

# IA Responsável, Plataformas de Mídia Social e Regulação no Brasil: Notas para Pensar a Participação Social em Tempos de Incertezas

## Francisco Leite

Doutor em Ciências da Comunicação pela Universidade de São Paulo (ECA-USP), com estágio de doutoramento PDSE/CAPES na Università di Bologna e na Università degli Studi di Trento, Itália; e pós-doutorado em Comunicação, Consumo e DEI pela USP/FAPESP. Professor no Instituto de Educação Continuada da Pontifícia Universidade Católica de Minas Gerais (PUC Minas). Vice-líder do grupo de pesquisa ArC2 – Estudos Antirracistas em Comunicação e Consumos, CNPq/USP. Pesquisador do grupo de estudos “Inteligência Artificial Responsável” (divisão IA, Comunicação e Artes) na Cátedra Oscar Sala, do Instituto de Estudos Avançados da USP. E-mail: leitefco@gmail.com

## Isabela Xavier Gonçalves

Doutoranda em Direito e Tecnologia na Vrije Universiteit Brussel (VUB), mestre em Direito Internacional e Europeu pela Brussels School of Governance (BSoG). Pesquisadora associada do grupo de estudos “Inteligência Artificial Responsável” da Cátedra Oscar Sala, do Instituto de Estudos Avançados da Universidade de São Paulo (USP). E-mail: isabela.xavierg@outlook.com

## Telma Luiza de Azevedo

Membro do grupo de pesquisa FIGAS – Feminismos, Imagens, Gêneros e Artes e da Colaborativa Artística JUnTEs. Doutora em Artes e mestre em Ciências pela Universidade de São Paulo (USP). Formação em Psicologia pela Universidade Estadual de São Paulo (UNESP – Bauru). Pesquisadora associada do grupo de estudos “Inteligência Artificial Responsável”, Cátedra Oscar Sala, do Instituto de Estudos Avançados da USP. E-mail: teluzvdo@alumni.usp.br

**Resumo:** Este artigo de contribuição conceitual, direcionado por uma pesquisa bibliográfica, tem como objetivo refletir e observar criticamente alguns impactos e transformações que a Inteligência Artificial (IA), sustentada por sistemas algorítmicos de aprendizagem de máquina, vem operando nas relações e processos socioculturais no contemporâneo brasileiro. Nessa direção, é intenção também explicar e delinear o conceito de IA responsável (IAR), ressaltando o papel-chave da participação social nos debates e processos de construções regulatórias, que estão abarcando o desenvolvimento, a aplicação e os usos sociais das neotecnologias de IA. Como resultado, se compartilha um quadro informativo (teórico e técnico) que sinaliza e discute, de modo crítico e interdisciplinar, alguns pontos de atenção e recomendações sobre as expressões da IA no Brasil que podem informar e apoiar intervenções e pesquisas futuras sobre as temáticas em foco.

**Palavras-chave:** inteligência artificial responsável; comunicação; participação social; regulação.

## IA responsable, plataformas de redes sociales y regulación en Brasil: notas para pensar la participación social en tiempos de incertidumbre

**Resumen:** A partir de una contribución conceptual y guiado por una investigación bibliográfica, este artículo tiene como objetivo reflexionar y observar críticamente algunos impactos y transformaciones que la inteligencia artificial (IA), apoyada en sistemas algorítmicos de aprendizaje automático, viene operando en las relaciones y procesos socioculturales en el Brasil contemporáneo. En esta dirección, también se pretende explicar y delinear el concepto de IA responsable (IAR), destacando el papel clave de la participación social en los debates y procesos de construcciones regulatorias, que abarcan el desarrollo, la aplicación y los usos sociales de la IA. Como resultado, se comparte un marco informativo (teórico y técnico) que señala y discute, de manera crítica e interdisciplinaria, algunos puntos de atención y recomendaciones sobre las expresiones de la IA en Brasil que pueden informar y apoyar intervenciones e investigaciones futuras sobre el tema en cuestión.

**Palabras clave:** IA responsable; comunicación; participación social; regulación.

## Responsible AI, Social Media Platforms, and Regulation in Brazil: notes to think about social participation in times of uncertainty

**Abstract:** This conceptual study, organized by a literature survey, aims to reflect and critically observe the impacts and transformations artificial intelligence (AI), supported by algorithm systems and machine learning, has brought to sociocultural relations and processes in contemporary Brazil. It also outlines the concept of responsible artificial intelligence, emphasizing the key role of social participation in the regulatory constructions encompassing the development, application, and social uses of AI. As a result, this study provides an informative theoretical and technical framework that signalizes and discusses some points of attention and recommendations on the expressions of AI in Brazil in a critical and interdisciplinary way. These recommendations can inform and support future interventions and research on the topics in focus.

**Keywords:** responsible AI; communication; social participation; regulation.

## Luciana Moherdau

Jornalista. Integra o grupo de pesquisadores de Inteligência Artificial da Cátedra Oscar Sala, do Instituto de Estudos Avançados da Universidade de São Paulo (IEA-USP), tem pós-doutorado pela Faculdade de Arquitetura da USP (FAU-USP), com estudo sobre os impactos das projeções no espaço urbano. É autora de *Guia de estilo web: produção e edição de notícias on-line*, pioneiro no Brasil, e *Jornalismo sem manchete: a implosão da página estática*, ambos publicados pela editora Senac de São Paulo.  
E-mail: luciana.moherdau@gmail.com

## Everaldo Pereira

Doutor em Comunicação Social pela Universidade Metodista de São Paulo, com foco de pesquisa em Comunicação Institucional e de Marketing. É professor e coordenador do curso de Design no Instituto Mauá de Tecnologia. Integra os grupos de pesquisa “LabDesign” no Instituto Mauá, “Inteligência Artificial Responsável” da Cátedra Oscar Sala, no Instituto de Estudos Avançados da USP e “Design Researchers Network” na Universidade de Palermo. É associado à Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação.  
E-mail: everaldo@maua.br

<sup>1</sup> Como explica Domingos, a ideia de *machine learning* (aprendizagem de máquina) “assume muitas formas diferentes e tem muitos nomes: reconhecimento de padrões, modelagem estatística, mineração de dados (*data mining*), descoberta de conhecimento, análise preditiva, ciência de dados, sistemas adaptativos, sistemas auto-organizados e muito mais. Cada um deles é usado por comunidades diferentes e tem associações diferentes. [Uso] o termo *machine learning* para me referir amplamente a todos eles” (Domingos, 2015, p. 8).

<sup>2</sup> Grosso modo, as RNA essencialmente são operações matemáticas com matrizes que usam nós interconectados inspirados em neurônios do cérebro humano.

<sup>3</sup> Para Dignum, na perspectiva sociológica, a agência pode ser compreendida como “a capacidade dos indivíduos de agir de forma independente e de fazer as suas próprias escolhas”. Essa ideia também é central para os estudos da economia, em particular quando se trata da relação entre um agente e o seu principal. O problema do agente principal ocorre quando uma pessoa ou entidade (o “agente”) é capaz de tomar decisões e/ou agir em nome de, ou que impactam, outra pessoa ou entidade: o ‘principal’. Na IA, é particularmente relevante compreender este dilema quando um dos dois participantes é uma máquina inteligente (Dignum, 2019, p. 16).

Este artigo de contribuição conceitual (MacInnis, 2011), direcionado por uma pesquisa bibliográfica, tem como objetivo refletir e observar criticamente alguns reflexos e transformações que a Inteligência Artificial (IA), sustentada por sistemas algorítmicos de aprendizagem de máquina, vem operando nas relações e processos socioculturais no Brasil contemporâneo. Nessa direção, é intenção também explicar o conceito de Inteligência Artificial Responsável (IAR), ressaltando o papel-chave da participação social, como princípio, nos debates civilizatórios e processos de construções regulatórias que estão abarcando o desenvolvimento, a aplicação e os usos sociais da IA no país.

A ideia de IA, minimamente, denota os sistemas ou artefatos construídos por pessoas para cumprir alguns objetivos e ou tarefas determinadas (Dignum, 2020). Em complemento, a partir das ciências computacionais, Baeza-Yates (2023) elucida que a IA pode ser compreendida, grosso modo, como a ciência e engenharia por trás do conceito de fazer máquinas “inteligentes”. No entanto, “o termo em si é muito amplo e abrange múltiplas técnicas, incluindo a mais popular, *machine learning*<sup>1</sup>. Essa que inclui as RNA<sup>2</sup>, que quando contêm muitas camadas são chamadas de *deep learning*” (Baeza-Yates, 2023, p. 407).

Hagerty e Rubinov (2019), a partir de uma perspectiva antropológica, definem a IA como um sistema tecnossocial. Nessa visada, se compreende que “os aspectos técnicos da IA estão intrínsecos e intimamente ligados aos seus aspectos sociais. Os valores e pressupostos sociais moldam a forma como percebemos, concebemos e utilizamos a IA, bem como informam as nossas percepções, expectativas e receios dessas tecnologias” (Hagerty & Rubinov, 2019, p. 3). Em adição, ainda segundo esses autores, a IA é moldada pelo contexto sociocultural em todas as suas fases de desenvolvimento, implementação e usos sociais.

Uma definição de IA bastante acolhida por diversas organizações e governos globalmente vem sendo elaborada e atualizada, de modo constante, pela *Organization for Economic Co-operation and Development* (OECD). Segundo essa entidade, a IA pode ser entendida como

um sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, a partir dos dados que recebe (*input*), como gerar resultados (*outputs*) como previsões, conteúdos, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam nos seus níveis de autonomia e adaptabilidade após a implantação. (Russell et al., 2023)

Com efeito, é ponto de atenção nas reflexões deste artigo, ao implicar a comunicação e cultura, tentar sinalizar como essas ideias se implicam, levando em consideração as intervenções sociotécnicas da IA na agência humana<sup>3</sup>, tendo em vista que essas neotecnologias estão permitindo, progressivamente, que sistemas algorítmicos decidam e ajam com autonomia. Logo, sem controle humano direto, interferindo dessa forma nas mentalidades (efeitos cognitivos e subjetivos), nas formas de decisão e nas práticas socioculturais, políticas e econômicas (Marres, 2017; Dignum, 2020; Pataranutaporn et al., 2023).

É importante salientar que esses sistemas algorítmicos, atualmente, segundo Baeza-Yates (2023), conformam a espinha dorsal da economia digital e dos processos automatizados, sendo a face pública da IA – comércio eletrônico, serviços de *streaming*, *chatbots* e tomada de decisão automatizada, plataformas de mídia social etc. – amplamente utilizada. Esse autor também ressalta que as IA vêm se tornando muito populares, tanto na literatura técnica (foco de pesquisas e intervenções), quanto no imaginário social há pouco mais de dez anos.

