

ESFERA PÚBLICA VIRTUAL E IA GENERATIVA:

riscos de “morte” da internet e erosão democrática?

VIRTUAL PUBLIC SPHERE AND GENERATIVE AI:

risks of “death” of the internet and democratic erosion?

Thales Alessandro Dias Pereira (T.A.D.P.)¹

SUMÁRIO. Introdução. 1. Da esfera pública presencial para a virtual. 2. Inteligência artificial generativa e o risco de “morte” da esfera pública virtual. Conclusão. Referências.

RESUMO

O artigo discute os riscos de erosão da esfera pública virtual devido aos avanços da inteligência artificial generativa, destacando a manipulação de comportamentos e a disseminação de conteúdos prejudiciais à democracia. O *cyberspaço* transformou-se em um importante meio de comunicação e mobilização social com a internet e as redes sociais, o que ocorreu em eventos como a Primavera Árabe e o *Occupy Wall Street*. No entanto, a IA generativa, aliada ao uso de *bots* e à aproximação das *big techs* ao espectro político de extrema-direita, traz riscos a esse potencial emancipatório, como o aumento da disseminação de desinformação, perfis falsos e manipulação de comportamentos.

Palavras-chave: esfera pública virtual; inteligência artificial generativa; teoria da internet morta.

ABSTRACT

The article discusses the risks of erosion of the virtual public sphere due to advances in generative artificial intelligence, highlighting the manipulation of behaviors and the dissemination of content harmful to democracy. Cyberspace has become an important means of communication and social mobilization with the internet and social networks, which has occurred in events such as the Arab Spring and Occupy Wall Street. However, generative AI,

¹ Centro Universitário UNDB, São Luís, Brasil. Doutorando em Teorias Jurídicas Contemporâneas pela Universidade Federal do Rio de Janeiro. Mestre em Direito, Estado e Constituição pela Universidade de Brasília. Professor licenciado do Curso de Direito do Centro Universitário UNDB. Defensor Público do Estado do Maranhão. E-mail: thalespereirajur@gmail.com.

combined with the use of bots and the approximation of big techs to the far-right political spectrum, brings risks to this emancipatory potential, such as the increase in the spread of disinformation, fake profiles, and behavior manipulation.

Keywords: virtual public sphere; generative artificial intelligence; dead internet theory.

INTRODUÇÃO

O escopo do presente artigo é discutir os riscos de erosão da esfera pública virtual diante dos avanços da inteligência artificial, especificamente da possibilidade de emprego da nascente inteligência artificial generativa para manipular comportamentos das pessoas, e disseminar conteúdos prejudiciais à democracia.

A esfera pública, como conjunto de fluxos comunicativos no sistema democrático, experimentou transformações significativas com a ascensão da internet e das redes sociais, que ultrapassaram a anterior superação das barreiras físicas já propiciada pela imprensa e pela radiodifusão. Hoje, o *cyberspaço* possibilita a troca de informações, opiniões e deliberações em larga escala e em tempo real.

As plataformas virtuais não apenas encurtam distâncias, mas também aumentam a velocidade e o alcance das trocas comunicativas, facilitando a mobilização social e a organização de movimentos coletivos. Exemplos como a Primavera Árabe e o *Occupy Wall Street* ilustram o papel vital das redes sociais na coordenação de protestos e na disseminação de informações em tempo real.

Todavia, a despeito de seu potencial emancipatório, as tecnologias do espaço virtual são encrustadas no contexto capitalista e colonial, e são sujeitas aos poderes hegemônicos. Redes sociais podem ser utilizadas para influenciar comportamentos e resultados eleitorais. Algoritmos e *bots* podem facilitar a propagação de *fake news* e a criar bolhas de informação que limitam o contato dos cidadãos com pontos de vista divergentes, ameaçando a diversidade e a complexidade da esfera pública virtual.

Nesse quadro, a inteligência artificial generativa, aliada ao uso de *bots* e à aproximação das *big techs* ao espectro político de extrema-direita, torna verossímil uma antiga teoria conspiratória surgida em redes sociais – a “teoria da internet morta”, segundo a qual a quase totalidade do conteúdo em circulação no mundo virtual seria produzido artificialmente.

Obviamente, a “morte” da esfera pública virtual é possibilidade distópica, mas a sua erosão é um problema do presente.

A discussão foi estruturada em duas seções. Primeiramente, é realizada uma breve abordagem da esfera pública virtual e seu potencial de mobilização social. Em sequência, são apresentadas explanações sobre a inteligência artificial generativa e os impactos dessa nova tecnologia sobre a democracia.

2. DA ESFERA PÚBLICA PRESENCIAL PARA A VIRTUAL

Em sua visão deliberativa da democracia, Habermas coloca como centro de seu funcionamento uma esfera pública como rede que permeia toda a sociedade, na qual circulam conteúdos, tomadas de posição e opiniões. Em outros termos, como uma “*estrutura comunicacional* do agir orientado pelo entendimento, a qual tem a ver com o *espaço social* gerado no agir comunicativo”², desligada de uma presença física, graças ao anonimato das pessoas que atuam como interlocutores nesse espaço difuso de comunicações não-especializadas. Esses fluxos comunicativos seriam porosos e sobrepostos, sendo episódicas (em bares, cafés, encontros na rua), de presença organizada (encontros de pais, público que frequenta espetáculos teatrais) ou abstratos (leitores, ouvintes e espectadores espalhados)³, em ligação contínua e sem regras que fixam seus limites de forma definida. O bom funcionamento da democracia dependeria da alimentação dos centros de poder político pelos impulsos advindos da periferia, captados pela esfera pública e satisfeitos pela produção de regulações jurídicas precedidas de deliberações que levem em conta tais impulsos.

Ainda que não se concorde com as bases teórico-discursivas de Habermas⁴, em outras diversas teorizações sobre a democracia a liberdade na circulação de informações entre

² HABERMAS, Jürgen. **Direito e Democracia**: entre facticidade e validade. Tradução de Flávio Beno Siebeneichler. Rio de Janeiro: Tempo Brasileiro, 2003, v. 2. p. 92, grifos no original.

³ HABERMAS, Jürgen. **Direito e Democracia**: entre facticidade e validade. v. 2..., p. 107.

⁴ Luhmann, exemplificativamente, entende que a “opinião pública” é critério orientador de operações no sistema político, possibilitador de observações de segunda ordem: “[a] política orienta suas operações pela opinião pública, para observar a repercussão de suas ações através dos olhos de outros observadores”. LUHMANN. **O direito da sociedade**. Tradução Saulo Krieger; tradução das citações em latim de Alexandre Agnolon. São Paulo: Martins Fontes - selo Martins, 2016. Livro eletrônico. 2,0 Mb; ePUB. p. 318.

os cidadãos é vista como um de seus principais pilares. Para que um sistema político seja considerado democrático, devem existir cadeias ou fluxos comunicativos de contestação e anuência públicas, que conectam os cidadãos a diferentes órgãos do sistema político⁵. A possibilidade de avaliação crítica dos transitórios detentores do poder estatal é um dos requisitos de equidade na alternância periódica dos governantes.

