



B1

ISSN: 2595-1661

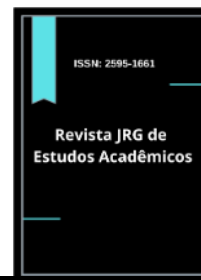
ARTIGO DE REVISÃO

Listas de conteúdos disponíveis em [Portal de Periódicos CAPES](#)

Revista JRG de Estudos Acadêmicos

Página da revista:

<https://revistajrg.com/index.php/jrg>



Democracia em xeque: inteligência artificial e *Deep Fake*

Democracy in question: artificial intelligence and Deepfake

DOI: 10.55892/jrg.v7i15.1743

ARK: 57118/JRG.v7i15.1743

Recebido: 02/12/2024 | Aceito: 09/12/2024 | Publicado *on-line*: 13/12/2024

Orlando Lyra de Carvalho Júnior¹

<https://orcid.org/0000-0002-0281-9832>

<https://lattes.cnpq.br/4578120012479961>

Universidade Federal de Juiz de Fora, UFJF, Brasil

E-mail: orlandolyra@gmail.com

Stela Silva Carvalho²

<https://orcid.org/0000-0002-9075-680X>

<https://lattes.cnpq.br/4578120012479961>

Universidade Federal de Juiz de Fora, UFJF, Brasil

E-mail: stelasilvacarvalho@gmail.com

Bárbara Andrade de Sousa³

<https://orcid.org/0009-0006-5429-7008>

<https://lattes.cnpq.br/5829486489256053>

Universidade Federal de Juiz de Fora, UFJF, Brasil

E-mail: barbaraandrade.trad@gmail.com



Resumo

Este artigo fornece uma estrutura conceitual que permite o mapeamento de áreas de contato importantes entre IA e democracia e conecta os usos e efeitos da IA a debates normativos relevantes na teoria política que podem fornecer orientação na avaliação dos usos e impactos da IA. No futuro, as amplas áreas de contato conceituais de autogoverno, igualdade, eleições e competição entre sistemas podem servir como grupos tópicos para trabalhos futuros. O trabalho futuro fornecerá um relato mais detalhado e avançará teorias que explicam os padrões de uso e efeito. Os cientistas sociais precisam considerar a IA em sua análise de características, perigos e potenciais da democracia contemporânea. Ao fazer isso, eles precisam refletir o funcionamento interno e os efeitos específicos do domínio da tecnologia subjacente. Ao mesmo tempo, cientistas da computação e engenheiros precisam considerar as consequências para a democracia no desenvolvimento e emprego da IA. Isso significa focar não apenas na análise da tecnologia em si, mas também considerar sua inserção em estruturas econômicas, políticas e sociais que mediam seus efeitos para melhor ou pior. Isso torna a análise do impacto da IA na democracia uma área importante de trabalho interdisciplinar futuro.

Palavras-chave: Democracia. *Deep Fake*. Inteligência artificial.

¹ Doutor em Ciências Sociais pela Universidade Federal de Juiz de Fora e em Política Comparada pela Otto-Friedrich Universität Bamberg, Alemanha. Pesquisador do Núcleo de Pesquisa em Comunicação Política da Universidade Federal de Juiz de Fora, tem experiência em Direito Público, sociologia e comunicação política.

² Mestre e doutorando em Psicologia. Universidade Federal de Juiz de Fora.

³ Bacharelado em Letras, Universidade Federal de Juiz de Fora (UFJF).

Abstract

This article provides a conceptual framework that enables the mapping of significant areas of intersection between AI and democracy and connects the uses and effects of AI to relevant normative debates in political theory that can guide the evaluation of AI's applications and impacts. In the future, the broad conceptual intersections of self-governance, equality, elections, and competition between systems may serve as thematic clusters for future research. Future work will provide a more detailed account and advance theories explaining patterns of use and effect. Social scientists need to consider AI in their analysis of the characteristics, dangers, and potentials of contemporary democracy. In doing so, they must reflect on the inner workings and domain-specific effects of the underlying technology. Simultaneously, computer scientists and engineers need to consider the consequences for democracy when developing and deploying AI. This involves focusing not only on analyzing the technology itself but also considering its integration into economic, political, and social structures that mediate its effects, for better or worse. This makes the analysis of AI's impact on democracy an essential area for future interdisciplinary work.