<sup>4</sup> Em comunicação pessoal, Almeida (2023) explicou que o “ChatGPT é um grande modelo de linguagem (*Large Language Model* – LLM), com cerca de 175 bilhões de parâmetros, treinado em uma imensa base de dados, capaz de aprender de forma autônoma e produzir textos sofisticados, aparentemente inteligentes. Os parâmetros são chave para os algoritmos de aprendizagem de máquina, pois representam aquilo que é aprendido pelo modelo com os dados de treinamento”. Os LLM são um tipo de modelo de aprendizagem de máquina que processa e gera texto em linguagem natural. Além do ChatGPT, outras ferramentas de IA fazem uso desse modelo, como: assistentes virtuais, *chatbots* ou outros geradores de texto, similares ao ChatGPT, como o Gemini, da Google, lançado em 6 de dezembro de 2023.

<sup>5</sup> Processamento de Linguagem Natural (PNL) é vertente da IA, que usa *machine learning* para ajudar as máquinas computacionais a entender, interpretar e manipular a linguagem humana.

<sup>6</sup> Epistemologicamente, Sodr  distingu  “comum” e “comunidade” destinando “ao primeiro termo o sentido de uma disposi o ontol gica origin ria e inerente   filog nese e   ontog nese do ser humano . . . . A comunidade, por sua vez, n o   a atualiza o institucional desse comum origin rio, mas algo em que sempre estamos na medida em que sempre nos comunicamos, no interior da distribui o dos lugares e das identifica es constitutivas do la o coesivo [o comum]. S o v rias, assim, as formas de comunidade . . . . Cada uma delas resulta de uma subjetiva o que, por sua vez, instaura um novo comum” (Sodr , 2014, p. 399-400).

Essa populariza o se acentuou na consci ncia p blica ap s o lan amento, em 30 de novembro de 2022, do ChatGPT<sup>4</sup>, um *chatbot* alimentado por IA generativa, que utiliza redes neurais para criar conte dos inovadores. Esse sistema de IA interage com humanos e permite obter respostas em linguagem natural<sup>5</sup> (LN) a partir de comandos (*prompts*) escritos. O ChatGPT foi desenvolvido pela empresa estadunidense OpenAI.

Englobando e interpelando esse contexto, este trabalho est  considerando como fontes, especialmente, n o exclusivamente, meta-an lises, revis es e estudos emp ricos recentes, que foram identificados e originados no campo de pesquisa da comunica o, da ci ncia da computa o, sociologia, artes, direito, entre outros. S o utilizados tamb m materiais informativos e t cnicos desenvolvidos por entidades da sociedade civil do Brasil e exterior, bem como proposi es legislativas. Ao reunir e articular intencionalmente essa produ o cient fica e t cnica, este texto constr i a sua fundamenta o integrando descobertas e perspectivas de autoras e autores de diferentes partes do mundo que, de contextos diversos de pesquisa e atua o, do Sul ao Norte Global (Santaella, 2022; Arun, 2020; Hagerty & Rubinov, 2019 etc.), podem falar de uma gama variada de pontos de partida socioculturais, te ricos, conceituais, t cnicos, regulat rios e metodol gicos. A literatura explorada neste texto   publicada em ingl s e portugu s.

Desse modo, este artigo est  organizado da seguinte maneira: em um primeiro momento se explana sobre algumas ideias-chave que envolvem e direcionam o racioc nio aqui proposto como, por exemplo, as no es conceituais de comunica o, *bios* virtual e midiatiza o, estrategicamente enfatizando nesse contexto as plataformas de m dia social, em aproxima o da perspectiva da IA em sua abordagem respons vel. Na sequ ncia, de modo expl cito, s o apresentadas as ideias de intelig ncia artificial respons vel (IAR), bem como as reflex es sobre o papel-chave da participa o social nesse enquadramento. Por fim, considerando as quest es regulat rias, de modo cr tico e propositivo, s o compartilhadas algumas reflex es, caminhos e desafios que est o sendo – e ou podem ser – desdobrados e enfrentados, no contexto brasileiro, para regular o desenvolvimento, implementa o e uso social seguro de uma IA centrada em valores humanos e democr ticos.

### Comunica o, Bios Virtual e IA

Como discutido em outra oportunidade (Leite, 2022), para assimilar a episteme comunicacional no contempor neo, conforme Sodr  (2014, 2021),   preciso prescindir do modelo sociol gico e lingu stico que define a comunica o como processo transmissor de informa es e pens -la a partir de um entendimento  tico-pol tico, que a considera como conex o ou organiza o origin ria do comum, como la o coesivo da comunidade<sup>6</sup>, fruto das experi ncias humanas e da sociabilidade. Dessa forma, como um campo geral dos processos de organiza o e reverbera o de influ ncias no  mbito da coes o social, a “comunica o significa, de fato, em sua radicalidade (assentada na no o latina de *communicatio*), o fazer organizativo das partilhas ou media es imprescind veis ao comum humano; portanto, a resolu o aproximativa das diferen as pertinentes em formas simb licas” (Sodr , 2021).

O saber da comunica o, portanto, pode ser compreendido “como uma ci ncia redescritiva do comum humano, que abrange desde o la o intersubjetivo inerente   coes o comunit ria at  as rela es sociais regidas por m dia [por algoritmos de IA etc.]”, tendo em vista que “a comunica o . . . n o   transmiss o de informa es nem di logo verbal, e sim uma forma modeladora (organiza o de trocas reais) e um processo (a o) de p r diferen as em comum” (Sodr , 2014). Dessarte, conforme Sodr  (2002), o objeto dos estudos da comunica o seria esse comum humano, respectivamente, desenvolvido e implicado pelas seguintes inst ncias ou n veis operativos: 1) relacional, 2) vinculativo e 3) cr tico-cognitivo ou metacr tico. Neste artigo, os dois primeiros n veis s o especialmente considerados.

O n vel relacional denota as “antropot cnicas eticistas ou pr ticas de natureza empresarial (privada ou estatal), voltadas para a rela o ou o contato entre os sujeitos sociais por meio das tecnologias da informa o [e comunica o], como imprensa escrita, r dio, televis o, publicidade [plataformas de m dia social] etc.”

(Sodré, 2002, p. 234). Esse nível abarca “as relações secundárias, rearranjadas e fabricadas por mídia, [é nessa instância, por exemplo, que se localizam a maioria dos] estudos de mídia, da economia interna da comunicação [...]” (Sodré, 2015, p. 124).

É também no nível relacional que se observa as ideias de midiaticização, isto é:

... o funcionamento articulado das tradicionais instituições sociais e dos indivíduos com a mídia. . . A midiaticização é . . . uma elaboração conceitual para dar conta de uma nova instância de orientação da realidade capaz de permear as relações sociais por meio da mídia e constituindo – por meio do desenvolvimento acelerado dos processos de convergência midiática – uma forma virtual ou simulativa de vida, a que já demos nome de *bios* midiático (ou *bios* virtual). (Sodré, 2014)

De modo específico, implicando os algoritmos de IA nesse contexto, Sodré (2023) pontua que no âmbito das neotecnologias, “palavras, sons, imagens e discursos podem ser convertidos em algoritmos, cuja função combinatória é fortemente organizativa sem dependência de processos escritos de significação”. Nesse sentido, “o *bios* midiático revela-se aí não apenas como uma forma de vida entre outras, mas como uma orientação existencial que tenta impor-se em termos universais a reboque do mercado” (Sodré, 2014). Assim, o “*bios* midiático ou *bios* virtual são expressões adequadas para o novo tipo de forma de vida caracterizado por uma realidade ‘imaginarizada’; isto é, feita de fluxos de imagens e dígitos, que reinterpretem continuamente com novos suportes tecnológicos as representações tradicionais do real” (Sodré, 2021).

Nessa linha, compreendendo que essa imaginariade virtual pode ser edificada também pelos algoritmos de IA, é importante ressaltar que, no pensamento de Sodré (2002), a comunicação é maior que a mídia, ou seja, ela não se confina à atividade midiática. Desse modo, abarcando esses sentidos, o nível vinculativo ou da vinculação envolve os laços invisíveis do comum “entre o eu e o outro, logo, a apreensão do ser-em-comum (individual ou coletivo), seja sob a forma de luta social por hegemonia política e econômica, seja sob a forma de empenho ético de reequilíbrio das tensões comunitárias” (Sodré, 2002, p. 223).

Com efeito, como dito, a vincularidade também “não passa necessariamente por mídia, embora a mídia possa estar presente na influência desse vínculo” (Sodré, 2015, p. 123), que é tratado por ela como relacional. Sendo assim, “a mídia trata a externalidade dessa vinculação, que é a relação” (Sodré, 2001, p. 2)<sup>7</sup>. Os sistemas de IA e as suas dinâmicas algorítmicas estão implicados também nesse enquadramento de sentido, como as plataformas de mídia social, tendo em vista que atualmente os algoritmos de IA fornecem deliberadamente o lastro ideológico para o *bios* virtual que envolve a sociedade midiaticizada.

Por fim, o terceiro nível operativo que conforma a ambiência comunicacional, de acordo com a teoria da comunicação sodreriana, é o crítico-cognitivo ou metacrítico. Ele refere-se aos fazeres teóricos pertinentes à:

posição de observação e sistematização das práticas de veiculação e das estratégias de vinculação. Aqui, a Comunicação emerge não como uma disciplina no sentido rigoroso do termo, mas como uma maneira de pôr em perspectiva o saber tradicional sobre a sociedade, portanto, como um *constructum* hipertextual (interface de saberes oriundos de diversos campos científicos) a partir de posições interpretativas. A “ciência” da comunicação impõe-se . . . como uma atividade crítica, só que voltada para a sociabilidade, a eticidade e as práticas de socialização pela cultura . . . (Sodré, 2002, p. 235)

Este artigo é um exemplo de esforço crítico que busca inscrever uma contribuição em linha com as perspectivas observadas nesse terceiro nível metacrítico da comunicação.