Com o advento da internet, tais redes comunicativas passaram a ultrapassar em muito os locais físicos de discussões de assuntos públicos. Ainda que tecnologias como a imprensa e a radiodifusão tenham encurtado distâncias comunicativas em momentos anteriores, o *cyberespaço*, o mundo virtual possibilitado pela rede mundial de computadores, tornou possível a criação de interações sociais à longa distância *em larga escala e em tempo real*. As redes sociais⁶, ao mesmo tempo em que trazem a difusão de informações além das barreiras espaciais, tal como a imprensa, são capazes de propiciar a *instantaneidade de trocas comunicativas*, de modo semelhante a uma conversa informal presencial. Por isso, apresentam grande potencial de mobilização social:

A internet transformou a ação política e os mecanismos há muito estabelecidos de organização, comunicação e mobilização de movimentos sociais de várias maneiras. A redução dos custos de comunicação permitiu uma distribuição mais rápida e fácil

⁵ Thornhill propõe uma definição *mínima* de democracia em dois níveis: no institucional, como um uma condição social na qual indivíduos ou grupos políticos designados participam de forma igualitária em um sistema representativo, no bojo do qual contribuem para a construção de um poder geral de governança e para a promulgação de leis que regulam as suas próprias ações; no normativo, instituições democráticas são assim definidas e legitimadas pelo fato de que condizem processos de elaboração coletiva de leis, de modo que os cidadãos a elas anuem por serem expressão de um comprometimento coletivo. THORNHILL, Chris. **The Sociology of Law and the Global Transformation of Democracy**. Cambridge University Press, 2018, p. 3-4.

⁶ Papacharissi, no entanto, aloca a esfera virtual em paralelo (ou até mesmo contraposta) à esfera pública geral: “Uma esfera virtual existe na tradição da esfera pública, embora seja radicalmente diferente dela. Esta esfera virtual é dominada por proprietários burgueses de computadores, muito parecido com aquele traçado por Habermas consistindo em proprietários burgueses. Nessa esfera virtual, diversos públicos de interesse especial coexistem e exibem suas identidades coletivas de dissenso, refletindo assim a dinâmica social do mundo real, como Fraser (1992) observou. Esta visão da verdadeira esfera virtual consiste em várias esferas de contrapúblicos que foram excluídos do discurso político dominante, mas que empregam a comunicação virtual para reestruturar a corrente dominante que os expulsou”. PAPACHARISSI, Zizi. The virtual sphere: the internet as a public sphere. p. 21. **New Media Society**, 2002, v. 4, n. 9. Sage Publications. DOI: 10.1177/14614440222226244. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/14614440222226244>. Acesso em: 30 jan. 2025. Tradução livre. No original: “A virtual sphere does exist in the tradition of, but radically different from, the public sphere. This virtual sphere is dominated by bourgeois computer holders, much like the one traced by Habermas consisting of bourgeois property holders. In this virtual sphere, several special interest publics co-exist and flaunt their collective identities of dissent, thus reflecting the social dynamics of the real world, as Fraser (1992) noted. This vision of the true virtual sphere consists of several spheres of counterpublics that have been excluded from mainstream political discourse, yet employ virtual communication to restructure the mainstream that ousted them”.

de informações dos movimentos e permitiu que os indivíduos mantivessem contato com mais pessoas, comunidades e causas diversas. Facilitou o desenvolvimento de ‘laços fracos’, que permitem aos ativistas estender e gerenciar melhor suas redes sociais e afiliar-se a grupos distantes (Bennett, Breunig, & Givens, 2008; Valenzuela, Arriagada, & Scherman, 2012). Isso resultou em um recrutamento mais eficaz para causas coletivas e na promoção de identidades coletivas em comunidades com queixas semelhantes, o que muitas vezes resultou em atividades de protesto transnacionais bem-sucedidas (Bennett, 2003; Della Porta & Mosca, 2005).⁷

Castells enfatiza o aspecto emancipatório das plataformas virtuais, nada obstante a sua propriedade por parte de grandes conglomerados empresariais, em razão da espontaneidade de trocas comunicativas⁸ entre seus usuários:

[...] mesmo que o universo da internet seja construído em torno do poder dos grandes conglomerados empresariais e seja, de alguma forma, regulado pelos governos, ele

⁷ THEOCHARIS, Yannis; LOWE, Will; DETH, Jan W. van; GARCÍA-ALBACETE, Gema. (2015) Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. **Information, Communication & Society**, 2015, v. 18, n. 02, p. 202-220, DOI: 10.1080/1369118X.2014.948035. Disponível em: <http://www.tandfonline.com/loi/rics20>. Acesso em: 10 jan. 2025. p. 204. Tradução livre. No original: “*The internet transformed political action and long established mechanisms of social movement organization, communication, and mobilization in various ways. The reduction of communication costs has allowed for the faster and easier distribution of movement information, and has enabled individuals to stay in touch with more people, communities, and diverse causes. It has facilitated the development of ‘weak ties’, which allow activists to extend and better manage their social networks and affiliate with distant groups (Bennett, Breunig, & Givens, 2008; Valenzuela, Arriagada, & Scherman, 2012). This resulted in more effective recruitment for collective causes and the promotion of collective identities across communities with similar grievances, which often led to successful transnational protest activity (Bennett, 2003; Della Porta & Mosca, 2005)*”.

⁸ Para Castells, “A comunicação livre é a prática mais subversiva de todas, pois desafia o poder dos relacionamentos incorporados às instituições e organizações da sociedade. A comunicação de cima para baixo, sob o controle das empresas e dos governos, caracterizou a história da humanidade. Qualquer nova tecnologia de comunicação, tal como a prensa de impressão, é um desafio à autoridade, pois as sementes da revolta existentes na maior parte dos indivíduos que estão incorporados às formas permanentes e injustas de organização social apenas podem crescer e florescer quando são conectadas a outros indivíduos, quebrando as barreiras da experiência individual para então se tornarem mobilização social e projetos alternativos de organização social — não necessariamente para o bem, em termos normativos, porque a história não tem um sentido predeterminado de (p. 30) experiência positiva, contra a suposição ideológica de progresso como sendo o destino inelutável da evolução humana. Instituições estabelecidas, em todos os domínios da vida, são desafiadas por aqueles que se sentem dominados, desvalorizados, explorados, humilhados e mal representados. Esses desafios precisam confrontar a capacidade coercitiva das instituições, bem como a habilidade persuasiva da mentalidade dominante que legitima as formas existentes de relações de poder. O reino da comunicação é a esfera social onde valores e interesses de atores conflitantes estão comprometidos em disputa e debate para reproduzir a ordem social, para subvertê-la, ou para acomodar novas formas resultantes de interação entre o velho e o novo, o passado de dominação cristalizado e o futuro de projetos alternativos para a existência humana promovidos por aqueles que aspiram a mudar o mundo e estão prontos para lutar por isso. Diferentemente da esfera institucional deliberativa, que é sistematicamente tendenciosa em relação à dominação existente, a esfera da comunicação é moldada pelos múltiplos insumos que recebe de uma diversidade de fontes, assim como por sua interação”. CASTELLS, Manuel. **O poder da comunicação**. Tradução de Vera Lúcia Mello Joscelyne. 4 ed. São Paulo: Paz e Terra, 2019. p. 31.

permanece sendo, de fato, um modo de comunicação muito distinto, caracterizado pela considerável autonomia dos sujeitos comunicantes em relação aos donos e reguladores da infraestrutura de comunicação. Isso acontece porque as tecnologias de redes digitais permitem que indivíduos e organizações gerem seus próprios conteúdos e mensagens e os distribuam no ciberespaço, evitando amplamente o controle de corporações e burocracias.⁹

Dois dos exemplos internacionais mais conhecidos de mobilização social por redes sociais foram a Primavera Árabe e o *Occupy Wall Street* nos anos de 2010 e 2011. Na primeira, o uso das redes sociais Facebook e Twitter por ativistas e cidadãos comuns para documentar abusos governamentais, organizar manifestações e campanhas de solidariedade, e disseminar informações em tempo real, tanto localmente quanto internacionalmente, permitiu uma mobilização social rápida e eficaz, derrubando regimes autoritários na Tunísia, Egito e Líbia¹⁰. Já no *Occupy*, a plataforma Twitter foi utilizada como meio de, dentre outros: facilitar protestos presenciais; possibilitar o repasse e disseminação ao vivo de informações sobre esses protestos; permitir a expressão de opiniões e discussões das pessoas sobre o movimento; estabelecer conexões entre os ativistas; e facilitar ações virtuais¹¹. No Brasil, um dos exemplos de mobilização social a partir do Twitter foi a série de protestos de 2013, impulsionados em grande parte por meio dessa rede social¹².