Keywords: Democracy. Deepfake. Artificial Intelligence.

1. Introdução

Há apenas alguns anos, a ferramenta mais utilizada na guerra global da desinformação era os assim chamados "memes",⁴ isto é, "sinais" com enorme potencialidade de se tornarem virais na Internet: uma simples imagem, um vídeo ou texto compartilhado nas redes sociais. Os memes são caracterizados como formas "fáticas" de interação midiática de baixo nível informacional. No entanto, observações sobre seus padrões de interação nas redes sociais revelaram uma densidade surpreendente de interações fáticas, através das quais os memes são compartilhados sem determinação a priori de significado. A alta "viralidade" dos memes, associada aos conteúdos "vazios", ou seja, não determinados a priori, passaram a ser utilizados como arma na guerra global da desinformação: a Rússia de Vladimir Putin usa memes para influenciar as eleições na América e Europa; a China emprega memes para denigrir opositores negacionistas de todo o mundo continuam questionar a eficácia de vacinas contra as pandemias globais através de memes.

A mudança paradigmática provocada pelo advento da inteligência artificial (IA), no entanto, fez surgir novos e poderosos métodos propulsores de desinformação nas redes sociais: as chamadas deepfakes, isto é, vídeos, áudios e imagens sinteticamente alterados através de IA. Atualmente a tecnologia de produção de deepfake encontra-se à disposição de qualquer indivíduo com conhecimento básico de computação. Ela tem sido empregada como arma publicitária para manchar a reputação de qualquer pessoa através de um recurso discursivo voltado a atrair atenções, holofotes, likes e visibilidade, sem relação mais efetiva com a realidade. Em um caso recente, ocorrido em Itararé, no interior de São Paulo, dezenas de jovens e adolescentes tiveram seus rostos colocados em nudes digitais que passaram a ser compartilhados por estudantes de uma escola estadual por meio das redes sociais.

Deepfakes, também conhecidos como mídia sintética ou IA generativa, capturaram a imaginação de alguns e causaram medo em outros. Deepfakes — imagem, áudio ou vídeo sintetizados automaticamente por um sistema de aprendizado de máquina — são as últimas de uma longa linha de técnicas usadas para manipular

⁴ O termo foi usado pelo biólogo Richard Dawkins em sua obra, "O Gene Egoísta" (1976). A ideia do autor se relacionava aos processos de replicação e evolução cultural de ideias, propondo uma espécie de evolucionismo cultural.

a realidade. Sua introdução, no entanto, apresenta novos riscos devido ao acesso democratizado a ferramentas sofisticadas de manipulação. Uma preocupação particular é como os deepfakes serão transformados em armas na política.

No Brasil, Deepfakes têm sido utilizadas contra líderes políticos, religiosos e celebridades. Na campanha eleitoral de 2024 para prefeitura da cidade de São Paulo, a candidata do Partido Socialista Brasileiro (PSB), Tabata Amaral, foi alvo de Deep Nude. As imagens manipuladas por AI mostravam uma mulher em poses eróticas, acompanhada de uma minibiografia, na qual a vítima era cinicamente descrita como "polítologa" e "ativista pela educação".

Tais desafios tornam imperativo uma revisão dos esforços em andamento para detectar e combater os deepfakes e a "infodemia", isto é, a pandemia de desinformação no Brasil. O presente artigo oferece uma breve revisão das principais tecnologias de IA, as quais fornecem a base para manipulação e divulgação de vídeos falsos, clonagem de voz e texto generativo. Destaca-se ainda as ameaças que tais tecnologias representam para as instituições democráticas, sobretudo para o sistema eleitoral.

2. A crise da democracia liberal

Desde a eleição de Donald Trump em 2016, e sua reeleição em 2024, uma onda de pessimismo se alastrou no Ocidente como fantasma ameaçador da democracia liberal. Ao contrário de golpes antidemocráticos tradicionais, hoje a corrosão ocorre dentro das próprias instituições democráticas, de forma gradual e sem violência. Em outras palavras, "a democracia chega ao fim pelo próprio mecanismo do voto" (Runciman, 2018, p. 7).

