasil em relação à questão racial, ainda é possível identificar alguns limites e dificuldades. Dada a dimensão do racismo no Brasil, em suas múltiplas manifestações, diretas ou indiretas, envolvendo diversas formas de violência e

²¹ Comentario_id 850.

²² Comentario_id 1183.

graves violações de direitos fundamentais em grande escala nos ambientes online, qualquer esforço regulatório que trate da moderação de conteúdo em grandes plataformas digitais tem a obrigação de tratar esse tema de maneira direta e explícita, devendo ser incisivo no enfrentamento e na adoção de medidas antirracistas de mitigação.

A postura do consulente na formulação das questões tendeu a ser mais genérica, sem indicar diretamente a questão racial. Tal abordagem contribui com o ocultamento do problema: a omissão ou a não ação impede o enfrentamento das discriminações raciais sofridas *por* e *no* ambiente digital adequadamente. Entendemos, assim, que as contribuições poderiam avançar mais nessa temática se a raça fosse endereçada na consulta de forma mais explícita e central.

Na análise das contribuições, destaca-se especialmente o eixo de riscos à democracia e aos Direitos Humanos no âmbito do tópico “O que regular?”, no qual identificamos 18 menções aos termos pesquisados, demonstrando uma preocupação significativa com a questão racial. As propostas ressaltam a importância de uma moderação de conteúdo mais plural e justa, guiada por valores como a transparência, a igualdade e a diversidade. Merece destaque a sugestão de medidas de mitigação, por exemplo a apresentação de relatórios de transparência que explicitem ações adotadas contra a discriminação racial. Alguns comentários, na linha defendida neste artigo, destacam a importância de uma postura ativa contra o racismo, envolvendo a análise de riscos sistêmicos, especialmente em um país estruturalmente racista como o nosso, e a necessidade de ações preventivas e do dever de cuidado.

Por fim, no tópico “Como regular?”, merece destaque a sugestão de criação de uma nova autoridade regulatória no Brasil, com composição diversa, levando em consideração a representação racial, um ponto relevante e frequentemente negligenciado em nosso país. A importância desse debate é ainda maior em razão da centralidade e da polêmica envolvida na discussão legislativa acerca de quem deveria assumir a responsabilidade regulatória nesse domínio. Uma autoridade regulatória pouco representativa em termos raciais é um mal começo, aumentando a tendência reprodutiva de nosso racismo estrutural.

Concluimos este estudo ressaltando a importância da iniciativa do CGI.br, que tem fomentado o debate qualificado sobre a regulação de plataformas digitais no Brasil e permitido que, entre outros temas, também a questão racial seja endereçada e enfrentada. Esperamos que novas consultas e debates sejam promovidos, abordando também o racismo algorítmico em suas diversas dimensões, *nas* e *por meio das* plataformas digitais, assim como a necessidade da adoção de medidas antirracistas.

Referências

- ACHIUME, Tendayi. Racial discrimination and emerging digital technologies: a human rights analysis. Geneva: UN, Human Rights Council, 2020. Disponível em: https://digitallibrary.un.org/record/3879751/files/A_HRC_44_57-EN.pdf?ln=en. Acesso em: 28 nov. 2024.
- ALVES, M.A.S.; MACIEL, E.H. O fenômeno das fake news: definição, combate e contexto. *Internet & Sociedade*, v. 1, n. 1, p. 144-171, 2020.
- BARBOSA, Arthur Almeida Meneses et al. Documento Preto: contribuições do AqualtuneLab para o debate sobre regulação de Inteligência Artificial no Brasil. 2022. Disponível em: <https://aqualtunelab.com.br/wp-content/uploads/2022/11/AQUALTUNELAB-DocumentoPreto-A5-V2-web.pdf>. Acesso em: 24 nov. 2024.
- BRASIL. Decreto n. 10.932, de 10 de janeiro de 2022. Promulga a Convenção Interamericana contra o Racismo, a Discriminação Racial e Formas Correlatas de Intolerância, firmado pela República Federativa do Brasil, na Guatemala, em 5 de junho de 2013. Brasília, DF: Presidência da República,

[2022]. Disponível em:

https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2022/Decreto/D10932.htm. Acesso em: 19 fev. 2024.

BRASIL. Projeto de Lei 2630/2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Brasília: Câmara dos Deputados, 2020. Disponível em:

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1909983&filename=PL%202630/2020. Acesso em: 30 nov. 2024.

COMITÊ GESTOR DA INTERNET. Consulta sobre Regulação de Plataformas Digitais. 2023. Disponível em: <https://dialogos.cgi.br/documentos/debate/consulta-plataformas/>. Acesso em: 28 nov. 2024.

GONZALEZ, Lélia. A categoria político-cultural de amefricanidade. Tempo Brasileiro, Rio de Janeiro, n. 92/93, jan./jun. 1988, p. 69-82.