Em síntese, é oportuno compreender a comunicação, na contemporaneidade, conforme Sodré, como um novo modo de organização das relações sociais, tendo a vinculação social, – o comum –, como objetos da sua episteme. Nesse sentido,

<sup>7</sup> Para Sodré, “. . . o que a sociologia hoje chama de ‘relação social’ é tão só a superfície visível desse laço ou vínculo, sob o qual se agitam as diferenças e as oposições, a latência das transformações e das passagens de uma forma simbólica a outra” (Sodré, 2014).

a comunicação é “. . . um discurso de organização, reorganização do mundo e da consciência” (Sodré, 2019, p. 880), que vem sendo fortemente impactado e conformado, em todos os seus níveis (cuja a manifestação mais evidente, como dito, é a dimensão midiática), por diversos dispositivos neotecnológicos dirigidos por sistemas de IA. O algoritmo é a linguagem estruturante desses sistemas (Sodré, 2021) como, por exemplo, as plataformas digitais de mídia social que são as mais patentes faces públicas da IA.

### **Plataformas de Mídia Social: Moderação e Amplificação de Conteúdo**

Os sistemas algorítmicos de IA utilizados nas plataformas de mídia social estão tornando cada vez mais fácil manipular informações e a comunicação, influenciando relações, comportamento e decisões de consumo. A opacidade dos sistemas de IA, como se pontuará *a posteriori*, aumenta as preocupações em torno dessas questões, considerando, especialmente, que essas plataformas são propriedades de poucas e grandes empresas globais de tecnologias identificadas como *big techs* e ou *tech giants* (por exemplo: Meta [Facebook], Alphabet [Google], Amazon, Apple, Microsoft etc.).

Atualmente, as tecnologias digitais desempenham um papel cada vez mais vital, facilitando as interações sociais por meio de uma variedade de plataformas. Essas plataformas se autodenominam intermediárias entre indivíduos, instituições, governos e anunciantes. No entanto, as plataformas de mídia social, em particular, não apenas fornecem perfis públicos e *feeds* de conteúdo, mas também interferem na forma e no compartilhamento de informações. Como tal, tornaram-se parte integrante da esfera pública, moldando também os discursos socioculturais e políticos em todos os níveis.

O ato de promover conteúdo online, por exemplo, por meio de plataformas de mídia social, conhecido como *amplificação de conteúdo*, tem se tornado cada vez mais comum nos últimos anos. No entanto, essa prática muitas vezes resulta na disseminação de desinformação e manipulação de informações, o que viola o direito humano de ser informado com precisão. Esse tratamento desigual prejudica o direito à liberdade de expressão não porque determinado conteúdo é removido da plataforma, mas porque as plataformas de mídia social amplificam um conteúdo específico que gera receita enquanto rebaixam outros que são considerados menos valiosos. Isso cria um ambiente online enviesado e injusto para os indivíduos que utilizam esses sistemas.

As plataformas de mídia social contam, como já pontuado, com sistemas algorítmicos de IA para moderar essa visibilidade de conteúdo, permitindo que recomendações rápidas do conteúdo escolhido sejam amplificadas. No entanto, há preocupações em torno do uso de IA e sistemas de aprendizagem de máquina para amplificação de conteúdo. Apesar de serem treinados por vastos conjuntos de dados, esses sistemas, por vezes, não identificam os chamados conteúdos problemáticos ao serem replicados generalizadamente em outros territórios, pois os dados de referência são provenientes de grupos étnico-culturais de localizações específicas do globo. Isso torna a moderação uma tarefa difícil de automatizar, especialmente quando se trata de tipos específicos de conteúdo. Além disso, o treinamento de dados de algoritmos pode suprimir vozes minoritárias (pessoas negras, LGBTQIA+, mulheres, indígenas etc.) nas mídias sociais, já que algoritmos tendenciosos podem impactar negativamente a liberdade de expressão.

Esses sistemas automatizados, usados pelas plataformas de mídia social para facilitar o controle das informações, são frequentemente chamados de “caixas pretas” (Pasquale, 2015), porque usam processos de tomada de decisão opacos que podem ser difíceis de explicar e serem entendidos pelo público geral e até mesmo por seus desenvolvedores. Essa falta de transparência pode levar a regulamentações internas sobre amplificação de conteúdo sem fornecer garantias públicas, deixando os usuários inconscientes das escolhas que afetam suas experiências. Isso levanta questões importantes sobre o poder das plataformas de mídia social para controlar o discurso, permitir resistência e causar exclusão. Também coloca em questão a autoridade das plataformas de mídias sociais sobre as relações e o papel das considerações comerciais na comunicação cotidiana.

Nesse horizonte, ao considerar o impacto das plataformas de mídias sociais nas informações que a sociedade consome, há dois fatores críticos a considerar (Rodotä, 2007). O primeiro é a racionalidade, que envolve a capacidade do indivíduo de compreender e interpretar a realidade para facilitar a análise e a programação (Rose, 1999; Rose & Miller, 2008; Foucault, 2020). O segundo fator é a tecnologia do poder, que engloba o uso de táticas e estratégias para moldar o comportamento e prevenir resultados indesejados (Foucault, 1993). Juntos, esses elementos constituem a *governamentalidade* – a dinâmica de poder em que uma parte emprega métodos particulares para influenciar as ações de outros em direção a um objetivo específico (Burchel *et al.*, 1991).

Como sociedade, as pessoas devem permanecer vigilantes na observação do impacto das tecnologias emergentes em suas vidas diárias, especialmente no caso das plataformas de mídia social. As observações de Delacroix (2019) sobre a capacidade dessas plataformas de regulamentar e manipular a comunicação e o comportamento dos usuários são particularmente dignas de nota. Esses novos poderes regulatórios, como destacado por Winner (1980), sobre o design da tecnologia, podem conceder autoridade e influência aos criadores e operadores dessas plataformas.

Embora os cidadãos agora desfrutem de acesso a grandes quantidades de informações e possam se conectar continuamente pelas tecnologias da internet, é importante avaliar se essa conectividade recém-descoberta os capacita ou apenas os doutrina. A democracia eletrônica deve oferecer um novo espaço público onde os processos democráticos possam se desenvolver (Rodotä, 2007). Nesse sentido, à medida que a tecnologia continua a avançar, torna-se cada vez mais importante integrar a governança da inovação tecnológica em um quadro democrático que leve em conta considerações éticas e não separe as questões técnicas de seu impacto social. A regulação das “tecnocomunicações”, em curso no Brasil e em diversas partes do mundo, deve ser democrática e ética (Rodotä, 2007).

Dessa forma, é pensando a comunicação e as suas expressões nesse contexto (Sodré, 2002) que se reivindica o seu papel-chave, em colaboração e fortalecimento com outros campos do saber como a cultura, a educação, o direito etc., na articulação de intervenções que reivindiquem e assegurem, como princípio e não apenas como fonte de consultas, a participação e diversidade social nos processos de construção e governança da IA.

A expectativa desse contramovimento é buscar mobilizar e garantir que os conhecimentos sobre os impactos positivos e ou negativos dos sistemas de IA sejam difundidos socialmente, bem como capacitar a sociedade a usufruir dos benefícios da IA e, igualmente, se proteger, se posicionar e combater os contextos adversos forjados pelo mau uso dessas neotecnologias. Portanto, o foco é capacitar e “conscientizar as pessoas de que elas podem participar na definição do desenvolvimento social . . . , a fim de garantir que os avanços inevitáveis da IA sejam acessíveis a todos e sirvam para o bem da sociedade” (Dignum, 2020, p. 220).

Estabelecida essa compreensão sobre as dimensões de sentido que envolvem a comunicação, as plataformas de mídia social e IA, se avança com as reflexões deste trabalho para, nos próximos tópicos, explorar as noções de IA em sua abordagem responsável, bem como os pontos de atenção acerca dos desafios e caminhos que podem ser considerados e estimulados frente a tais intervenções sociotécnicas na agência humana.

Nessa direção, também é oportuno ainda ressaltar que é a partir desse ponto de vista sobre as expressões na agência social que as discussões sobre a responsabilidade e questões éticas no contexto dos sistemas de IA vêm ganhando expressão, especialmente, frente aos “dilemas relativos a acesso a dados, viés algorítmico, ética e transparência, e responsabilidade legal por decisões resultantes de IA” (Santaella, 2022, p. 73). Em complemento a essa lista de preocupações, a Academia Brasileira de Ciências, em novembro de 2023, divulgou o documento *Recomendações para o avanço da Inteligência Artificial no Brasil*, no qual, entre outros pontos, ao reconhecer os potenciais benefícios e oportunidades da IA, chama atenção especial

para os danos e riscos éticos e sociais evidentes que esses sistemas podem trazer para indivíduos, grupos, sociedades e para o planeta.

Entre esses dilemas a publicação destaca: violações de privacidade, criação de ambientes anticompetitivos, manipulação de comportamentos e ocorrência de desastres ambientais, viés algorítmico de opressão racista e outras formas de discriminação, efeitos diretos em vários perfis de trabalhadores (por exemplo no monitoramento de sua atuação e automação excessiva), entre outras. Esses dilemas vêm enredando a todos, especificamente, tendo em vista a capacidade, direta ou indireta, que sistemas algorítmicos de IA têm de interferir e influenciar os processos e as relações socioculturais (Cordeiro, 2021), como já pontuado.

Portanto, é na emergência desses debates e embates acerca tanto dos potenciais benefícios e oportunidades, quanto dos riscos e danos que as tecnologias de IA podem oferecer para indivíduos, grupos, sociedades e para o planeta, que a ideia de inteligência artificial responsável vem se difundindo e se estabelecendo como possibilidade de intervenção crítica nesse horizonte de incertezas que enreda a todos.

### **Inteligência Artificial Responsável**

O conceito de inteligência artificial responsável (IAR), segundo Brown (2023), denota um conjunto de princípios, políticas, ferramentas e processos para garantir que os sistemas de IA sejam desenvolvidos e operados a serviço do bem social, ao mesmo tempo em que alcançam um impacto transformador nos negócios. Por sua vez, Baeza-Yates (2023) observa que o objetivo da IAR é criar sistemas que beneficiem os indivíduos, as sociedades e o meio ambiente. Ela abrange todos os aspectos éticos, legais e técnicos ao longo do desenvolvimento (análise, projeto, construção, implantação e avaliação) de tecnologias benéficas de IA, bem como busca garantir que esses sistemas não interfiram na agência humana, não causem riscos e danos, não discriminem/oprimam ou desperdicem recursos ambientais.

Essa abordagem responsável da IA deve ser suportada, segundo Dignum (2017, 2019, 2020), por atributos éticos e valores humanos que, minimamente, reflitam os seguintes princípios básicos: 1) responsabilidade, 2) transparência e 3) *accountability* (prestação de contas). Para Dignum (2020), tais princípios também devem enquadrar todo o sistema sociotécnico da IA. Isto é, a condução desses princípios orienta um tratamento sociotécnico para “o projeto, a implantação e o uso de sistemas, entrelaçando soluções de *software* com governança e regulação. Além disso, embora cada um desses princípios possa ser aplicado a todos os aspectos dos sistemas de IA, cada um é imperativo para uma característica específica” (Dignum, 2020, p. 408). Logo, ainda de acordo com a autora (2020), não se pode ter autonomia dos sistemas de IA sem alguma forma de responsabilidade, interação sem *accountability* ou adaptabilidade sem transparência.