Segundo Timothy Snyder (2019), a guinada autoritária global foi fortemente influenciada pela máquina de desinformação do autocrata russo, Vladimir Putin. O autor enfatiza o modo como a Rússia vêm atuando nas redes sociais através de robôs, em particular no Telegram, X (Twitter) e Facebook, para minar as democracias liberais através de difamação, calúnia e discurso de ódio. O autor chegou à conclusão de que, hoje, a desinformação em massa representa o maior risco para o regime liberal, dado que "o autoritarismo se fortalece não porque os eleitores dizem o que pensam, mas porque perdem a capacidade de distinguir entre a verdade dos fatos e a falsidade da mentira" (idem, p. 21).

Há evidências de que a crise da democracia é universal. O relatório de 2021, do *Freedom House*, revela que apenas 21 países, dos 193 listados pela ONU, alcançaram a chamada "democracia plena". Em contrapartida, 63 outros foram considerados "democracias falhas", 40 "regimes híbridos", e 69 "regimes autoritários". É considerado "democracia plena" o país onde as liberdades civis e as políticas fundamentais são respeitadas e reforçadas por uma cultura política que conduz ao fomento dos valores democráticos. O arcabouço institucional dessas democracias compreende, entre outros, alternância de poder, eleições livres, sistema efetivo de freios e contrapesos, poder judiciário independente, governos funcionais, mídia diversificada e independente (Repucci e Slipowitz, 2021).

O que se verifica atualmente é que autocratas de todo mundo exploram os conflitos inerentes ao jogo democrático, sobretudo no que respeita as contradições do neoliberalismo econômico, para vencer as eleições e, uma vez no poder, tentam minar os pilares da própria democracia, como a independência do Judiciário e a liberdade de imprensa e opinião. Em muitos países as centelhas de esperança por liberdade se extinguem paulatinamente, contribuindo assim para o surgimento de um novo *status*

quo global, em que os atos de repressão ficam impunes e os defensores da democracia são gradualmente aliados da cena política.

Segundo o professor e pesquisador da Universidade John Hopkins, Yacha Mounk (2018), a desigualdade social e a concentração de renda das elites econômicas mundiais criaram um fosso intransponível que as isolou ainda mais da população em geral. O resultado é que o eleitor médio se sente cada vez mais alienado da política e dos políticos. Ainda que a participação popular perfeita nunca tenha sido uma realidade plena, o pesquisador afirma que tais questões põem a nu os dilemas do liberalismo, bem como a capacidade dos eleitores influenciarem a política. Daí o forte apelo populista em todo mundo e o conseqüente fim do “mito fundador da democracia”.

Ainda segundo Mounk (2018), Donald Trump representa um tipo diferente de populista: ao invés de propor políticas públicas demagógicas, que aumentem as oportunidades socioeconômicas dos menos favorecidos, em oposição às elites financeiras, Trump seria um “sadopopulista”, cujas políticas se destinam a prejudicar a parte mais vulnerável do seu próprio eleitorado. Para o autor, tal “sadopopulismo” põe em xeque os fundamentos do sistema democrático norte-americano (Mounk, 2018, p. 43-54).

É neste contexto de crise institucional que a presente pesquisa pretende analisar as predisposições autoritárias geradas pela cascata de desinformação e pela sedução que as Deep Fakes exercem sobre milhões de pessoas em todo mundo. Tal sedução encontra raízes em um fenômeno profundo da psiquê humana conhecido como “viés de confirmação” (Jussim, 1986). Trata-se da inclinação que todos temos de acreditar naquilo que nos convém.

Tal fenômeno torna a desinformação um problema complexo, uma vez que as pessoas mal informadas raramente estão dispostas a aceitar a evidência dos fatos. Para muitos, mudar as crenças pessoais põe em risco a própria identidade, além de ameaçar o sentimento de pertença a um determinado grupo, pois implica romper laços sociais arraigadas, sobretudo em um ambiente de polarização política. Assim, a aceitação ou recusa das Fake News e das Deep Fake não depende da “verdade” dos fatos, mas de como ela é percebida (Nyhan e Reifler, 2010).