⁹ CASTELLS, Manuel. **O poder da comunicação...** p. 30.

¹⁰ HOWARD, Philip N.; HUSSAIN, Muzammil M. **Democracy's Fourth Wave?: Digital Media and the Arab Spring**. Oxford University Press: New York, 2013. Os autores apresentam um modelo de seis estágios de mobilização social por redes sociais: “[...] uma fase de capacitação que envolve a difusão e o entrenchamento da mídia digital ao longo de muitos anos em comunidades locais e diaspóricas; uma fase de preparação que envolve ativistas aprendendo a usar a mídia digital de maneiras criativas para se encontrarem, construir solidariedade em torno de queixas compartilhadas e identificar objetivos políticos coletivos; uma fase de ignição que envolve algum incidente incitante, geralmente ignorado pela grande mídia controlada pelo Estado, que enfurece o público e é alavancada por grupos da sociedade civil; uma fase de protestos de rua que são coordenados digitalmente; uma fase de adesão internacional, durante a qual a mídia digital é usada para atrair governos internacionais, diásporas globais e agências de notícias estrangeiras; uma fase de clímax em que ou o Estado reprime, ou os manifestantes e elites chegam a um impasse, ou as demandas públicas são atendidas; e então um desfecho final de uma guerra de informação pós-protesto entre vencedores e perdedores na mudança social”. HOWARD, Philip N.; HUSSAIN, Muzammil M. **Democracy's Fourth Wave?: Digital Media and the Arab Spring...**, p. 124.

¹¹ PENNEY, J; DADAS, C. (Re)Tweeting in the service of protest: digital composition and circulation in the Occupy Wall Street movement. **New Media & Society**, p. 1-17, 2013. Disponível em: https://www.researchgate.net/publication/259823399_ReTweeting_in_the_service_of_protest_Digital_composition_and_circulation_in_the_Occupy_Wall_Street_movement. Acesso em: 21 fev. 2025.

¹² RECUERO, Raquel; ZAGO, Gabriela; BASTOS, Marco Toledo. O discurso dos #ProtestosBR: análise de conteúdo do Twitter. **Galáxia**, São Paulo, v. 14, n. 28, p. 199-216, Dec. 2014. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1982-25532014000200017&lng=en&nrm=iso. DOI: <https://doi.org/10.1590/1982-25542014217911>. Acesso em: 21 jan. 2025.

Com o tempo, o potencial de mobilização da esfera pública virtual se confirmou. Mas, de fato, o quão realmente livre são as comunicações no *cyberespaço*? Investigações científicas e jornalísticas desvelaram que o mundo virtual é palco de manipulações comportamentais que, se não inéditas, nunca foram tão sofisticadas e individualizadas. O uso de redes sociais e o acesso a *sites* da internet trazem o risco de coleta involuntária ou semivoluntária de diversos tipos de dados de seus usuários, possibilitando, por suas companhias proprietárias ou terceiros, desde uma gestão mais lucrativa de anúncios até o reconhecimento de preferências políticas para fins de campanhas eleitorais. Relembre-se o escândalo da *Cambridge Analytica*. Essa empresa minerou dados pessoais de milhões de perfis da rede *Facebook* com a finalidade elaborar estratégias digitais para influenciar a eleição estadunidense em favor de Donald Trump em 2016. Inclusive, parte desses dados foi obtida através de um questionário gratuito (o *quiz app* “*thisisyourdigitallife*”) que fazia entender que os usuários estariam fornecendo informações pessoais para pesquisa acadêmica. A revelação foi feita pelos jornais *The New York Times*, *The Observer* e *The Guardian* em março de 2018¹³.

Além disso, redes sociais são permeadas por algoritmos que determinam quais tipos de postagens devem ter mais visibilidade aos usuários. Mensagens automatizadas por *bots* podem facilitar a propagação de *fake news* e a perpetuação de bolhas de desinformação, minando a autonomia de vontade das pessoas. O uso de tecnologias de geração de imagem e som propicia a criação de conteúdos falsos, influenciando sobretudo pessoas mais vulneráveis¹⁴. Os usuários passam não ter contato com pontos de vista divergentes, suprimindo-se um dos principais pontos positivos da esfera pública virtual: diversidade e complexidade de posicionamentos e valores que refletem a sociedade moderna.

Pode-se dizer que as *fake news* e a manipulação algorítmica de postagens (criando as “bolhas”) concretizaram, em sede virtual, a chamada comunicação sistemicamente

¹³ LAPAIRE (Bordeaux Montaigne), Jean-Rémy. Why content matters. Zuckerberg, Vox Media and the Cambridge Analytica data leak. **Antares: Letras e Humanidades**, 10(20), 2018, p. 88–110. Disponível em: <https://sou.ucs.br/etc/revistas/index.php/antares/article/view/6583>. Acesso em 20 fev. 2025. p. 89.

¹⁴ BASTOS, Elísio Augusto Velloso. *et al.* Algoritmos, Inteligência Artificial e novas formas de “fazer política”: análise da influência da IA nos processos eleitorais democráticos na contemporaneidade. **Revista Direito e Justiça**, v. 21, n. 41, p. 175-193, 2021. <https://san.uri.br/revistas/index.php/direitojustica/article/view/635/320>. Acesso em 30 jan. 2025.

distorcida¹⁵. Trata-se de uma comunicação racional¹⁶ na aparência, porém, em essência, o poder econômico ou político distorce a comunicação, uma vez que a tecnologia, tão infiltrada na vida cotidiana, bombardeia as subjetividades a ponto de fazer com o que o autoengano dos agentes (ideologia) torne impossível a referência a um mundo objetivo comum.

Nesse contexto, os poderes hegemônicos contam com mais uma arma para domar a espontaneidade comunicativa das redes virtuais: a inteligência artificial generativa, capaz de gerar conteúdos escritos ou mídia audiovisual de forma rápida e autônoma. Isso será abordado no próximo tópico.

3. INTELIGÊNCIA ARTIFICIAL GENERATIVA E O RISCO DE “MORTE” DA ESFERA PÚBLICA VIRTUAL

Embora subsista certa controvérsia para estabelecer-se um conceito de inteligência artificial, é possível considerá-la como o conjunto de sistemas artificiais eletrônicos que se destinam à resolução de problemas gerais ou específicos, simulando características da inteligência humana, como a possibilidade de decidir entre diferentes alternativas e capacidade de autoaperfeiçoamento¹⁷.

¹⁵ BUSCHMAN, John. Fake News as Systematically Distorted Communication. **Journal of Documentation**, vol. 80, n. 1, p. 203-217. Disponível em: http://works.bepress.com/john_buschman/101/. Acesso em 24 fev. 2025.

¹⁶ Comunicação para Habermas é usar a linguagem para fins de entendimento mútuo, por meio de razões ou argumentos, com referência a um mundo objetivo (fatos), subjetivo (intenções) e social (normas). O agir comunicativo se contrapõe ao agir estratégico, pelo qual o sujeito não busca convencer racionalmente o seu interlocutor, mas sim influenciar o seu comportamento por meio de cálculos egocêntricos (prêmios, ameaças ou manipulações). O mundo da vida é a esfera social da interação cotidiana dos agentes, no qual a cultura se reproduz pelo agir comunicativo. Sistemas, por outro lado, são emaranhados de ações estratégicas coordenadas pelo dinheiro (sistema econômico) ou pelo poder burocrático (sistema administrativo). HABERMAS, Jürgen. **Teoria do Agir Comunicativo**. Tradução de Paulo Astor Soethe. São Paulo: WMF Martins Fontes, 2012. v. 1.