Mas frente a esse cenário de inquietudes e incertezas, como o poder público brasileiro vem se movimentando para regular o desenvolvimento e o uso da IA no país, em diálogo, com as iniciativas internacionais? Como a participação social está sendo ou não abarcada nesse contexto?

Para o exercício de articular orientações de respostas para tais questões, é frutífero inscrevermos inicialmente algumas reflexões-chave acerca da *Artificial Intelligence Act* (AI ACT – Lei de IA), a primeira regulação geral sobre a IA no mundo, aprovada pela União Europeia, em 2024. Estabelecido esse breve contexto, posteriormente, considerando o cenário brasileiro, são observados alguns projetos de leis e resoluções, direcionados a preencher essa lacuna, que estão em trâmites nos poderes legislativo e judiciário.

Nessa direção, a seguir, de modo introdutório, algumas dessas iniciativas são apresentadas para se entender o cenário regulatório da IA no país e refletir sobre os seus impactos sociais. Entre as proposições regulatórias destacam-se neste texto: o Projeto de Lei (PL) nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial (atualmente em trâmite no Senado Federal) e as normativas aprovadas

<sup>8</sup> Em 24 de abril de 2024, (24), a Comissão Temporária sobre Inteligência Artificial (CTIA) do Senado Federal apresentou o texto preliminar de um substitutivo para o PL 2338/2023. Conforme o senador Eduardo Gomes (PL-TO), relator da comissão, a expectativa é que o marco regulatório brasileiro de IA seja aprovado até o final de 2024. “Uma das principais novidades do documento é a proposta de um sistema híbrido de fiscalização, com a criação de um Sistema Nacional de Regulação e Governança de IA (SIA), coordenada por uma autoridade competente, a ser nomeada pelo poder executivo, mas que conte também com a participação das agências e órgãos reguladores existentes, que ficarão responsáveis pela fiscalização da IA em suas respectivas áreas de atuação. . . . Outra novidade do projeto foi a adição de um capítulo que trata sobre o fomento público de IA. O trecho cria diretrizes programáticas para que o poder público invista em pesquisa e educação, sem desconsiderar o impacto ambiental. Também está prevista a criação de mecanismos de certificação de sistemas, como a criação de selos de IA segura, para incentivar e assegurar boas práticas de governança da tecnologia” (Soares, 2024).

<sup>9</sup> European Commission (2024a).

<sup>10</sup> European Commission (2024b).

<sup>11</sup> European Commission (2019).

<sup>12</sup> European Commission (2019).

recentemente pelo Tribunal Superior Eleitoral (TSE), em 27 de fevereiro de 2024, que regulamentam o uso da IA nas eleições brasileiras<sup>8</sup>.

Nesta oportunidade, é digno de nota reconhecer os esforços e debates investidos para a articulação do PL nº 2630/2020. Esse documento tinha como foco instituir a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Contudo, em 09 de abril de 2024, o presidente da Câmara dos deputados, Arthur Lira, comunicou, justificando o não consenso nos debates sobre o seu conteúdo, que o referido PL, também popularmente conhecido como PL das *fake news*, deverá ser reescrito e ter novo relator (Moherdau, 2024a).

### A Lei da IA Estabelecida pela União Europeia

No dia 12 de julho de 2024, o Conselho da União Europeia aprovou e publicou no seu Jornal Oficial a primeira lei abrangente de IA do mundo, a *Artificial Intelligence Act (AI Act – Lei da Inteligência Artificial)*<sup>9</sup>, com o objetivo de harmonizar as regras sobre inteligência artificial na UE. A *AI Act* foi formulada com base no artigo 114 do Tratado sobre o Funcionamento da União Europeia (TFUE) e visa melhorar o funcionamento do mercado interno por meio do estabelecimento de regras sobre o desenvolvimento, comercialização, venda e a utilização de produtos e serviços que empreguem tecnologias de IA.

As discussões sobre a necessidade de criação de regras para regular a IA na UE começaram em 2018, momento em que a Comissão Europeia lançou a sua estratégia de IA, que conduziu a criação do *High-Level Expert Group (HLEG)*<sup>10</sup>, o qual apresentou suas diretrizes éticas para um Sistema de IA Confiável, que nos referimos neste artigo, como IA Responsável (IAR). De acordo com as diretrizes do *HLEG*, uma IAR deve ser “(1) regulada – respeitando todas as leis e regulamentos aplicáveis; (2) ética – respeitando os princípios e valores éticos; e (3) robusta – tanto do ponto de vista técnico quanto levando em consideração o seu ambiente de aplicação”<sup>11</sup>.

A Comissão propôs uma abordagem que coloca as pessoas no centro do desenvolvimento da IA (IA centrada no ser humano [Comissão Europeia, 2018]). Essa estratégia apoia uma IA ética, com base em sete diretrizes: 1) ação e supervisão humanas; 2) solidez técnica e segurança; 3) privacidade e governança dos dados; 4) transparência; 5) diversidade, não discriminação e equidade; 6) bem-estar ambiental e social; e 7) responsabilização<sup>12</sup>.

O artigo 3 (1) da *AI Act* define um sistema de IA como um sistema baseado em aprendizado de máquina para funcionar com diferentes níveis de autonomia. Além disso, o sistema pode produzir resultados como previsões, decisões, conteúdo ou recomendações que podem afetar o ambiente físico ou online, de acordo com as metas definidas por humanos. Essa definição abrange os sistemas inteligentes utilizados pelas plataformas digitais, para amplificar, ordenar e moderar informação, na medida em que envolve o uso de algoritmos de aprendizado de máquina que produzem saídas destinadas a oferecer recomendações ou previsões, que podem manipular opiniões, pensamentos e comportamentos. Portanto, as regras contidas na *AI Act* devem ser obedecidas pelas plataformas digitais, quando utilizarem sistemas de IA para moderar e amplificar conteúdo dentro da UE.

Para garantir a transparência no funcionamento dos sistemas de IA, o Artigo 9 da *AI Act* exige uma avaliação de risco dos sistemas de IA de alto risco. Essa avaliação visa identificar e avaliar riscos potenciais à saúde, segurança e direitos fundamentais associados a sistemas de IA de alto risco. Essas exigências abrangem mais do que apenas a proteção do consumidor e o direito à privacidade, estendendo-se a projetos algorítmicos que resultem em problemas comportamentais, como vício e perda de controle do uso de mídias sociais, por exemplo.

Além das obrigações de transparência<sup>13</sup>, os sistemas de IA sob a *AI Act* estão sujeitos a proibições que visam proteger os indivíduos contra a manipulação algorítmica. A Lei proíbe o uso de sistemas de IA que utilizem “técnicas subliminares” ou técnicas intencionalmente manipuladoras ou enganosas além da consciência de uma pessoa,

<sup>13</sup> Ver, também, o artigo 13 da *AI Act*.

com o objetivo ou efeito de distorcer significativamente o comportamento de um indivíduo ou grupo<sup>14</sup>.

A proibição de técnicas manipuladoras, enganosas ou subliminares que operem além da consciência de uma pessoa é especialmente significativa no contexto da manipulação de informações e da comunicação no ambiente online. Essas técnicas são frequentemente utilizadas para persuadir os usuários a tomarem decisões que talvez não tivessem tomado se não manipulados por algoritmos treinados para tanto.

Porém, as formas sofisticadas e diferenciadas de influenciar os usuários com a ajuda da IA tornam extremamente desafiador implementar regras que possam abordar e regular adequadamente as tecnologias utilizadas para curadoria das informações online, o que pode implicar desafios para o “enforcement” da *AI Act*, principalmente para proibir sistema de IA que empreguem técnicas subliminares.

### Efeito Bruxelas da AI Act no Brasil e o Projeto de Lei nº 2338/2023

A regulação da IA na UE já está impactando o mundo todo, incluindo o Brasil. Esse fenômeno é denominado de Efeito Bruxelas<sup>15</sup>. A proposta de Lei de IA brasileira adota uma abordagem baseada em risco que está alinhada com a *AI Act*. O Projeto de Lei nº 2.338/2023 (PL 2338/2023) estabelece regras nacionais para o desenvolvimento, implementação e, principalmente, utilização responsável de sistemas de IA no Brasil. Seu objetivo central é salvaguardar os direitos fundamentais e garantir a adoção de sistemas confiáveis e seguros que atendam aos melhores interesses dos indivíduos, à democracia e ao avanço científico e tecnológico. Da mesma forma como a lei da UE para a IA, as regras do PL 2338/2023 também se aplicam aos sistemas de IA que organizam, priorizam e recomendam informação online.

Em resumo, o PL 2338/2023 estabelece, em primeiro lugar, assegurar os direitos da parte mais vulnerável, ou seja, o indivíduo impactado pelos sistemas de IA. Isso implica protegê-lo contra publicidade direcionada, recomendações de conteúdo e avaliações de sua elegibilidade para crédito e políticas públicas. Em segundo lugar, o projeto se esforça para estabelecer instrumentos de governança eficazes e arranjos institucionais para inspeção e supervisão.

Desde o início das discussões sobre a regulação da IA no Brasil, em meados de maio de 2023, o texto legislativo sofreu várias mudanças significativas. A última redação do PL 2338/2023, de 28 de novembro de 2024, manteve, de forma geral, algumas das regras de transparência previamente estabelecidas. Dentre as alterações percebidas, as regras de transparência, antes concentradas no artigo 5º, da versão preliminar do PL 2338/2023, de 2023, foram divididas em duas seções: “Seção I: Dos Direitos da Pessoa e Grupos Afetados por Sistema de IA” (Artigo 5º) e a “Seção II: Dos Direitos da Pessoa e Grupos Afetados por Sistema de IA que Produz Efeitos Jurídicos Relevantes ou de Alto Risco” (Artigo 6º).

Isso significa que, antes dessa alteração no texto legal, todas as pessoas e grupos afetados por qualquer tipo de sistemas de IA deveriam receber uma explicação para quaisquer decisões, recomendações ou previsões feitas por esses sistemas inteligentes. A explicação deveria incluir detalhes sobre os critérios e procedimentos utilizados, bem como os principais fatores que influenciavam a decisão ou previsão, como informações sobre a racionalidade e lógica do sistema e os resultados esperados da decisão (Projeto de Lei nº 2338, 2023, art. 8º).