No campo científico, ocorre a marginalização e perda de autonomia dos cientistas na era da “pós-verdade”: a interpretação de fatos e informações de acordo com os vieses e as crenças de cada indivíduo. Por exemplo, desinformação sobre as mudanças climáticas, percebidas como farsa; lâmpadas fluorescentes que provocam câncer e vacinas contra o Coivd19 que causam autismo nas crianças. Segundo Pivaro e Giroto Júnior (2022), a era da pós-verdade atinge principalmente as instituições e os cientistas que dependem de observações objetivas para garantir a confiabilidade de seus achados.

3. Inteligência Artificial e Deepfake

O sucesso gerado pela criação de grandes modelos de linguagem na esteira do ChatGPT, em diversas plataformas globais, aumentou as especulações sobre seus possíveis benefícios e malefícios econômicos, sociais e políticos. Atualmente, os softwares que usam algoritmos de Inteligência Artificial (AI) podem facilmente manipular vídeos de qualquer indivíduo e posta-los nas redes sociais, seja para promover ou destruir reputações: são conhecidos pela expressão inglesa “deepfakes”, um amálgama de “Deep learning” (aprendizagem profunda) e “fake” (falso), isto é, a manipulação por inteligência artificial (IA) de vídeos, áudios ou imagens com o intuito de lhes conferir uma falsa aparência de realidade.

É preciso ressaltar que cada novo passo no desenvolvimento e aplicação da Inteligência Artificial (AI) vem acompanhado de intenso debate sobre os riscos eventuais que uma inteligência “sobre-humana” representaria para a própria sobrevivência do *homo sapiens*. Tais expectativas tem alimentado uma acirrada discussão sobre o real impacto da IA, a qual é pautada tanto por receios desfocados quanto por entusiasmos utópicos. Por isso, é necessário distinguir os diversos tipos de IA e suas potencialidades.

A primeira distinção envolve a Inteligência Artificial Geral (IAG) e a Inteligência Artificial Restrita (IAR). Essa última refere-se a um sistema de máquinas com cognição de nível humano e capacidade de definir metas de forma autônoma, além de “raciocinar” e se adaptar a circunstâncias imprevistas. Os avanços na capacidade de autoaprendizado (Self learning) e aprendizado profundo (Deep learning) têm causado preocupações no que respeita a diferentes aspectos do sistema democrático, incluindo possíveis impactos sobre as condições de autogoverno e sobre as oportunidades de os eleitores exercê-lo através de eleições livres, periódicas e competitivas.

Outro fator de preocupação é o fácil acesso às tecnologias de IA que geram deepfakes e deepnudes. De acordo com pesquisadores, até mesmo as interfaces de programação de aplicativos (APIs) considerados seguros já foram ludibriadas por deepfake. Há relatos de que 78% de conteúdos deepfake conseguiram penetrar com sucesso os APIs da Microsoft e da Amazon.com. Até hoje, as APIs da Microsoft e da Amazon têm lutado para diferenciar entre impostores deepfake e conteúdos reais.

Vídeos deepfake incluem filmagens modificadas sinteticamente que apresentam alterações nos rostos ou corpos dos sujeitos. As imagens desses vídeos sintéticos são desenvolvidas por meio de redes adversárias generativas (Generative Adversarial Networks - GANs). Tianxiang Shen et al., (2018, p. 25) fornecem uma excelente descrição de como as GANs funcionam para criar conteúdo sintético:

“O sistema GAN consiste em uma fonte geradora que produz imagens a partir de ruídos aleatórios e um discriminador que julga se uma imagem de entrada é autêntica ou produzida pelo gerador. Os dois componentes são funcionalmente adversários e desempenham papéis opostos, como o de um criminoso e de um policial. Após um período de treinamento, o gerador pode produzir imagens falsas com alta fidelidade”.