¹⁷ O Projeto de Lei nº 2338, de 2023, do Senado Federal (atualmente em tramitação na Câmara dos Deputados), que pretende estabelecer o marco regulatório da IA no Brasil, traz a seguinte definição: “sistema baseado em máquina que, com graus diferentes de autonomia e para objetivos explícitos ou implícitos, infere, a partir de um conjunto de dados ou informações que recebe, como gerar resultados, em especial previsão, conteúdo, recomendação ou decisão que possa influenciar o ambiente virtual, físico ou real”. BRASIL. Senado Federal. **Projeto de Lei nº 2338, de 2023 do Senado Federal**. Brasília, 2023. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9881643&ts=1738768184212&disposition=inline>. Acesso em: 24 fev. 2025.

Nos últimos anos, um novo marco no desenvolvimento de sistemas de IA foi alcançado: com avanços significativos na área de aprendizado de máquina¹⁸, particularmente no uso de redes neurais profundas¹⁹, foram disponibilizados para uso do público geral modelos de IA generativa. Eles são capazes de identificar e replicar padrões complexos nos dados de treinamento, permitindo a criação autônoma de novos conteúdos.

Um dos conceitos-chave na IA generativa é o dos grandes modelos de linguagem (*large language models* – LLMs), que são algoritmos treinados em vastos conjuntos de dados para processar demandas em linguagem natural. Podem prever a próxima palavra em uma sequência, construir frases completas ou até mesmo redigir parágrafos coerentes sobre tópicos específicos. Esses modelos utilizam arquiteturas sofisticadas, como as redes transformadoras, que permitiram avanços significativos na capacidade de entender e gerar linguagem natural. Os conhecidos e largamente utilizados GPT's (*generative pre-trained transformers* – transformadores generativos pré-treinados) são uma espécie de LLM, desenhados para entender e gerar texto imitativos de textos humanos com base nos *inputs* recebidos²⁰.

A evolução dessas tecnologias pode ser traçada cronologicamente por meio de algumas inovações. Em 2014, a introdução das Redes Geradoras Adversárias (GANs) por Ian Goodfellow e outros pesquisadores representou um amplo avanço da IA generativa²¹. Essas redes consistem em duas redes neurais que competem entre si: uma geradora que cria novos

¹⁸ Aprendizado de máquina é a aplicação de algoritmos que podem ser “treinados”, ou seja, que podem aperfeiçoar o seu comportamento através da extração de padrões a partir de dados não processados, até o nível em que possam realizar satisfatoriamente tarefas novas, ou seja, em relação às quais ainda não tiveram experiência. O aprendizado pode ser supervisionado (em que o ser humano intervém inserindo “rótulos” nos dados de treinamento), não-supervisionado (os dados não são rotulados) e por reforço (em que há estímulos para o bom desempenho da máquina). No caso da abordagem do aprendizado profundo (*deep learning*), a evolução da máquina é realizada pelo uso de redes neurais de múltiplas camadas e de conjuntos amplos de dados (*large datasets*). Redes neurais são um tipo de sistema computacional inspirado nas propriedades básicas dos neurônios, sendo compostas por muitas unidades individuais interligadas, em que cada uma recebe comandos de entrada (*inputs*) e envia comandos de saídas (*outputs*) a outras, tendo cada ligação um determinado “peso” modificado pela experiência. PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. **Inteligência Artificial e Direito**. Curitiba: Alteridade, 2019, p. 88-100.

¹⁹ Veja-se a nota anterior.

²⁰ STACKADEMIC. Understanding the Difference Between GPT and LLM: A Comprehensive Comparison. 2023. Disponível em: <https://stackademic.com/blog/understanding-the-difference-between-gpt-and-llm-a-comprehensive-comparison-1f624c713507>. Acesso em 14 fev. 2025.

²¹ GOODFELLOW, Ian J; *et al.* **Generative Adversarial Nets**. 2014. Disponível em: <https://arxiv.org/abs/1406.2661v1>. Acesso em: 20 fev. 2025.

dados e uma discriminadora que tenta distinguir esses dados dos dados reais²². Posteriormente, em 2018, a OpenAI lançou o GPT (*Generative Pre-trained Transformer*), um modelo de linguagem que utiliza uma arquitetura transformadora²³ para prever a próxima palavra em uma sequência de texto²⁴. O GPT-2, lançado em 2019, e o GPT-3, em 2020, expandiram essa capacidade, permitindo a geração de texto coerente e contextualizado em uma variedade de tópicos²⁵. Em 2021, a OpenAI também apresentou o DALL-E, um modelo que pode gerar imagens a partir de descrições textuais, combinando a compreensão da linguagem natural com a geração de imagens²⁶.

Esses novos patamares foram possibilitados pela crescente disponibilidade de conjuntos de dados massivos e pelo aumento do poder computacional, permitindo que a IA generativa alcance níveis cada vez maiores de sofisticação. Hoje, essa tecnologia é amplamente utilizada em diversos setores, como produção textual, entretenimento, arte e design, em uma abrangência de modelos, que podem ser proprietários ou fechados (códigos-fonte apenas parcialmente abertos ou totalmente fechados para o público de desenvolvedores), como os da OpenAI, Google, Microsoft, IBM, Apple, Deepmind, Adept, Anthropic, Cohere, ou abertos, como GPT4ALL, LLama e OPT (*Open pre-trained Transformer*) da Meta, OpenNMT, Koala,

²² “Na estrutura de redes adversárias proposta, o modelo generativo é colocado contra um adversário: um modelo discriminativo que aprende a determinar se uma amostra é da distribuição do modelo ou da distribuição de dados. O modelo generativo pode ser pensado como análogo a uma equipe de falsificadores, tentando produzir moeda falsa e usá-la sem detecção, enquanto o modelo discriminativo é análogo à polícia, tentando detectar a moeda falsa. A competição neste jogo impulsiona ambas as equipes para melhorar seus métodos até que as falsificações sejam indistinguíveis das amostras genuínas”. GOODFELLOW, Ian J; *et al.* **Generative Adversarial Nets...** p. 1, tradução livre. No original: “*In the proposed adversarial nets framework, the generative model is pitted against an adversary: a discriminative model that learns to determine whether a sample is from the model distribution or the data distribution. The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles*”.

²³ Esse tipo de arquitetura possibilita que o modelo considere o contexto de cada palavra em uma frase ao gerar textos.

²⁴ RADFORD, Alec; *et al.* **Improving Language Understanding by Generative Pre-Training**. 2018. OpenAI. Disponível em: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf. Acesso em 20 fev. 2025.

²⁵ BROWN, Tom. B.; *et al.* **Language Models are Few-Shot Learners**. 2020. ArXiv. Disponível em: <https://arxiv.org/abs/2005.14165v4>. Acesso em 20 fev. 2025.

²⁶ RAMESH, Aditya; *et al.* (2021). **Zero-Shot Text-to-Image Generation**. 2021. Arxiv. Disponível em: <https://arxiv.org/abs/2102.12092v2>. Acesso em 21 fev. 2025.

Open Assistant, Alpaca-LoRA, Vicuna 1.3 Meta OPT, Pythia e Cerebras-GPT²⁷, DeepSeek-R1²⁸, dentre diversos outros LLM's.

Contudo, esses avanços tecnológicos não são isentos de impactos negativos e nem se traduzem necessariamente em avanços civilizacionais. Nenhum desenvolvimento tecnológico é politicamente neutro. Ao tratarem do colonialismo digital²⁹, apontam Deivison Faustino e Walter Lippold que:

Ainda que pareça contraintuitivo, dados, códigos e programas virtuais são entes sujeitos às leis da física e, sobretudo, aos interesses sociais e aos projetos de poder que lhes dão fundamento e existência. Como tudo o que existe, se movimentam no tempo e no espaço ou são armazenados a certa velocidade e intensidade, fisicamente definidas pela infraestrutura existente mediante determinado contexto social. Toda via de transmissão - rodovia, rede de esgoto, fibra óptica etc. - tem um limite posto por suas características físicas e disposição geográfica, e elas próprias são ou não construídas em alguns lugares a partir de decisões políticas e econômicas socialmente determinadas.³⁰

Por isso é importante trazer à lume as narrativas dos *impactados*, que normalmente são soterradas pelas narrativas do progresso. Shakir Mohamed, Marie-Therese Png e William Isaac utilizam o termo *colonialismo algorítmico* para designar como o colonialismo está presente nas seguintes práticas: vieses em sistemas de decisão algorítmica; exploração laboral

²⁷ SINGH, Pankaj. Explore These 10 GPT-4 Open-Source Alternatives. 20 nov. 2024. **Analythics Vidhya**. Disponível em: <https://www.analyticsvidhya.com/blog/2024/04/gpt-open-source-alternatives/>. Acesso 27 fev. 2025.