Os usuários também teriam o direito de saber a extensão da contribuição feita pelo sistema de IA para o processo de tomada de decisão, os dados processados e sua fonte, os critérios para a tomada de decisão e a ponderação aplicada à situação do usuário afetado pelo sistema (Projeto de Lei nº 2338, 2023, art. 8º, II). Essas regras, mesmo que mais custosas de cumprimento, aparentemente garantiam mais explicabilidade aos usuários dos sistemas de IA, até certo ponto, do que as regras de transparência da *AI Act*, por exigirem mais explicações e por sua aplicabilidade estender-se a todos os sistemas de IA e não somente àqueles considerados de alto risco.

<sup>14</sup> Ver artigo art. 5º, 1, (a), da *AI Act*.

<sup>15</sup> O conceito “efeito Bruxelas” foi introduzido por Bradford, em 2012, referindo-se ao impacto inadvertido da UE no estabelecimento e aplicação global de regulamentos. Segundo ela, a UE influenciou a adoção das suas regras por corporações globais e estados transfronteiriços, resultando em duas categorias de Efeito Bruxelas: “Efeito Bruxelas de fato” e “Efeito Bruxelas de jure”. O primeiro denota a integração de regulamentos da UE por corporações globais, enquanto o último envolve a implementação de “regulamentos de estilo da UE” por estados transfronteiriços. Em 2016, a UE estabeleceu padrões e normas internas de proteção de dados pessoais, que resultaram na criação do GDPR. Esse regulamento é amplamente considerado como a referência para as leis de proteção de dados e influenciou políticas em todo o mundo, inclusive nos EUA e na China. Além disso, o GDPR foi adotado como um regulamento padrão em vários países, como Brasil, Japão, Quênia, Coreia do Sul e Nigéria. A nova onda do Efeito Bruxelas pode ser vista com a aplicação das novas regras de IA da UE em outros países, como o Brasil (Bradford, 2020).

A nova proposta legislativa brasileira alinha-se à *AI Act* para determinar que somente as pessoas afetadas por Sistema de IA de Alto Risco, acrescentando aqueles que produzirem efeitos Jurídicos Relevantes, poderão solicitar informações limitadas e tão somente ao direito à explicação sobre a decisão, recomendação ou previsão feitas pelo sistema (Projeto de Lei nº 2338, 2023, art. 6º, I). Essa mudança no texto legal compromete a efetividade da transparência para que os usuários desses sistemas possam entender como a curadoria das informações realizadas pelas mídias sociais afetam suas opiniões, pensamentos e comportamentos.

Essa divisão nas regras de transparência se justifica, talvez, pela inclusão do inciso XIII no artigo 14 (que deve, portanto, ser lido e interpretado em conjunto com o artigo 6º, I da nova redação do PL 2338/2023), que, diferente da proposta aprovada pela UE<sup>16</sup>, classifica o uso de sistemas de IA por provedores de aplicação, para produção, curadoria, difusão, recomendação e distribuição, como um sistema de IA de alto risco. Essa consideração é crucial para a preservação dos direitos humanos e da democracia no Brasil, particularmente após os eventos antidemocráticos ocorridos e os impactos negativos dos sistemas de moderação e recomendação de IA sobre os indivíduos brasileiros, principalmente no que se refere à disseminação de desinformação em ambiente online.

Porém, vale ressaltar que a disposição traz uma ressalva. Esses sistemas somente serão considerados de alto risco se aplicados em grande escala e com uso significativamente automatizado, com o objetivo de maximização do tempo de uso e engajamento das pessoas ou grupos afetados, e somente quando o funcionamento desses sistemas puder representar riscos relevantes à liberdade de expressão e acesso à informação e aos demais direitos fundamentais. (*Projeto de Lei nº 2338, 2023, art. 14, XIII*).

Isso implica dizer que esses sistemas não são, por natureza, considerados, pela lei, de alto risco, mas somente quando essa produção, curadoria, recomendação e distribuição de conteúdo não passar por revisão humana e impactar negativamente de forma relevante um volume considerado alto de usuários. A redação parece, portanto, sugerir o reconhecimento de impacto por meio do alcance em grande escala. Porém, verifica-se que o impacto não deve ser medido somente pelo alcance de pessoas que aquele conteúdo prejudicial atingiu. Estudos (Amnesty Internacional, 2023) e casos reais<sup>17</sup> demonstram que a recomendação de conteúdo direcionada possui um impacto significativamente negativo para os usuários das plataformas de mídias sociais.

Além disso, a consideração limitada da finalidade desses sistemas somente para a maximização do tempo de uso e engajamento, também prejudica a classificação de alguns sistemas de IA, utilizados para recomendação direcionada, como de alto risco. Isso se torna preocupante, porque os sistemas de recomendação podem se tornar problemáticos por atuarem como “*sticky traps*” [armadilhas pegajosas], com o objetivo de manipular as decisões e motivações dos usuários (Milano, Taddeo & Floridi, 2020, p. 962). Essa manipulação pode criar “bolhas de filtro” que limitam a exposição dos usuários a uma gama maior de informações (Pariser, 2011; Bruns, 2019). Os usuários podem nem perceber que seu conteúdo recomendado é restrito, o que pode causar incerteza na tomada de decisões. Essa situação torna difícil para os indivíduos compreenderem toda a extensão de suas opções dentro do espaço recomendado, modificando seu comportamento (Jannach & Adomavicius, 2016).

Quando um sistema de IA, dentro das regras estabelecidas pelo PL 2338/2023, não é considerado de alto risco, regras mais robustas de transparência, mitigação de riscos e avaliação de impacto sobre os direitos fundamentais, como as determinadas pelos artigos 18 e 25, da última proposta legislativa para regular a IA no Brasil, não são exigidas. A exemplo dos casos jurídicos estadunidenses, como o caso *Gonzalez vs. Google e Twitter v. Taamneh* (“Twitter, Inc.”, 2023), além das mais variadas pesquisas envolvendo estudos sobre o impacto negativo da recomendação de conteúdo, e os depoimentos de vários ex-funcionários das *big techs*, como a Frances Haugen, esses sistemas são projetados, desde a sua concepção, para manipular comportamentos.

<sup>16</sup> Ver Artigo 6º e, em complemento, o Anexo III, da *AI Act*.

<sup>17</sup> Para saber mais, ver Goggin (2024) e Carville (2023).

Além disso, muitas plataformas evitam assumir a responsabilidade em recomendar conteúdo, baseando suas desculpas no comportamento do usuário no ambiente social, alegando que essas pessoas têm sua própria agência ao interagir online (Facebook, 2020). Leiser (2016) sugere que isso ocorre porque os usuários não são apenas “passive dots”, mas agentes ativos em suas atividades na internet. No entanto, essas plataformas limitam as opções dos utilizadores, especialmente quando se trata de receber recomendações personalizadas, ou culpam o próprio algoritmo de ter tomado a decisão sem intervenção humana.

Dessa forma, a aplicação de regras mais robustas de transparência, mitigação de riscos e avaliação de impacto são cruciais para evitar os efeitos adversos desses sistemas de IA. Nesse sentido, aparentemente o legislador escolheu agir de forma reativa, em vez de proativa, o que pode contribuir para uma aplicação ineficaz das regras da lei brasileira para regular sistemas de IA utilizados com finalidades e contextos em ambientes sociais online.

Pensando em preencher essa lacuna, o legislador introduziu o artigo 13, incisos I e II, que, respectivamente, proíbem sistemas de IA que possam “induzir o comportamento da pessoa natural ou de grupos de maneira que cause danos à saúde, segurança ou outros direitos fundamentais próprios ou de terceiros”, ou “explorar quaisquer vulnerabilidades de pessoa natural ou de grupos com o objetivo ou o efeito de induzir o seu comportamento de maneira que cause danos à saúde, segurança ou outros direitos fundamentais próprios ou de terceiros.”

No entanto, interpretando em conjunto e de forma comparativa os artigos 13, incisos I e II, e o 14, inciso XIII, considerando que para classificar um sistema de IA, utilizado por plataforma digital para moderar, amplificar e recomendar conteúdo, há a necessidade de comprovação de representação de riscos relevantes à liberdade de expressão e acesso à informação e aos demais direitos fundamentais, esses sistemas seriam proibidos de utilização. Essa análise incorre o risco desses sistemas de IA, utilizados em contextos de moderação e amplificação de conteúdos, manterem-se desregulados pelas múltiplas possibilidades de interpretação do texto legal, ao mesmo tempo em que não necessariamente são considerados sistemas de alto risco ou proibidos por lei.

Ainda, em análise ao inciso XIII do artigo 14, verifica-se um ganho significativo para os cidadãos nacionais, em que pese a vantagem do legislador brasileiro de aproveitar as redações dos regulamentos já existentes na União Europeia, nos Estados Unidos e no Reino Unido; e os estudos já realizados para a elaboração de um texto mais eficaz, com a retirada do termo “técnicas subliminares” do inciso I do artigo 13, da nova proposta legislativa. Essa mudança é expressivamente importante, porque estudiosos, durante o processo legislativo da *AI Act*, já levantavam preocupações sobre a utilização do termo “técnicas subliminares” para a proibição de sistemas de IA manipuladores (Franklin et al., 2022). Essa definição foi mantida na lei europeia<sup>18</sup>.

<sup>18</sup> Ver artigo 5º, 1 (a), da *AI Act*.

As críticas, em torno do tema, sugerem que as tecnologias subliminares devem ser substituídas por uma gama mais ampla de técnicas de manipulação (Uuk, 2022). Isso porque as práticas de mudança de comportamento são manipuladoras quando não se envolvem com a capacidade de escolha das pessoas. No entanto, nem todos os aspectos da arquitetura de escolha são igualmente influentes (Mertens et al., 2021). Isso levanta questões sobre quais aspectos da arquitetura de escolha são manipuladores, e, portanto, dignos de proibição (Franklin et al., 2022). Além disso, uma vez que a noção de “técnicas subliminares” não é um conceito bem estabelecido na lei de IA da União Europeia, muitos sistemas de IA manipuladores podem ficar fora do escopo de proibição da *AI Act*.

Por derradeiro, a última proposta legislativa brasileira retirou uma disposição dedicada ao exercício do direito à não discriminação e à correção de vieses discriminatórios diretos, indiretos, ilegais ou abusivos (*Projeto de Lei n° 2338*, 2023, art. 12). Essa disposição incluía o uso de dados pessoais sensíveis ou impactos desproporcionais com base em características pessoais, como raça, etnia, gênero, orientação sexual, classe socioeconômica, idade, deficiência, religião ou opiniões políticas (*Projeto de Lei n° 2338*, 2023, art. 12, I). O projeto também proibia o

estabelecimento de desvantagens ou agravamento da situação de vulnerabilidade de pessoas pertencentes a um grupo específico, ainda que fossem utilizados critérios aparentemente neutros (*Projeto de Lei n° 2338, 2023, art. 12, II*).