Desde que Ian Goodfellow e colegas criaram o sistema GAN em 2014, os vídeos deepfake se tornaram cada vez mais convincentes. Na primavera de 2021, uma conta do TikTok (Tom [@deptomcruise], 2021) lançou uma série de vídeos deepfake altamente realistas do que parecia ser Tom Cruise falando. Naquela época, o vídeo teve mais de 15,9 milhões de visualizações e provocou intenso debate no meio acadêmico sobre os possíveis impactos que a nova tecnologia poderia causar no campo da desinformação. Os deepfakes de Tom Cruise surgiram na esteira de uma série de vídeos que apresentavam Barack Obama usando palavras e um discurso de Richard Nixon que ele nunca fez. A cada geração, a qualidade dos deepfakes se torna cada vez mais realista dado que seus componentes sintéticos são mais difíceis de detectar a olho nu.

Alguém poderia perguntar: por que os marqueteiros usariam imagens falsas em campanhas eleitorais? Por que a alternativa seria usar imagens subtraídas de pessoas reais, mas há ferramentas tecnológicas que identificam imagens de tais perfis. Especificamente, é possível usar a busca reversa de imagens do Google para

escanear a internet em busca de uma foto suspeita e identificar sua progênie. Assim, usar fotos manipuladas por AI permite contornar as barreiras defensivas de fotos autênticas.

Por exemplo, várias páginas da web agora oferecem gratuitamente acesso a serviços deepfake e Deep Nude. Sites populares incluem: <https://fakeyou.com/ai-live-portrait> e <https://muke.ai/undress-ai/> permitem desnudar fotos de qualquer indivíduo; <https://www.myheritage.com.br/> anima fotos de parentes falecidos; Mais notoriamente, a web Deep Nude (<https://easywithai.com/ai-nsfw-tools/n8ked-deepnude-generator/>), disponibiliza fotos reais de homens e mulheres para serem manipuladas e transformadas em Deep Nude.

4. Inteligência Artificial, Governança, Governabilidade e Accountability

A governança e a governabilidade estão diretamente relacionadas: enquanto a governabilidade é a capacidade política de governar e equilibrar as demandas sociais e transformá-las em políticas públicas que cumpram com os objetivos institucionais, a governança é a competência técnica de implementá-las com eficiência e eficácia. Na ausência de condições políticas necessárias para a governabilidade, a governança se torna inviável. Em contrapartida, uma má governança gera perda da legitimidade e pode até provocar uma queda do governo.

Outro princípio básico da democracia é a accountability, isto é, governantes devem prestar contas a quem os elegeu. Princípio normativo de legitimidade, a accountability se baseia na ideia prática de que a tomada de decisão compartilhada é superior a outras formas mais centralizadas de governo. O termo diz respeito ao dever de prestação de contas pelo administrador público e tem como premissa a transparência dos atos de gestão do governo, permitindo a responsabilização dos gestores pelos resultados entregues à sociedade. Tal transparência permite o controle por parte da população, que deve fiscalizar a gestão dos recursos públicos (Dahl, 1998).

A legitimidade da democracia, portanto, está intimamente ligada à ideia de que os eleitores sejam capazes de tomar decisões informadas para si mesmos e para suas comunidades. Isso depende em larga medida do ambiente informacional no qual estão inseridos. Ora, as redes sociais e as deepfakes geradas pela Inteligência Artificial afetam diretamente os fundamentos informativos de qualquer governo. Isto inclui a forma como as pessoas são expostas à informação na mídia tradicional e nas redes sociais, dado que a modelagem algorítmica da informação produz um enviesamento limitante, sobretudo do debate político.

Além disso, a IA pode também impactar a produção e distribuição de informações políticas, ao alterar as condições econômicas de sua produção. Muitas agências de notícias passaram a usar a IA para produzir textos, imagem e vídeo gerados por IA. Isto pode acelerar as tendências já existentes de geração automatizada de conteúdo de ultra baixo custo, o que aumenta potencialmente a pressão sobre o salário dos jornalistas, bem como sobre a audiência e os lucros das empresas noticiosas (Nielsen, 2020; Scharrow et al., 2020).