²⁸ O lançamento do modelo DeepSeek-R1, da chinesa Deepseek, causou enorme impacto no mercado, uma vez que, além de ter o seu código aberto, foi anunciado que o seu desenvolvimento se deu com muito menos recursos econômicos e impactos ambientais do que os modelos da OpenAI. BAPTISTA, Eduardo. What is DeepSeek and why is it disrupting the AI sector?. **Reuters**, 2025. Disponível em: <https://www.reuters.com/technology/artificial-intelligence/what-is-deepseek-why-is-it-disrupting-ai-sector-2025-01-27/>. Acesso em: 27 fev. 2025. Em outras palavras, o DeepSeek pode significar uma ameaça ao monopólio tecnológico estadunidense em IA.

²⁹ “Sua existência se materializa a partir de duas tendências. A primeira é a emergência de uma nova partilha territorial do globo terrestre entre os grandes monopólios da indústria da informação: as chamadas “big techs”, majoritariamente concentradas no vale do Silício, mas não apenas - partilha essa que atualiza o imperialismo, o subimperialismo e o neocolonialismo tardio ao reduzir o chamado Sul global a mero território de mineração extrativista de dados informacionais. [...] A segunda tendência, também nomeada colonialismo de dados, é aquela que subsume cada vez mais a vida humana, o ócio, a criatividade, a cognição e os processos produtivos às lógicas extrativistas, automatizadas e panópticas do colonialismo digital. Não se trata, aqui, de uma simples alteração dos ritmos de vida ou mesmo da percepção humana sobre a introdução de novas tecnologias, como poderia se presumir, mas, sim, da manipulação intencional da cognição humana por grandes corporações empresariais a partir dessas tecnologias, com vistas à ampliação da acumulação de capitais”. FAUSTINO, Deivison; LIPPOLD, Valter. **Colonialismo Digital**: por uma crítica hacker-fanoniana. São Paulo: Boitempo, 2023, p. 24.

³⁰ FAUSTINO, Deivison; LIPPOLD, Valter. **Colonialismo Digital**: por uma crítica hacker-fanoniana... p. 24.

na classificação de dados por trabalhadores [do Sul global] a baixos valores; testagem-beta de novas tecnologias de alto impacto primeiramente no Sul e em populações vulneráveis, tal como era feito com técnicas médicas experimentais; supremacia do Norte global no desenvolvimento de políticas regulatórias de IA, tendentes a favorecer os países desenvolvidos.³¹

Recentemente, foi noticiado mesmo na grande mídia o alinhamento *expresso* das *big techs*³² ao governo conservador, considerado de extrema-direita³³, de Donald Trump. Em sua cerimônia de posse (20 de janeiro de 2025), sentaram-se na mesma fileira Mark Zuckerberg, chefe da Meta, Jeff Bezos (fundador da Amazon), Elon Musk (dono da rede social X) e outros executivos de grandes empresas de tecnologia como Tim Cook, chefe da Apple – controladores de companhias avaliadas em trilhões de dólares. Meses antes da mudança governamental, essas empresas já encampavam algumas bandeiras de Trump, especialmente a não-regulação das redes sociais³⁴.

O que se seguiu foi um pacote regulatório que explicitou as diretrizes do governo estadunidense em relação à IA: assegurar a liderança tecnológica de empresas estadunidenses e reduzir regulações consideradas incompatíveis com o ideário conservador atualmente hegemônico. Em 20 de janeiro de 2025, foi revogada a Ordem Executiva n. 14.110 –

³¹ MOHAMED, Shakir; PNG, Marie-Therese; ISAAC, William. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. **Philosophy & Technology**, n. 33, p. 659–684, 2020. Disponível em: <https://doi.org/10.1007/s13347-020-00405-8>. Acesso em 20 maio 2024.

³² “As chamadas big techs - grandes corporações do ramo da tecnologia digital - representam um dos elos fundamentais do atual estágio de acumulação capitalista. Juntas, as corporações do vale do Silício valem mais de 10 trilhões de dólares. Das dez empresas mais valiosas do mundo, somente duas (Aramco e Hathaway) não atuam direta ou indiretamente no ramo da indústria digital'. Só as chamadas Big Five (Apple, Amazon, Alphabet, Microsoft e Facebook) somaram quase 900 bilhões em receita em 2019”. FAUSTINO, Deivison; LIPPOLD, Valter. **Colonialismo Digital**: por uma crítica hacker-fanoniana... p. 72.

³³ LEVITISKY, Steven; ZIBLATT, Daniel. **How democracies die**. Crown: New York, 2018, p. 57-58. Alguns dos elementos que caracterizariam o primeiro governo de Trump como de extrema-direita seriam: nacionalismo exacerbado, com adoção de medidas xenofóbicas; conservadorismo moral projetado no sistema legal, com redução de direitos de minorias, como mulheres, população LGBTQI+, negros, indígenas e imigrantes; liberalismo entendido como redução da estrutura do Estado de Bem-Estar; descridibilização seletiva do sistema eleitoral; discursos populistas; idealização do passado nacional. LIMA, Melina Moreira Campos. O Direito Internacional sob governos de extrema direita: uma comparação entre as administrações de Donald Trump e Bolsonaro. **Sequência Estudos Jurídicos e Políticos**, Florianópolis, v. 43, n. 90, p. 1–28, 2022. DOI: 10.5007/2177-7055.2022.e66065. Disponível em: <https://periodicos.ufsc.br/index.php/sequencia/article/view/66065>. Acesso em: 21 fev. 2025. p. 6-10.

³⁴ PRAZERES, Leandro. Como aliança entre Trump e big techs aumenta pressão sobre governo Lula e STF. **Portal BBC (portal eletrônico)**. Brasília, 21 de janeiro de 2025. Disponível em: <https://www.bbc.com/portuguese/articles/cx2k80ekl3ko>. Acesso em 20 fev. 2025.

Desenvolvimento e Uso Seguro, Protegido e Confiável da Inteligência Artificial –, que determinava aos desenvolvedores de IA o compartilhamento de informações de suas tecnologias com o Governo Federal antes de sua disponibilização ao público. No dia seguinte, 21 de janeiro, foi anunciado na Casa Branca o *Stargate Project*, um empreendimento privado envolvendo *big techs* que planeja um investimento de 500 bilhões de dólares em infraestrutura de IA. Em 23 de janeiro, a Ordem Executiva 14.179 – Removendo Barreiras à Liderança Americana em Inteligência Artificial – entrou em vigor, reduzindo ainda mais a regulação da IA nos Estados Unidos³⁵.

No que concerne à IA generativa, Fernanda Rodrigues aponta quatro tipos de impactos negativos dessa tecnologia: baixa qualidade dos dados de treinamento, que podem levar a erros factuais e diversos tipos de vieses, como de raça, gênero e classe, além de violações de propriedade intelectual e direitos autorais; exploração laboral de treinadores de IA, sujeitos a baixa remuneração e contato com discursos de ódio; preocupações quanto à redução de vagas de trabalhadores em decorrência da aplicação dessa tecnologia nos ambientes laborais; fortes impactos ambientais dessa tecnologia, fator quase invisibilizado nas abordagens do tema³⁶; a reprodução de estereótipos nocivos e disseminação de desinformação com o auxílio desses modelos de IA, de modo intencional ou não³⁷.