Ainda que a nova redação do PL 2338/2023, de 28 de novembro de 2024, traga disposições legais para a proteção contra a discriminação direta, indireta, ilegal ou abusiva, mais uma vez, as regras para sua mitigação serão apenas consideradas para sistemas de IA classificados como de alto risco. Em que pese os altos índices, de caráter estrutural e sistêmico, de discriminação no Brasil, a adoção de uma postura preventiva para a proteção dos direitos fundamentais individuais dos brasileiros, e para a construção de uma sociedade mais igualitária, seria o ideal. As decisões automáticas, muitas vezes opacas, podem ter efeitos discriminatórios, por exemplo, quando um sistema de IA aprende com dados que refletem decisões humanas tendenciosas.

Além disso, o Brasil como utilizador, em sua maioria, de tecnologia, e não como desenvolvedor, a importa de países onde aspectos discriminatórios não são tão evidentes. Com isso, as ferramentas estrangeiras que não são desenvolvidas com considerações, em primeiro lugar, antidiscriminatórias são, muitas vezes, implantadas em solo brasileiro, exacerbando a segregação já existente no país.

Diante disso, os instrumentos jurídicos mais relevantes para mitigar os riscos de discriminação, impulsionada por IA, são leis de não discriminação e de proteção de dados (Heinrichs, 2021). As características intrínsecas dessas leis são a adoção de (a) uma postura proativa, em vez de reativa, para antecipação do risco para sua prevenção, ou seja, a incorporação de práticas não discriminatórias ao design e na arquitetura de desenvolvimento do sistema desde sua concepção (*by design*); e (b) como parte inerente do sistema, produto ou tecnologia, por padrão (*by default*). Esses princípios podem ser vistos na Lei Geral de Proteção de Dados (LGPD) brasileira. A mesma lógica se aplicaria para as regras de transparências mencionadas anteriormente.

### O Uso de IA nas Eleições Brasileiras

O poder judiciário brasileiro também debateu em audiência pública no Supremo Tribunal Eleitoral (TSE), em 25 de janeiro de 2024, uma resolução que exigia informar material adulterado e responsabilizava as *big techs* pela moderação de conteúdo falso ou “gravemente descontextualizado”<sup>19</sup>. Essas e outras propostas regulatórias, recentemente, foram discutidas e aprovadas pelo TSE, em 27 de fevereiro de 2024. Entre as diretivas aprovadas estão:

Exigência de rótulos de identificação de conteúdo multimídia fabricado – qualquer material visual feito por meio de inteligência artificial deverá trazer o aviso explícito sobre o uso da tecnologia; Restrição ao uso de *chatbots* e avatares para intermediar a comunicação da campanha – fica proibido simular conversas com o candidato ou outro avatar que aparente ser uma pessoa real; Vedação absoluta, seja contra ou a favor de candidato, do uso de *deep fake* – conteúdo fabricado em formato de áudio, vídeo ou combinação de ambos e que tenha sido gerado ou manipulado digitalmente para criar, substituir ou alterar imagem ou voz de pessoa viva, falecida ou fictícia; Paralelamente, os provedores de aplicações na internet (redes sociais e aplicativos de mensagem, por exemplo) ficam obrigados a retirar do ar, sem a necessidade de ordem judicial, contas e materiais que promovam condutas e atos antidemocráticos e também discursos de ódio, como racismo, homofobia, fascismo e qualquer tipo de preconceito. (Pontes, 2024)

Resoluções como as do TSE têm efeito paliativo desde a última campanha eleitoral, em que o presidente Luiz Inácio Lula da Silva venceu o candidato à reeleição Jair Bolsonaro. Essas normativas visam, grosso modo, combater a desinformação e regular as práticas de candidatos e partidos políticos no contexto eleitoral. Embora estudos, como o *Detecting misinformation: Identifying false news spread by political leaders in the Global South*, indiquem que somente 1% das publicações de políticos brasileiros contêm desinformação (Mello, 2024a), o Parlamento precisa avançar. Não se trata, contudo, da quantidade de *posts* nas plataformas de mídias sociais,

<sup>19</sup> Resolução n° 23.732, de 27 de fevereiro de 2024 (2024).

mas de seus efeitos, de seus alcances, viabilizados pelos sistemas de IA oferecidos pelas plataformas de mídias sociais das *big techs*.

Na invasão ao Capitólio, em 6 de janeiro de 2021, nos Estados Unidos, foi um tuíte (o nome da rede do bilionário Elon Musk ainda não havia mudado para X) do então presidente Donald Trump que incitou apoiadores a impedir a diplomação de Joe Biden, eleito para comandar o país: “Grande protesto em D. C. no dia 6 de janeiro. Esteja lá, vai ser selvagem”, escreveu em 19 de dezembro de 2020<sup>20</sup>. Nesse contexto, o Brasil, citado no relatório *Digital Risks to the 2024 Elections: Safeguarding Democracy in the Era of Disinformation* (Barrett et al., 2024), é indicado como um exemplo do fracasso das plataformas na moderação de conteúdo que levou aos ataques antidemocráticos ao Palácio do Planalto, Supremo Tribunal Federal (STF) e Congresso Nacional, em 8 de janeiro de 2022, após a diplomação do presidente Lula.

<sup>20</sup> Ver Trump (2020).

Portanto, a reflexão para elaborar regramentos em campanhas eleitorais orientadas por IA tem de, obrigatoriamente, observar as características territoriais locais e não detalhar as normas de modo a se tornarem obsoletas, a exemplo do TSE – uma resolução a cada nova tecnologia. Em vez disso, conforme Barrett et al. (2024), é preciso efetivamente pensar sobre como abordar juridicamente os métodos bem estabelecidos de distribuição de material digital prejudicial. Assim, o:

principal risco relacionado com a tecnologia que as próximas eleições enfrentam não decorre de novas formas como os malfeitores podem utilizar a IA para criar conteúdos *online* prejudiciais. Em vez disso, **provém de métodos bem estabelecidos de distribuição de material digital prejudicial** [ênfase adicionada] – nomeadamente, através das principais plataformas de redes sociais como Facebook, Instagram, YouTube, TikTok e X (anteriormente conhecido como Twitter). (Barrett et al., 2024, p. 5)

Com efeito, ao observar de modo crítico os movimentos regulatórios da IA, especificamente no Brasil, é possível perceber fortemente que o país está buscando construir essa regulação levando em consideração apenas as questões políticas. Nesse contexto, é preciso observar e discutir que a regulação da IA, bem como das plataformas de mídia social, vai muito além da regulação para as eleições, ou apenas com vieses políticos. É urgente construir no cenário brasileiro uma legislação que, com participação social, foque regular a moderação e a amplificação de conteúdo, estabelecendo explicitamente direitos e deveres para os usuários e para as plataformas proprietárias da IA.

Ainda no enquadramento desse cenário, é verdade que as citações do impacto de um tuíte de Donald Trump ou da organização em rede de apoiadores de Bolsonaro em invasões e nas depredações do Capitólio e em prédios da Esplanada dos Ministérios, conforme relatado *a priori* neste texto, não são observadas do ponto de vista da produção de conteúdo eleitoral, viabilizado pelas neotecnologias da IA generativa, mas de exibição, distribuição e alcance, tão ou mais danosa que a criação. Esses processos impõem reflexões em mais de uma abordagem das novas tecnologias. Não é sem razão que o *IA Act*, as normativas do TSE e a proposta de regulação de IA em tramitação no Congresso incluem exibição, distribuição e alcance e exigem transparência das plataformas sociais. Há, inclusive, levantamento recente no qual comprova que a desplataformização tem efeito expressivo na redução de postagens falsas e incitação à violência. Após o caso do Capitólio, por exemplo, a suspensão de contas do ex-presidente estadunidense e de 70 mil apoiadores reduziu os danos on-line, de acordo com McCabe et al. (2024), em publicação na revista *Nature*.

Quanto ao conteúdo (texto, foto, vídeo e áudio, entre outros) desenvolvido por IA generativa, muito se vem discutindo por pesquisadores e especialistas em todo mundo acerca dos seus efeitos em eleições como, por exemplo, as repercussões desses conteúdos em pleitos na Índia, nos Estados Unidos e no Brasil (“Many AI”, 2024; Stokel-Walker, 2024; Chhabria, 2024; Garimella & Chauchard, 2024; Moherdau. 2024b etc.). Apesar de contribuir para campanhas eleitorais menores, com baixo orçamento, ajudar na redação de e-mails, pedidos de doações e discursos, uma preocupação constante acerca da IA generativa é a sua aplicação em larga escala para forjar voz de candidato, espalhar montagens de fotos e vídeos e espalhar falsidades

por textos, uma vez que a resposta desse sistema algorítmico é apresentada rapidamente e ainda não é possível mensurar matematicamente as consequências – influência de voto e resultado da disputa.

Nos Estados Unidos, candidatos usaram IA para traduzir discursos em inglês para outros idiomas, como o espanhol, de modo a atingir um número maior de pessoas. Mas também houve quem se passou pelo candidato à reeleição daquele país, o democrata Joe Biden, para espalhar desinformação. Decreto de Biden orienta candidatos a aplicar marcas d'água a materiais feitos com IA. Na Índia, anunciantes fantasmas e de aluguel, *deep fakes*, premiação de youtubers governistas e armação de debates com falsos eleitores comuns foram algumas das técnicas utilizadas este ano, ainda que sejam limitados os estudos relacionados às implicações (McCabe et al., 2024).

Muitas das táticas dos indianos – grupos de WhatsApp para disseminar desinformação eleitoral, aplicativos de candidatos que coletam dados dos eleitores e segmentação de mensagens – foram usadas nas eleições do Brasil e dos EUA. . . . A comissão eleitoral instruiu os partidos a removerem os deep fakes em até três horas após notificação das autoridades, mas a determinação não tem poder de lei. (Mello, 2024b)

No Brasil, há obrigatoriedade de alertar, sob pena de a candidatura ser cassada, em caso de distribuição de *deep fake* com desinformação eleitoral. Gullino (2024), em reportagem de O Globo, mostra que a Justiça começou a punir o uso irregular de inteligência artificial (IA) na pré-campanha das eleições municipais do segundo semestre de 2024. De acordo com o jornal, a medida é inédita, uma vez que tais decisões traçam um limite do que poderá ser feito:

Juízes de ao menos quatro tribunais regionais eleitorais já aplicaram multas ou determinaram a remoção de conteúdos, a maioria dos casos por uso de *deep fakes*, quando imagens ou áudios são manipulados para inventar declarações ou situações. As decisões seguem o entendimento do Tribunal Superior Eleitoral (TSE), que aprovou regras para o uso da IA. (Gullino, 2024)

Posto isso, de modo geral, o atual estado da IA no Brasil reflete desafios significativos relacionados à qualidade dos dados, autonomia da IA, competição global etc. Estratégias para melhorar a organização de dados, promover uma abordagem responsável à IA e enfrentar os desafios competitivos devem ser considerados de forma integrada.