JURISTAS NEGRAS. Contribuição JURISTAS NEGRAS. 10 jun. 2022. Disponível em:

<https://legis.senado.leg.br/sdleg-getter/documento/download/2bf6209f-f6eb-4d49-b447-e626e7c55a77>. Acesso em: 25 jun. 2023.

KREMER, Bianca. Racismo algorítmico. Rio de Janeiro: CESeC, 2023.

LIMA, Brunno Ewerton de Magalhães; LIMA, Carla Fernanda de. Racismo algorítmico: vivências e percepções de influenciadores(as) digitais negros(as). Revista África e Africanidades, ano 16, ed. 47-48, ago./nov. 2023. Disponível em:

https://africaeaficanidades.com.br/documentos/ARTLIV03_ed.47-48.pdf. Acesso em: 20 nov. 2024.

MOURA, Clóvis. Sociologia do negro brasileiro. São Paulo: Perspectiva, 2019.

NASCIMENTO, Abdias do. O genocídio do negro brasileiro: processo de um racismo mascarado. São Paulo: Perspectiva, 2016.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR (ed.). Sistematização das contribuições à consulta sobre regulação de plataformas digitais. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2023.

REGULAÇÃO de plataformas digitais: relatório do CGI.br mapeia consensos e dissensos entre setores. CGI.br. 13 dez. 2023. Disponível em: <https://cgi.br/noticia/releases/regulacao-de-plataformas-digitais-relatorio-do-cgi-br-mapeia-consensos-e-dissensos-entre-setores/>. Acesso em: 26 nov. 2024.

SILVA, Fernanda dos Santos Rodrigues; GERTRUDES, Júlia Maria Caldeira. Governança da moderação de conteúdo online: percepções sobre o papel dos atores e regimes. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2023. Disponível em: <https://bit.ly/3WHMUIg>. Acesso em: 20 jan. 2024.

SILVA, Fernanda dos Santos Rodrigues. “Nada mais sobre nós sem nós”. Escurecendo o Debate sobre a Regulação de IA no Brasil e Pensando Mecanismos de Combate ao Racismo Algorítmico. Relatório de Pesquisa do Programa Líderes LACNIC, 2023. Disponível em:

<https://www.lacnic.net/innovaportal/file/6974/1/regulacao-de-ia-e-discriminacao-algoritmica-informe-de-investigacion-pt.pdf>. Acesso em: 24 nov. 2024.

SILVA, Tarcízio. Racismo algorítmico: inteligência artificial e discriminação nas redes digitais. São Paulo: Edições Sesc São Paulo, 2022.

SILVA, Tarcízio. Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código. In: SILVA, Tarcízio (org.). Comunidades, algoritmos e ativismos digitais: olhares afrodiaspóricos. São Paulo: Literarua, 2020, p. 121-137.

SOARES, Matheus. CGI abre consulta pública sobre regulação de plataformas digitais no Brasil. *desinformante**, 25 abr. 2023. Disponível em: <https://desinformante.com.br/cgi-consulta-publica/>. Acesso em: 26 nov. 2024.

SUGESTÕES ao marco regulatório da IA serão recebidas até 13 de maio. Agência Senado, 4 abr. 2022. Disponível em: <https://www12.senado.leg.br/noticias/materias/2022/04/04/sugestoes-ao-marco-regulatorio-da-ia-serao-recebidas-ate-13-de-maio>. Acesso em: 24 nov. 2024.

TARDELLI, Breno. Adilson Moreira: 'Twitter explora economicamente racismo e misoginia?'. *Carta Capital*, 25 ago. 2020. Disponível em: <https://www.cartacapital.com.br/justica/adilson-moreira-twitter-explora-economicamente-racismo-e-misoginia/>. Acesso em: 24 nov. 2024.

THAMARA, Ariel. Como os vieses raciais no Instagram impactam na vida de criadores de conteúdo negros? *In: CUSTÓDIO, Leonardo et al (orgs.). Práticas de comunicação antirracista no Brasil. 1 ed. Rio de Janeiro: Criola, 2023.*

UNESCO. Diretrizes para a governança das plataformas digitais: salvaguardar a liberdade de expressão e o acesso à informação com uma abordagem multissetorial. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000387560>, pp. 21 - 22.

UNITED NATIONS (UN). Pact for the future, Global Digital Compact, and Declaration on Future Generations. Summit of the Future, Outcome Documents, september 2024. Disponível em: https://www.un.org/sites/un2.un.org/files/sotf-pact_for_the_future_adopted.pdf. Acesso em: 30 nov. 2024.

Mini-bios

Design enganoso e manipulativo em plataformas digitais: um mapeamento de estudos empíricos e reflexões para regulação

George Augusto Valença Santos é professor associado I do Departamento de Computação da UFRPE, onde lidera o grupo ASPAS. Atualmente, atua como cientista-chefe de convênios de cooperação técnica com o Instituto Alana, na área de design ético na educação, e com o TCE-PE, na área de inovação aberta, segurança da informação e IA. Seu grupo de pesquisa investiga os desafios sociotécnicos trazidos por plataformas desenvolvidas por Big Techs. Nos últimos anos, tem dedicado atenção especial a reflexões sobre uma TI ética e inclusiva, com estudos sobre a proteção de pessoas trans e não-binárias, bem como de crianças no ambiente digital, analisando temas como racismo algorítmico e design de futuro, ferramentas de controle parental, garantia dos direitos da criança por design e problemas causados pelo reconhecimento facial para pessoas trans. No pós-doutorado, investigou padrões de design manipulativo na PUC-SP.