³⁵ BARNES & THORNBURG LLP. A Look at U.S. Government's Changed Approach to Artificial Intelligence Development and Investments. **Barnes & Thornburg LLP (portal eletrônico)**. Disponível em: <https://btlaw.com/insights/alerts/2025/a-look-at-us-governments-changed-approach-to-artificial-intelligence-development-and-investments>. Acesso em 19 fev. 2025.

³⁶ “Além de grandes bases de dados, o desenvolvimento, treinamento e operação dessa tecnologia demanda alta capacidade computacional. Isto, por sua vez, aumenta a demanda por data centers (servidores), já sobrecarregada pelas frequentes ondas de calor, que podem causar disrupções. Segundo a OCDE, atualmente, estes data centers estão entre os que mais consomem água nos EUA, que precisa ser utilizada para refrigeração e geração de energia. / Em West Des Moines, Iowa (EUA), moradores locais entraram com uma ação judicial que, em julho de 2022, revelou que o cluster de data center que atendia ao modelo mais avançado da OpenAI até então, o GPT-4, consumia em torno de 6% da água do distrito logo antes de terminar o treinamento do modelo. Tanto Google quanto Microsoft também registraram grandes picos de utilização de água, de 20% e 34% respectivamente, quando treinavam seus grandes modelos de linguagem (Bard e Bing)⁷⁰. Até 2040, estima-se que os parques de servidores serão responsáveis por até 14% das emissões globais de carbono. A busca crescente por mais espaço pode fazer, ainda, com que haja um deslocamento da indústria para áreas subpovoadas, em busca de menor custo e mais fontes de recursos naturais, por exemplo, o que pode fazer com que países menos desenvolvidos acabem arcando mais com os custos para o crescimento de grandes empresas de tecnologia”. RODRIGUES, Fernanda. Riscos e danos associados à I.A. generativa e a busca por uma tipologia. In: SILVA, Tarcizio (org.). **Inteligência Artificial Generativa: discriminação e impactos sociais**, p. 31-38. Online: Desvelar, 2024. Disponível em desvelar.org. Acesso em: 24 fev. 2025. p. 35.

³⁷ RODRIGUES, Fernanda. Riscos e danos associados à I.A. generativa e a busca por uma tipologia... p. 31-38.

Obviamente, a IA, pelo menos até o momento, não “compreende” o mundo da mesma forma que o ser humano. Como resgata Tarcizio Silva, as pesquisadoras Timnit Gebru e Margaret Mitchell³⁸ definiram os grandes modelos de linguagem (LLM’s) como *papagaios estocásticos*, já que “aleatoriamente costuram sequências de formas linguísticas observadas em seus vastos dados de treinamento, de acordo com informação probabilística sobre como combinam, mas sem nenhuma referência a significado”³⁹. Trata-se não de compreensão, mas de *computação* de elementos com maior probabilidade de acerto conforme o treinamento da máquina e a sofisticação de sua programação. Para o pesquisador, o maior risco da IA generativa é o empobrecimento epistêmico do ser humano:

Parece-nos que a maior parte dos danos gerados pela I.A. generativa – ou, como prefiro, por sistemas algorítmicos derivativos, envolvem a erosão da integridade, confiança e diversidade epistêmica. Ou seja, desvalorização e apagamento de avanços na produção de conhecimentos políticos, científicos e humanísticos alcançados a muito custo globalmente.⁴⁰

Para a esfera pública virtual, isso pode significar a sua erosão em razão do impulsionamento sem precedentes de desinformação e discursos de ódio pela IA generativa. Basta imaginar não apenas respostas simples e automatizadas em redes sociais para replicar ou reforçar *fake news* (isso já ocorre hoje), mas matérias jornalísticas inteiras, vídeos, apresentações, áudios – tudo veiculando desinformação com qualidade suficiente para tornar o artificial indiscernível do real, ao menos para a grande parte das pessoas. Multiplique-se o dano com a automatização por *bots*⁴¹ para se chegar à disseminação desses conteúdos em tempo real, em diversos idiomas e em diferentes tipos de mídias, de modo altamente adaptável ao

³⁸ BENDER, Emily M. et al. On the dangers of stochastic parrots: Can language models be too big?. In: **Proceedings of the 2021 ACM conference on fairness, accountability, and transparency**, 2021, p. 610-623. Disponível em: <https://dl.acm.org/doi/10.1145/3442188.3445922>. Acesso em 20 fev. 2025.

³⁹ SILVA, Tarcizio. Papagaios estocásticos: a revanche da epistemologia da ignorância. In: SILVA, Tarcizio (org.). **Inteligência Artificial Generativa: discriminação e impactos sociais**, p. 57-61. Online: Desvelar, 2024. Disponível em desvelar.org. Acesso em: 24 fev. 2025. p. 57.

⁴⁰ SILVA, Tarcizio. Papagaios estocásticos: a revanche da epistemologia da ignorância... p. 58.

⁴¹ *Bots* são algoritmos programados para realizar tarefas autonomamente. Tais tarefas podem não ter intenção prejudicial a terceiros (embora possam prejudicá-los indiretamente), como a indexação de conteúdos para *engines* de busca (Google, Yahoo, Jusbrasil), ou podem ser ataques virtuais para derrubar sites, furtar dados de terceiros e disseminar informações falsas. Segundo o *Bad Bot Report* de 2024 da Imperva, 49,6% do tráfego da internet em 2024 foi composto por *bots*, dos quais 32% foram com intenção maliciosa. IMPERVA. **2024 Bad Bot Report**. 2024. Disponível em: <https://www.imperva.com/resources/resource-library/reports/2024-bad-bot-report/>. Acesso em 19 fev. 2025.

comportamento das pessoas, ou em um feedback-loop reprodutor de absurdos racistas, sexistas, classistas e outros tantos existentes nos cantos mais sombrios das redes virtuais.

Nesse cenário, cidadãos não teriam poder maior do que o de *espectadores* manipulados, dentro de um ambiente em que *bots* falam com *bots* que respondem para *bots*, não se sabendo de fato quem são as pessoas *reais*. Essa é a nova versão da “*dead internet theory*” ou “teoria da internet morta”. Essa “teoria” começou como posts conspiratórios em fóruns 4Chan, no final dos anos 2010, que sustentavam ter a atividade humana na internet cessado, passando a ser a maior parte das interações on-line artificiais⁴².

Apesar do caráter hiperbólico da “teoria”, com a IA generativa ela passou atualmente a ter verossimilhança. Muito em breve, será praticamente impossível discernir sem exame apurado conteúdos textuais e audiovisuais genuínos de conteúdos gerados por IA. Como aponta Yoshija Walter:

Dez anos atrás, a teoria costumava ser bastante especulativa, mas com o despertar da IA generativa, agora pode ser observada em primeira mão e destaca uma tendência perturbadora: as linhas tênues entre as interações humanas e orientadas pela IA. À medida que a IA prolifera na geração de conteúdo, ela molda cada vez mais as percepções e comportamentos do usuário, direcionando a narrativa digital para seus objetivos determinados por algoritmos, muitas vezes em desacordo com a natureza inerentemente imprevisível do discurso humano. O problema é exacerbado pelo fato de que muitos podem ter dificuldade em discernir entre dados “reais” e “falsos” (ou, nesse caso, “gerados por humanos” e “gerados por IA”). Portanto, o surgimento de *deepfakes*, um subproduto dos avanços da IA, apresenta desafios para a confiança digital (Walter 2022). Essas ferramentas orientadas por IA, capazes de criar conteúdo altamente realista, mas fabricado, representam uma ameaça significativa à integridade das informações online, impulsionando a desinformação e corroendo a base de confiança essencial para interações digitais saudáveis.⁴³

⁴² DI PLACIDO, Dani. The Dead Internet Theory, Explained. **Forbes (portal eletrônico)**. 16 de janeiro de 2024. Disponível em: <https://www.forbes.com/sites/danidiplacido/2024/01/16/the-dead-internet-theory-explained/>. Acesso em 05 mar. 2025.