A transparência, colaboração e reflexão filosófica sobre o papel da IA na sociedade, bem como a fundamental participação social no desenvolvimento e usos da IA são princípios centrais para orientar o desenvolvimento, o emprego e utilização dessas neotecnologias, tendo em vista que apenas as necessárias definições legislativas observadas podem não ser suficientes para assegurar a aplicabilidade, sustentabilidade e os impactos sociais benéficos de uma IA. Mas como envolver e fomentar a participação social nesse cenário de modo a fortalecer essas iniciativas regulatórias?

### **Participação Social, Caminhos e Desafios**

A partir do campo da sociocultural, por exemplo, se pode recomendar o impulsionamento de iniciativas de IA nas cadeias culturais alternativas, visando um crescimento significativo e diversificado. Outra possibilidade poderia ser a criação de um “selo de qualidade em IA Responsável”, incentivando organizações a adotarem práticas responsáveis e éticas no desenvolvimento e aplicação de sistemas de IA, com reflexos na conscientização e participação social. Esse selo não apenas reconheceria as boas práticas de responsabilidade, mas também promoveria a confiança do público e estimularia uma concorrência saudável. Ainda nesse sentido, se vislumbra programas de IAR em instituições de ensino; espaços de colaboração virtual em código aberto; ensino de IA por *lowcode*; salientar a importância de acordos internacionais em IA; fomentar a remuneração dadas às pessoas, individualmente, requerendo uma transposição do termo “usuário”

para “produtor” dos dados; se pode pensar numa redistribuição social do lucro de dados.

Entretanto, a realidade concreta e complexa que atravessa o cotidiano é que os meios e usos dos sistemas de IA, como as plataformas de mídia social e outros aplicativos, que são facilmente ofertados pelas *big techs* do Norte Global – por mais pontos de resistência que possam permear essas plataformas –, até agora, têm garantido o acúmulo de dinheiro e poder e subtração de recursos naturais e mão de obra das periferias globais, sem perspectivas concretas de diminuição da desigualdade, violência e vulnerabilidades sociais que se desdobram exponencialmente nos países do Sul Global, como o Brasil (Penteado et al., 2023).

Nesse quadro, é urgente a intervenção regulatória governamental e, como princípio, a participação social para reivindicar e estabelecer criticamente parâmetros de responsabilidades nesse contexto que, de modo democrático, efetivamente beneficiem, ampliem e fortaleçam uma gestão de cuidados centrada no humano, na valorização da comunicação e cultura, tanto no âmbito global quanto dos territórios locais. Por exemplo, os resultados e desdobramentos do avanço tecnológico fomentados em equipamentos culturais – pontos de cultura, bibliotecas, centros culturais, museus, plataformas digitais atuantes no Brasil –, como meios de produção cultural e de processos comunicativos que permitem o trabalho administrativo e também correspondem à interface de acesso entre o ambiente e a sociedade, são recomendáveis e estão disponíveis em diversos levantamentos referentes à cultura e tecnologia no Brasil, do Núcleo de Informação e Coordenação do Ponto BR – NIC.br. Essa entidade vem oferecendo mapeamentos das práticas referentes a essas interfaces com os sistemas algorítmicos de IA no território nacional e inscrevendo medidas necessárias para a promoção da diversidade e inclusão social, considerando dimensões analíticas que implicam questões de acesso, econômicas, da diversidade, criativa, regulatórias, entre outras.

Em um desses levantamentos, Ghezzi (2022) aponta que por mais sedutor que o discurso de liberdade e acesso ilimitado à informação e à comunicação, amplamente difundido na sociedade, esteja atrelado a avanços tecnológicos e a novas possibilidades de acesso e produção cultural, essas:

. . . novas possibilidades dependem de novos acordos de governança, que precisam ser vislumbrados antes de serem pactuados. Algumas questões, por mais que pareçam impossíveis na configuração da Internet atual . . . podem ser norteadoras de uma nova governança na Internet . . . Estas estão relacionadas à transparência e ao papel de indivíduos na posse e uso de seus dados. (Ghezzi, 2022, p. 200)

Santaella (2021) agrega também a essa discussão propondo uma série de medidas, como a inserção da IA no desenvolvimento socioeconômico da região da América Latina e Caribe, a inclusão da cultura nas estratégias de desenvolvimento da IA, a promoção da diversidade cultural nas grandes plataformas, a criação de ecossistemas de dados locais, mapeamento de atores e articulação em rede, e a promoção de pesquisa e capacitação.

Em adição, para além das necessárias medidas regulatórias, se acredita que o poder público também tem que fomentar iniciativas e intervenções culturais e comunicativas de ecossistemas de soluções locais com incentivos governamentais, prestação de serviços ligados aos territórios, investimento e acesso a equipamentos e capacitação técnica para a sociedade manejar e operacionalizar a sua participação cidadã na construção de uma IAR. Esses ecossistemas de soluções territoriais locais, ou pontos de cultura, devem ser administrados, por exemplo, por ativistas e organizações da sociedade civil (Ministério da Cultura, 2024), que se posicionam criticamente diante das gigantes empresas globais de tecnologia, sob o prisma da soberania brasileira sobre os recursos naturais e ambientes de trabalhos criados por *software* livre como, por exemplo, faz a cooperativa plataforma EITA!<sup>21</sup>.

<sup>21</sup> Disponível em: <https://eita.coop.br/>.

Essa cooperativa agrega, dentro de um ambiente virtual, diversas outras cooperativas e comunidades ligadas à reforma agrária, movimento sindical, luta feminista, juventude,

agroecologia, economia solidária, movimento negro, movimento LGBTQIA+, direito à saúde, direito à educação, direito à comunicação etc. Construir esse caminho mais diverso, apoiado por organizações, semelhantes à plataforma EITA!, que respondam e reivindicam a essa lógica de abordagem responsável das neotecnologias da IA, bem como valorizem a participação social como princípio, precisam ser mapeadas e estruturadas de modo urgente e sustentável no Brasil. Nessa direção, é importante também entender, fomentar e difundir conhecimentos sobre como fazer um ecossistema de cultura livre que possa se sustentar aproximando e agregando os movimentos sociais, a sociedade em geral e o desenvolvimento do vínculo participativo, o comum, como ensina Sodr  (2014), a partir dos territ rios locais.

  guisa das considera es finais deste texto, aprender a lidar com esses sistemas algor tmicos de IA e a fomentar a participa o social, como princ pio, para interpell los criticamente   parte da produ o e difus o cultural local. Enquanto o Brasil enfrenta esses desafios,   essencial adotar e estimular uma abordagem proativa e colaborativa para garantir que a IA contribua positivamente para a sociedade e democracia, sendo mobilizada pelas pr ticas comunicativas e culturais que sensibilizam e direcionam, como aponta Sodr  (2014, 2021), para reorganiza o do mundo e das consci ncias. Essa interven o   central nos horizontes desses tempos de incertezas nos quais, com e para al m das neotecnologias de intelig ncia, a produ o da ignor ncia vem sendo fortemente estimulada como for a econ mica e de desestabiliza o social antidemocr tica.

### Refer ncias

Academia Brasileira de Ci ncias. (2023). *Recomenda es para o avan o da intelig ncia artificial no Brasil: GT-IA da Academia Brasileira de Ci ncias*. Academia Brasileira de Ci ncias.

Almeida, V. (2023). *IA Generativa: oportunidades, riscos e governan a*. UFMG. <https://bit.ly/3OdPfJk>

Arun, C. (2020). AI and the Global South: designing for other worlds. In Dubber, M. D., Pasquale, F., & Das, S. (Eds.), *The Oxford handbook of ethics of AI* (pp. 588-606). Oxford Academic.

Baeza-Yates, R. (2023). Lecture held at the Academia Europaea Building Bridges Conference 2022: an introduction to responsible AI. *European Review*, 31(4), 406-421. <https://doi.org/10.1017/S1062798723000145>

Barrett, P. M., Hendrix, J., & Richard-Carvajal, C. (2024). *Digital risks to the 2024 elections: safeguarding democracy in an era of disinformation*. Stern – Center for Business and Human Rights. <https://bit.ly/3wRCIW5>

Bradford, A. (2020). *The Brussels effect: how the European Union rules the world*. Oxford University Press. <https://doi.org/10.1093/oso/9780190088583.001.0001>

Brown, S. (2023, 21 de setembro.). Third-party AI tools pose increasing risks for organizations. *MIT Management Sloan*. <https://bit.ly/42Awu8O>

Bruns, A. (2019). *Are filter bubbles real?* Polity Press Cambridge.

Carville, O. (2023, 20 de abril). TikTok's algorithm keeps pushing suicide to vulnerable kids. *Bloomberg*. <https://www.bloomberg.com/news/features/2023-04-20/tiktok-effects-on-mental-health-in-focus-after-teen-suicide>

Chhabria, P. (2024, 29 de jan.). The big election year: how to protect democracy in the era of AI. *World Economic Forum*. <https://www.weforum.org/agenda/2024/01/ai-democracy-election-year-2024-disinformation-misinformation/>

Comiss o Europeia. (2018, 7 de dezembro). Coordinated plan on artificial intelligence. *European Commission*. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0795#footnote60>

Cordeiro, V. D. (2021). Novas questões para sociologia contemporânea: os impactos da Inteligência Artificial e dos algoritmos nas relações sociais. In F. G. Cozman, G. A. Plonski, H. Neri (Orgs.), *Inteligência artificial: avanços e tendências* (pp. 204-224). Instituto de Estudos Avançados.

Delacroix, S. (2019, 1 de fevereiro). Beware of 'Algorithmic Regulation'. *SSRN Electronic Journal*. <https://ssrn.com/abstract=3327191>

Dignum, V. (2017). Responsible Autonomy. In C. Sierra, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI – 17)* (pp. 4698-4704). AAAI Press.