Mais problemática ainda é o uso da IA para produzir deepfake em grande escala. Isso poderia prejudicar o exercício da cidadania não só através de falsificações destinadas a enganar as pessoas, mas também pela inundação dos ambientes informacionais com uma enxurrada de conteúdos não confiáveis gerados por IA. Isso diluiria o acesso às informações cruciais para o exercício da accountability, governança e governabilidade.

5. Inteligência Artificial e Eleições

As democracias dependem de eleições livres e periódicas para o gerenciamento dos conflitos políticos, ao fornecer aos diversos partidos a oportunidade de chegar ao poder dentro de uma estrutura institucional. Tal arcabouço legal só funciona se houver, de fato, a possibilidade da conquista do poder dentro de um sistema de "incerteza organizada" (Przeworski, 1991, p. 13). O emprego da IA pode diminuir tal incerteza ao prevê com máxima probabilidade quem ganhará e quem perderá uma determinada eleição.

Deepfakes podem aparecer em qualquer plataforma de mídia social. De acordo com o ditado de que "mentiras viajarão pelo mundo antes que a verdade possa calçar seus sapatos", deepfakes podem ser amplamente compartilhados mesmo antes que qualquer medida legal seja tomada. Uma vez notificadas pelas agências reguladoras, as plataformas de mídia social demoram para agir ou contestam a ordem judicial alegando a liberdade de expressão.

A Lei federal especificamente sobre desinformação é escassa em muitos países sede das grandes plataformas, como o X, Youtube, Facebook, TikTok, entre tantas. Alguns países da União Europeia aprovaram leis que tornam crime criar e compartilhar Fake News e Deepfakes com motivação política durante a temporada eleitoral. Não está claro se essas leis se manteriam nos Tribunais, especialmente se os criadores puderem argumentar que suas postagens se qualificam como sátira protegida constitucionalmente. Mesmo que a condenação se mantenha, as penalidades só ocorreriam depois que o dano já tivesse sido feito.

No Brasil, o Supremo Tribunal Federal, através da Resolução nº 742 de 27 de agosto de 2021, instituiu o Programa de Combate à Desinformação no âmbito do Supremo Tribunal Federal, estabelecendo que:

Art. 1º: Fica instituído o Programa de Combate à Desinformação (PCD) no Supremo Tribunal Federal (STF), com a finalidade de enfrentar os efeitos negativos provocados pela desinformação e pelas narrativas odiosas à imagem e à credibilidade da Instituição, de seus membros e do Poder Judiciário, a partir de estratégias proporcionais e democráticas, a fim de manter a proteção da Corte acerca das liberdades de comunicação (BRASIL, 2021).

Esse programa de combate as Fake News, é resultado da abertura do Inquérito das Fake News, ocorrido em 2019, pelo Supremo Tribunal Federal, com fundamento no Regimento Interno da Corte, que diz:

Art. 43. Ocorrendo infração à lei penal na sede ou dependência do Tribunal, o Presidente instaurará inquérito, se envolver autoridade ou pessoa sujeita à sua jurisdição, ou delegará esta atribuição a outro Ministro. (BRASIL, 2023).

Uma inovação muito comentada recentemente é a incorporação de marcas d'água em imagens, vídeos e áudio gerados por IA. Mas marcas d'água são facilmente falsificadas e agências de inteligência estrangeiras não as usariam em hipótese alguma. Além disso, o uso generalizado de marcas d'água pode ter um resultado perverso. Os usuários podem presumir que a ausência de uma marca d'água seja prova de autenticidade prova de uma imagem ou áudio. A verdade é que ainda desconhecemos o impacto real das deepfakes geradas por inteligência artificial durante períodos eleitorais.

5.1 O Exemplo americano

Em fevereiro de 2024, uma chamada automática se passando pelo presidente dos EUA, Joe Biden, foi enviada aos eleitores de New Hampshire, aconselhando-os a não votarem nas eleições primárias presidenciais do estado. A voz, gerada por inteligência artificial, soou bem real. "Guarda seu voto para a eleição de novembro", declarou a voz, afirmando falsamente que uma votação nas primárias impediria os eleitores de participarem das eleições gerais de novembro. O incidente da chamada automática reflete uma preocupação crescente de que a IA generativa tornará mais barato e fácil espalhar informações falsas e executar campanhas de desinformação. A Comissão Federal de Comunicações dos Estados Unidos tornou ilegais vozes geradas por IA em chamadas automáticas. O incidente confirma a preocupação de que a IA generativa tornará mais barato e fácil espalhar informações falsas e executar campanhas de desinformação (Mirza, 2024).