⁴³ WALTER, Yoshija. Artificial influencers and the dead internet theory. **AI & Society** (2025) 40:239–240. Disponível em: <https://doi.org/10.1007/s00146-023-01857-0>. Acesso em: 22 fev. 2025. p. 239. No original: “Ten years ago, the theory used to be rather speculative, but with the wake of generative AI, it can now be observed first-hand, and it highlights a disturbing trend: the blurring lines between human and AI-driven interactions. As AI proliferates in generating content, it increasingly shapes user perceptions and behaviors, directing the digital narrative toward its algorithmically determined objectives, often at variance with the inherently unpredictable nature of human discourse. The problem is exacerbated by the fact that many may have difficulty discerning between “real” and “fake” data (or, for that matter, “human-generated” and “AI-generated” ones). Hence, the emergence of deepfakes, a by-product of AI’s advancements, presents challenges to digital trust (Walter 2022). These AI-driven tools, capable of creating highly realistic yet fabricated content, pose a significant threat to the integrity of information online, propelling misinformation and eroding the foundation of trust essential for healthy digital interactions”.

Nesse quadro, seriam praticamente natimortas as chances de se estabelecerem comunicações fortes o suficiente para dar suporte a mobilizações sociais com as de 2010 a 2013. Cederia a esfera pública virtual ao peso de sua estruturação em torno de plataformas construídas e mantidas pelos poderes econômicos e políticos hegemônicos. A esfera pública seria limitada, uma vez mais, aos seus locais físicos de encontro, no cenário em que a credibilidade de informações só seria atingida se fornecida presencialmente e como relato pessoal (“eu vi”, “eu estava lá”).

Evidentemente, o cenário distópico apontado é o ponto culminante de um futuro possível. Apesar dos exemplos concretos de mobilização social de cunho democrático, viu-se também que o mundo virtual é meio de disseminação de ideais autocráticos, ao permitir que muitos indivíduos vivam em bolhas de desinformação, sem contato com pontos de vista divergentes.

Nessa direção, a instantaneidade da comunicação social virtual pode até manter a sua capacidade mobilizatória, mas o senso crítico e o pluralismo são suprimidos. Em seu lugar, fica o mero reflexo da aceleração social, o verdadeiro “presentismo”, entendido como “uma transformação social na experiência da temporalidade, ou seja, uma transformação no modo como a sociedade experimenta a diferença temporal, que é a distinção, que sempre acontece no presente, entre passado e futuro”⁴⁴. O cidadão-reduzido-a-espectador-da-rede fica perdido, sem senso de contexto histórico e sem perspectiva de futuro, totalmente à mercê da manipulação dos algoritmos.

A mobilização social passa a ocorrer em torno uma realidade paralela, artificialmente construída, em que toda sorte de paradoxos passa a fazer sentido, como *protestar* por *menos* liberdade política, *manifestar-se* publicamente pela instauração de regime *autocrático*, defender um *nacionalismo submisso* a outras nações, um discurso de *lei-e-ordem* acolhedor de certos *banditismos* convenientes, e assim por diante. Com a aproximação das *big*

⁴⁴ NEUENSCHWANDER, Juliana; GIRALDES, Marcus. **Democracia sem povo e o novo populismo de direita no Brasil**. In: DE GIORGI, Raffaele *et al.* Direito, Futuro e Risco. D’Placido: São Paulo, 2022, p. 155.

techs ao regime de extrema-direita estadunidense atual, avesso à regulação de redes sociais, não é difícil imaginar a tendência dos conteúdos que se sobressairão.

O que poderia frear, então, a morte da esfera pública virtual? Talvez fazer o que não desejam os grandes conglomerados tecnológicos: regular redes sociais e o desenvolvimento e uso de sistemas de inteligência artificial. Estipular regras firmes contra os discursos de ódio e desinformação, banindo contas que veiculem esse tipo de conteúdo, e restringido o uso de *bots* de automatização de mensagens e coleta de dados dos usuários. Reduzir os vieses, fiscalizar a qualidade de dados de treinamento e restringir a capacidade de geração de conteúdos danosos pela IA generativa. A receita, *in abstracto*, é simples. A sua concretização já depende de outros fatores, inclusive o fortalecimento da resistência democrática fora das redes.

CONSIDERAÇÕES FINAIS

A análise apresentada destaca os riscos inerentes ao avanço da inteligência artificial generativa na esfera pública virtual, especialmente no que tange à disseminação de desinformação e discursos de ódio. A capacidade dessas tecnologias de produzir conteúdos que se assemelham ao real desafia a confiança e a credibilidade das informações *online*, impactando diretamente a formação de opiniões e a mobilização social. A erosão da autenticidade e a proliferação de interações artificiais trazem à tona a “teoria da internet morta”, que, embora anteriormente especulativa, ganha contornos verossímeis diante do cenário atual.

A crescente dificuldade em discernir entre conteúdo gerado por humanos e por IA intensifica a manipulação dos algoritmos sobre os usuários, resultando em uma narrativa digital fabricada que pode distorcer percepções e comportamentos. A emergência das *deepfakes* exemplifica como o avanço da IA pode comprometer a integridade da informação, exacerbando a desconfiança digital e minando as bases de interações saudáveis na rede. Em um ambiente onde a desinformação se propaga com facilidade e velocidade, a capacidade crítica e o pluralismo encontram-se ameaçados, resultando em um público vulnerável à manipulação.

Além disso, a prevalência de bolhas de desinformação e a instantaneidade das redes sociais contribuem para a perda da capacidade de mobilização em torno de causas sociais autênticas, de forma que o potencial emancipatório das redes sociais resta comprometido pela

artificialidade das interações e pela falta de um contexto histórico compartilhado. Nesse sentido, a esfera pública virtual enfrenta o risco de se tornar um reflexo de interesses autocráticos e de um controle algorítmico que suprime a diversidade de vozes.

Para mitigar esses riscos, a regulamentação das redes sociais e dos sistemas de inteligência artificial surge como uma medida necessária. Estabelecer normas contra discursos de ódio e desinformação, banir contas maliciosas e restringir o uso de *bots* são passos fundamentais para preservar a integridade da esfera pública virtual. Além disso, a fiscalização rigorosa dos dados de treinamento e a limitação da geração de conteúdos danosos pela IA são cruciais. A concretização dessas medidas requer um esforço coletivo e o fortalecimento da resistência democrática fora das redes, assegurando que a esfera pública virtual continue a ser um espaço de diálogo e mobilização genuína.

REFERÊNCIAS

BAPTISTA, Eduardo. What is DeepSeek and why is it disrupting the AI sector?. **Reuters (portal eletrônico)**, 2025. Disponível em: <https://www.reuters.com/technology/artificial-intelligence/what-is-deepseek-why-is-it-disrupting-ai-sector-2025-01-27/>. Acesso em: 27 fev. 2025.

BASTOS, Elísio Augusto Velloso. *et al.* Algoritmos, Inteligência Artificial e novas formas de “fazer política”: análise da influência da IA nos processos eleitorais democráticos na contemporaneidade. **Revista Direito e Justiça**, v. 21, n. 41, p. 175-193, 2021. Disponível em: <https://san.uri.br/revistas/index.php/direitojustica/article/view/635/320>. Acesso em 30 jan. 2025.

BARNES & THORNBURG LLP. A Look at U.S. Government’s Changed Approach to Artificial Intelligence Development and Investments. **Barnes & Thornburg LLP (portal eletrônico)**. Disponível em: <https://btlaw.com/insights/alerts/2025/a-look-at-us-governments-changed-approach-to-artificial-intelligence-development-and-investments>. Acesso em 19 fev. 2025.