Significados do discurso de ódio: desafios regulatórios e disputas discursivas na visão de diferentes setores sociais

Eliza Bacheга Casadei é professora titular do Programa de Pós-Graduação em Comunicação e Práticas de Consumo da Escola Superior de Propaganda e Marketing. Doutora e mestre em Ciências da Comunicação pela ECA-USP, coordena o Grupo de Pesquisa História, Comunicação e Consumo. É bolsista produtividade em pesquisa do CNPq, nível 2.

Gabriela Agostinho Pereira é doutoranda em Comunicação e Práticas do Consumo pelo PPGCOM ESPM - SP e mestre pela mesma instituição. Bolsista CAPES/Prosup, pesquisadora do Grupo de Pesquisa História, Comunicação e Consumo e integrante da Cátedra Maria Aparecida Baccega. Atua nas áreas de comunicação, consumo, plataformas de mídias sociais e discurso de ódio.

Regulação da desinformação nas plataformas digitais e o contexto eleitoral brasileiro

Giovanna Milanese Tavares é advogada, especialista em Proteção de Dados pela Universidade de Lisboa. Bacharel em Direito pela Universidade de Brasília (UnB). Pesquisadora do Centro de Tecnologia e Sociedade da FGV Direito Rio (CTS-FGV), do Laboratório em Governança e Regulação em Inteligência Artificial (LIA-IDP) e do Centro de Direito, Internet e Sociedade (CEDIS-IDP).

Leandro Léo Rebelo é mestrando em Teoria do Estado e Direito Constitucional na PUC-Rio. Diplomado no Programa de Estudos Avançados em Direitos Humanos da American University WCL. Graduado em Direito pela FGV Direito Rio. Pesquisador do Centro de Tecnologia e Sociedade da FGV Direito Rio (CTS-FGV).

Marina Giovanetti Lili Lucena é doutora em Direito Civil pela Universidade do Estado do Rio de Janeiro (UERJ). Mestre e graduada em Direito pela Universidade Federal de Juiz de Fora (UFJF). Pesquisadora do Centro de Tecnologia e Sociedade da FGV Direito Rio (CTS-FGV) e no Legal Grounds Institute.

Atenção, dados e poder: uma análise crítica da infodemia e da desinformação

Leandro Modolo é doutor em Ciências Sociais (Unesp) e doutorando em Saúde Coletiva (Unicamp). Professor da Faculdade de Ciências Médicas da Santa Casa de São Paulo

(FCMSCSP). Pesquisador da iniciativa de prospecção Saúde Amanhã/Estratégia Fiocruz para agenda 2030.

Leonardo Castro é mestre em Psicologia Social (UERJ) e doutor em Antropologia (UERJ). Pesquisador no Departamento de Ciências Sociais da Escola Nacional de Saúde Pública Sérgio Arouca/ Fundação Oswaldo Cruz. Coordenador executivo da iniciativa de prospecção Saúde Amanhã/ Estratégia Fiocruz para a Agenda 2030.

Luis Gonçalves é mestre (UFRJ) e doutor (PUC-SP) em Psicologia Social. Foi pesquisador em saúde digital pela Fiocruz de 2021 a 2024. Faz pós-doutorado (bolsista FAPESP) no Centro de Pesquisa sobre Comunicação e Trabalho (CPCT / ECA-USP). É membro da Câmara de Conteúdos e Bens Culturais do Comitê Gestor da Internet no Brasil (CGI.br) e coordenador do GT Tecnologias Digitais e Subjetividade da Associação Brasileira de Psicologia Social (ABRAPSO).

A questão negra na regulação das plataformas digitais: uma análise da consulta realizada pelo CGI.br

Fernanda dos Santos Rodrigues Silva é doutoranda em Direito, Tecnociências e Interdisciplinaridade pela Universidade Federal de Minas Gerais (UFMG). Mestre em Direitos na Sociedade em Rede e graduada em Direito pela Universidade Federal de Santa Maria.

Marco Antônio Sousa Alves é professor adjunto de Teoria e Filosofia do Direito da Universidade Federal de Minas Gerais (UFMG). Subcoordenador do Programa de Pós-Graduação em Direito (PPGD/UFMG). Bolsista de Produtividade em Pesquisa FAPEMIG/CNPq.

Mariana Alves Araújo Lopes é mestranda em Direito, Tecnociências e Interdisciplinaridade pela Universidade Federal de Minas Gerais (UFMG). Facilitadora e participante do Programa Youth Brasil @ CGI.br (edições 2024 e 2022).