Dignum, V. (2019). *Responsible Artificial Intelligence: how to develop and use AI in a responsible way*. Springer.

Dignum, V. (2020). Responsibility and Artificial Intelligence. In M. D. Dubber, F. Pasquale, S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 214-231). Oxford Academic. <https://doi.org/10.1093/oxfordhb/9780190067397.013.12>

Domingos, P. (2015). *The master algorithm: how the quest for the ultimate learning machine will remake our world*. Basic Books.

European Commission. (2019, 8 de abr.). Ethics guidelines for trustworthy AI. *European Union*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

European Commission. (2024a, 13 de junho). Regulation (EU) 2024/1689 of the European Parliament and of the Council. *European Union*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>

European Commission. (2024b, 14 de outubro). High-level expert group on artificial intelligence. *European Union*. <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>

Facebook. (2020, 11 de agosto). Our commitment to safety. *Facebook Business*. <https://www.facebook.com/business/news/our-commitment-to-safety>

Foucault, M. (1993). About the beginning of the hermeneutics of the self: two lectures at Dartmouth. *Political Theory*, 21(2), 198-227.

Foucault, M. (2020). *Discipline and punish: the birth of the prison*. Penguin Modern Classics.

Franklin, M., Ashton, H., Gorman, R., & Armstrong, S. (2022). Missing mechanisms of manipulation in the EU AI Act. *The International FLAIRS Conference Proceedings*, 35, 1-4. <https://doi.org/10.32473/flairs.v35i.130723>

Garimella, K., & Chauchard, S. (2024). Kiran Garimella & Simon Chauchard, 2024. How prevalent is AI misinformation? What our studies in India show so far. *Nature*, 630(8015), 32-34. <https://doi.org/10.1038/d41586-024-01588-2>

Ghezzi, D. R. Inteligência Artificial e cultura: uma mirada de longo prazo In Núcleo de Informação e Coordenação do Ponto BR (Ed.), *Inteligência Artificial e cultura: perspectivas para a diversidade cultural na era digital* (pp. 191-204). Comitê Gestor da Internet no Brasil. 2022.

Goggin, B. (2024, 22 de abril). Did TikTok videos inspire a teen's suicide? His mom says she found graphic evidence. *NBC news*. <https://www.nbcnews.com/tech/social-media/tiktok-suicide-videos-lawsuit-social-media-self-harm-rcna146680>

Gullino, D. (2024, 15 de julho). Justiça traça primeiros limites ao uso de inteligência artificial nas eleições municipais e pune deepfakes. *O Globo*. <https://x.gd/eJpao>

- Hagerty, A., & Rubinov, I. (2019). Global AI Ethics: a review of the social impacts and ethical implications of Artificial Intelligence. *arXiv:1907.07892*. <https://arxiv.org/abs/1907.07892>
- Heinrichs, B. Discrimination in the Age of artificial intelligence. *AI & SOCIETY*, 37, 143-154, 2022. <https://doi.org/10.1007/s00146-021-01192-2>
- Jannach, D., & Adomavicius, G. (2016). Recommendations with a Purpose. In *RecSys '16: Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 7-10). Association for Computing Machinery.
- Ministério da Cultura. (2024, 26 de janeiro). Conferência Temática da Cultura Digital – Permacultura Digital: “Começo, Meio e Começo” – 3º Dia. [Vídeo]. YouTube. <https://www.youtube.com/watch?v=yI6drHFmT9w>
- Leiser, M. (2016). The problem with ‘dots’: questioning the role of rationality in the online environment. *International Review of Law, Computers & Technology*, 30(3), 191-210.
- Leite, F. (2022). *Comunicação Antirracista: notas crítico-reflexivas e propositivas* [Apresentação de trabalho]. 31º Encontro Anual da Compós, Imperatriz, Minas Gerais, Brasil.
- MacInnis, D. J. (2011). A framework for conceptual contributions in marketing. *Journal of Marketing*, 75(4), 136-154. <https://doi.org/10.1509/jmkg.75.4.136>
- Many AI researchers think fakes will become undetectable. (2024, 17 de janeiro). *The Economist*. <https://www.economist.com/science-and-technology/2024/01/17/many-ai-researchers-think-fakes-will-become-undetectable>
- Marres, N. (2017). *Digital Sociology: the reinvention of Social Research*. Polity Press.
- McCabe, S. D., Ferrari, D., Green, J., Lazer, D. M. J., & Esterling, K. M. (2024). Post-January 6th deplatforming reduced the reach of misinformation on Twitter. *Nature*, 630, 132-140.
- Mello, P. C. (2024a, 11 de fevereiro). Políticos do Brasil raramente postam fake news em suas redes, mostra estudo. *Folha de S.Paulo*. <https://x.gd/2C3tA>
- Mello, P. C. (2024b, 22 de junho). Inovações da Índia em campanha eleitoral suja na internet acendem alerta para outros países. *Folha de S.Paulo*. <https://x.gd/qgmji>
- Mertens, S., Herberz, M., Hahnel, U. J. J., & Brosch, T. (2021). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences*, 119(1). <https://doi.org/10.1073/pnas.2107346118>
- Moherdau, L. (2024a, 11 de abril). Fake news: “o xadrez da política legislativa move peças”. *Poder 360*. <https://www.poder360.com.br/opiniaofake-news-o-xadrez-da-politica-legislativa-move-pecas/>
- Moherdau, L. (2024b, 29 de fevereiro). Norma das eleições do TSE é falível diante do mar de fakes. *Poder 360*. <https://www.poder360.com.br/opiniaonorma-das-eleicoes-do-tse-e-falivel-diante-do-mar-de-fakes/>
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin.
- Pasquale, F. (2015). *The black box society: the secret algorithms that control money and information*. Harvard University Press.
- Pataranutaporn, P., Liu, R., Finn, E., & Maes, P. (2023). Influencing human – AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence*, 5, 1076-1086.

- Penteado, C., Pellegrini, J., & Silveira, S. A. (Orgs). (2023). *Plataformização, inteligência artificial e soberania de dados: tecnologia no Brasil 2020-2030*. Ação Educativa.
- Pontes, F. (2024, 3 de março). Entenda regras do TSE para uso de inteligência artificial nas eleições. *Agência Brasil*. <https://bit.ly/3PnD0uf>
- Projeto de Lei nº 2338, de 2023*. (2023). Dispõe sobre o uso da Inteligência Artificial. Senado Federal. <https://legis.senado.leg.br/sdleg-getter/documento?dm=9347593&ts=1720798347613&disposition=inline>
- Projeto de Lei nº 2630, de 2020*. (2020). Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Senado Federal. <https://bit.ly/3v9vLZH>
- Resolução nº 23.732, de 27 de fevereiro de 2024*. (2024, 27 de fevereiro). Altera a Res.-TSE nº 23.610, de 18 de dezembro de 2019, que dispõe sobre propaganda eleitoral, utilização e geração do horário gratuito e condutas ilícitas em campanha eleitoral. Tribunal Superior Eleitoral. <https://x.gd/wurcJ>
- Rodotä, S. (2007). Democracy, innovation, and the information society. In P. Goujon, S. Lavelle, P. Duquenoy, K. Kimppa & V. Laurent (Eds.), *The information society: innovation, legitimacy, ethics and democracy* (Vol. 233, pp. 17-26). Springer.
- Rose, N. (1999). *Powers of freedom: reframing political thought*. Cambridge University Press.
- Rose, N., & Miller, P. (2008). *Governing the present: administering economic, social and personal life*. Polity Press.
- Russell, S., Perset, K., & Grobelnik, M. (2023, 29 de novembro). Updates to the OECD's definition of an AI system explained. *OECD.AI Policy Observatory*. <https://bit.ly/3Tgaa0c>
- Santaella, L. (2021). *Inteligência Artificial e cultura: oportunidades e desafios para o Sul Global*. Unesco. <https://bit.ly/3TlucGD>
- Santaella, L. (2022). Inteligência Artificial e cultura: oportunidades e desafios para o Sul Global. In Núcleo de Informação e Coordenação do Ponto BR (Ed.), *Inteligência Artificial e cultura: perspectivas para a diversidade cultural na era digital* (pp. 69-95). Comitê Gestor da Internet no Brasil.
- Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & Soc*, 35, 957-967. <https://doi.org/10.1007/s00146-020-00950-y>
- Soares, M. (2024, 24 de abril). Nova proposta para regulação de IA prevê fiscalização híbrida. *Desinformante*. <https://desinformante.com.br/regulacao-ia-fiscalizacao/>
- Sodré, M. (2001). Objeto da comunicação é a vinculação social [Entrevista concedida a Desirée Rabelo]. *PCLA*, 3(1), 1-6. <https://bit.ly/49itG2t>
- Sodré, M. (2002). *Antropológica do espelho: uma teoria da comunicação linear e em rede*. Vozes.
- Sodré, M. (2014). *A ciência do comum – Notas para o método comunicacional*. Vozes.
- Sodré, M. (2015). A comunicação eletrônica é epistemóloga [Entrevista concedida a Rafael Grohmann]. *Parágrafo*, 1(3), 120-128.
- Sodré, M. (2019). Do lugar de fala ao corpo como lugar de diálogo: raça e etnicidades numa perspectiva comunicacional [Entrevista concedida a Roberto Abib]. *Reciis – Revista Eletrônica de Comunicação Informação & Inovação em Saúde*, 13(4). <https://doi.org/10.29397/reciis.v13i4.1944>
- Sodré, M. (2021). *A sociedade incivil: mídia, iliberalismo e finanças*. Vozes.

Sodré, M. (2023). A ruptura paradigmática da comunicação. *MATRIZES*, 17(3), 19-27.

Stokel-Walker, C. (2024, 18 de janeiro). How AI will manipulate elections in 2024. *I News*. <https://inews.co.uk/news/ai-clear-danger-2024-elections-2859465>

Trump, D. J. [@realDonaldTrump]. (2020, 19 de dezembro). Peter Nvarro releases 36-page report alleging election fraud 'more than sufficient' to swing victory to Trump . . . . X. <https://x.gd/iTzsM>

Twitter, Inc. v. Taamneh. (2023). *Harvard Law Review* 402, 137(1), 400-409. <https://harvardlawreview.org/print/vol-137/twitter-inc-v-taamneh/>

Uuk, R. (2022). Manipulation and the AI act. *Future of Life Institute*. [https://futureoflife.org/wp-content/uploads/2022/08/FLI-Manipulation\\_AI\\_Act.pdf](https://futureoflife.org/wp-content/uploads/2022/08/FLI-Manipulation_AI_Act.pdf)

Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121-136. <http://www.jstor.org/stable/20024652>