BENDER, Emily M. et al. On the dangers of stochastic parrots: Can language models be too big?. In: **Proceedings of the 2021 ACM conference on fairness, accountability, and transparency**, 2021, p. 610-623. Disponível em: <https://dl.acm.org/doi/10.1145/3442188.3445922>. Acesso em 20 fev. 2025.

BRASIL. Senado Federal. **Projeto de Lei nº 2338, de 2023 do Senado Federal**. Brasília, 2023. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9881643&ts=1738768184212&disposition=inline>. Acesso em: 24 fev. 2025.

BROWN, Tom. B.; *et al.* **Language Models are Few-Shot Learners**. 2020. ArXiv. Disponível em: <https://arxiv.org/abs/2005.14165v4>. Acesso em 20 fev. 2025.

BUSCHMAN, John. Fake News as Systematically Distorted Communication. **Journal of Documentation**, vol. 80, n. 1, p. 203-217. Disponível em: http://works.bepress.com/john_buschman/101/. Acesso em 24 fev. 2025.

CASTELLS, Manuel. **O poder da comunicação**. Tradução de Vera Lúcia Mello Joscelyne. 4 ed. São Paulo: Paz e Terra, 2019

DI PLACIDO, Dani. The Dead Internet Theory, Explained. **Forbes (portal eletrônico)**, 16 de janeiro de 2024. Disponível em: <https://www.forbes.com/sites/danidiplacido/2024/01/16/the-dead-internet-theory-explained/>. Acesso em 05 mar. 2025.

FAUSTINO, Deivison; LIPPOLD, Valter. **Colonialismo Digital**: por uma crítica hacker-fanoniana. São Paulo: Boitempo, 2023

GOODFELLOW, Ian J; *et al.* **Generative Adversarial Nets**. 2014. Disponível em: <https://arxiv.org/abs/1406.2661v1>. Acesso em: 20 fev. 2025.

LAPAIRE (Bordeaux Montaigne), Jean-Rémy. Why content matters. Zuckerberg, Vox Media and the Cambridge Analytica data leak. **Antares: Letras e Humanidades**, 10(20), 2018, p. 88–110. Disponível em: <https://sou.ucs.br/etc/revistas/index.php/antares/article/view/6583>. Acesso em 20 fev. 2025. p. 89.

LEVITISKY, Steven; ZIBLATT, Daniel. **How democracies die**. Crown: New York, 2018.

LIMA, Melina Moreira Campos. O Direito Internacional sob governos de extrema direita: uma comparação entre as administrações de Donald Trump e Bolsonaro. **Sequência Estudos Jurídicos e Políticos**, Florianópolis, v. 43, n. 90, p. 1–28, 2022. DOI: 10.5007/2177-7055.2022.e66065. Disponível em: <https://periodicos.ufsc.br/index.php/sequencia/article/view/66065>. Acesso em: 21 fev. 2025.

LUHMANN. **O direito da sociedade**. Tradução Saulo Krieger; tradução das citações em latim de Alexandre Agnolon. São Paulo: Martins Fontes - selo Martins, 2016. Livro eletrônico. 2,0 Mb; ePUB.

HABERMAS, Jürgen. **Direito e Democracia**: entre facticidade e validade. Tradução de Flávio Beno Siebeneichler. Rio de Janeiro: Tempo Brasileiro, 2003, v. 2.

HABERMAS, Jürgen. **Teoria do Agir Comunicativo**. Tradução de Paulo Astor Soethe. São Paulo: WMF Martins Fontes, 2012. v. 1.

HOWARD, Philip N.; HUSSAIN, Muzammil M. **Democracy's Fourth Wave?: Digital Media and the Arab Spring**. Oxford University Press: New York, 2013

IMPERVA. **2024 Bad Bot Report**. 2024. Disponível em: <https://www.imperva.com/resources/resource-library/reports/2024-bad-bot-report/>. Acesso em 19 fev. 2025.

MOHAMED, Shakir; PNG, Marie-Therese; ISAAC, William. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. **Philosophy & Technology**, n. 33, p. 659–684, 2020. Disponível em: <https://doi.org/10.1007/s13347-020-00405-8>. Acesso em 20 maio 2024.

NEUENSCHWANDER, Juliana; GIRALDES, Marcus. Democracia sem povo e o novo populismo de direita no Brasil. In: DE GIORGI, Raffaele *et al.* **Direito, Futuro e Risco**. D'Placido: São Paulo, 2022.

PAPACHARISSI, Zizi. The virtual sphere: the internet as a public sphere. p. 21. **New Media Society**, 2002, v. 4, n. 9. Sage Publications. DOI: 10.1177/14614440222226244. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/14614440222226244>. Acesso em: 30 jan. 2025.

PRAZERES, Leandro. Como aliança entre Trump e big techs aumenta pressão sobre governo Lula e STF. **Portal BBC**. Brasília, 21 de janeiro de 2025. Disponível em: <https://www.bbc.com/portuguese/articles/cx2k80ekl3ko>. Acesso em 20 fev. 2025.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. **Inteligência Artificial e Direito**. Curitiba: Alteridade, 2019, p. 88-100.

RADFORD, Alec; *et al.* **Improving Language Understanding by Generative Pre-Training**. 2018. OpenAI. Disponível em: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf. Acesso em 20 fev. 2025.

RAMESH, Aditya; *et al.* (2021). **Zero-Shot Text-to-Image Generation**. 2021. Arxiv. Disponível em: <https://arxiv.org/abs/2102.12092v2>. Acesso em 21 fev. 2025.

RECUERO, Raquel; ZAGO, Gabriela; BASTOS, Marco Toledo. O discurso dos #ProtestosBR: análise de conteúdo do Twitter. **Galáxia**, São Paulo, v. 14, n. 28, p. 199-216, Dec. 2014. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1982-25532014000200017&lng=en&nrm=iso. DOI: <https://doi.org/10.1590/1982-25542014217911>. Acesso em: 21 jan. 2025.

RODRIGUES, Fernanda. Riscos e danos associados à I.A. generativa e a busca por uma tipologia. In: SILVA, Tarcizio (org.). **Inteligência Artificial Generativa: discriminação e impactos sociais**, p. 31-38. Online: Desvelar, 2024. Disponível em desvelar.org. Acesso em: 24 fev. 2025.

SILVA, Tarcizio. Papagaios estocásticos: a revanche da epistemologia da ignorância. In: SILVA, Tarcizio (org.). **Inteligência Artificial Generativa: discriminação e impactos sociais**, p. 57-61. Online: Desvelar, 2024. Disponível em desvelar.org. Acesso em: 24 fev. 2025.

SINGH, Pankaj. Explore These 10 GPT-4 Open-Source Alternatives. 20 nov. 2024. **Analythics Vidhya**. Disponível em: <https://www.analyticsvidhya.com/blog/2024/04/gpt-open-source-alternatives/>. Acesso 27 fev. 2025.

STACKADEMIC. Understanding the Difference Between GPT and LLM: A Comprehensive Comparison. 2023. Disponível em: <https://stackademic.com/blog/understanding-the-difference-between-gpt-and-llm-a-comprehensive-comparison-1f624c713507>. Acesso em 14 fev. 2025.

THEOCHARIS, Yannis; LOWE, Will; DETH, Jan W. van; GARCÍA-ALBACETE, Gema. (2015) Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. **Information, Communication & Society**, 2015, v. 18, n. 02, p. 202-220, DOI: 10.1080/1369118X.2014.948035. Disponível em: <http://www.tandfonline.com/loi/rics20>. Acesso em: 10 jan. 2025.

THORNHILL, Chris. **The Sociology of Law and the Global Transformation of Democracy**. Cambridge University Press, 2018.

WALTER, Yoshija. Artificial influencers and the dead internet theory. **AI & Society** (2025) 40:239–240. Disponível em: <https://doi.org/10.1007/s00146-023-01857-0>. Acesso em: 22 fev. 2025.