O agora eleito Presidente Donald Trump, quando ainda em campanha, quase desistiu de uma entrevista em agosto de 2024 com um grupo de jornalistas negros depois de saber que eles planejavam verificar "fatos" que ele viesse a relatar. No mês seguinte, ele e seus aliados reclamaram repetidamente sobre a verificação de fatos que ocorreu durante seu debate com a vice-presidente Kamala Harris, repreendendo jornalistas e executivos de notícias no meio do debate televisionado. O mesmo ocorreu com a recusa de participar de uma entrevista com o programa "60 Minutes" da rede de televisão CBS News, o qual verifica em tempo real fatos relatados por seus entrevistados (Verma e Zakrzewski, 2024).

**As abordagens baseadas em dados são limitadas na previsão do comportamento individual dos eleitores. Embora o comportamento de votação de partidários comprometidos possa ser previsto com alguma probabilidade (Hersh, 2015; Nickerson & Rogers, 2014) — pelo menos em sistemas bipartidários — prever o comportamento de pessoas que estão apenas fracamente envolvidas com a política é muito mais difícil. As pessoas nem sempre votam e, quando o fazem, o contexto pode variar muito. Suas escolhas de voto não estão disponíveis para modeladores, tornando a previsão automática do comportamento de votação um problema para o qual a IA não é adequada. A incerteza das vitórias eleitorais permanecerá, portanto, no futuro previsível.

Mas as campanhas podem desenvolver outros modelos eleitorais relevantes baseados em dados, como a probabilidade de alguém votar ou doar dinheiro (Hersh, 2015; Issenberg, 2012; Nickerson & Rogers, 2014), o que pode dar às campanhas uma vantagem competitiva. Qualquer vantagem desse tipo provavelmente é passageira, embora que poderia dar às campanhas uma vantagem competitiva. Qualquer vantagem desse tipo provavelmente é passageira, no entanto, dada a ampla disponibilidade de ferramentas baseadas em IA e organizações de campanha aprendendo com os sucessos e fracassos de outros (Kreiss, 2016). Empresas e governos também podem buscar usar IA para prever resultados eleitorais ou as mudanças de humor do eleitorado e possivelmente intervir. Esses esforços são limitados pelos mesmos desafios levantados acima, mas a impressão pública dessa capacidade pode ser suficiente para minar e deslegitimar eleições e dar aos perdedores eleitorais um pretexto para contestar os resultados em vez de admitir.

O suposto papel da Cambridge Analytica na votação do Brexit no Reino Unido e na eleição presidencial dos EUA de 2016 previu alguns dos desafios. Embora haja pouca indicação de que a segmentação psicológica baseada em dados tenha sido amplamente usada ou tenha tido efeitos consideráveis, esses episódios ainda aparecem na imaginação pública como um exemplo do poder percebido da IA na

manipulação eleitoral (Jungherr et al., 2020, pp. 124–130). Podemos esperar que o uso generalizado da IA na vida econômica, política e social mude as expectativas das pessoas sobre seus usos e abusos na campanha eleitoral, independentemente de seus usos reais ou limitações inerentes.

No geral, o impacto da IA nas eleições parece limitado, dada a relativa escassez da atividade prevista — a votação. Embora efeitos indiretos sejam possíveis por meio de oportunidades potenciais para diferenciação competitiva, é duvidoso que isso possa se traduzir em uma mudança consistente e sistêmica de poder, dada a ampla disponibilidade de ferramentas de IA. Mais provável é o impacto indireto mencionado acima: que ao transpor as expectativas sobre os supostos poderes da IA da indústria e da ciência para a política, o público pode vir a acreditar que a IA é realmente capaz de compensar a "incerteza organizada" das eleições democráticas. Isso por si só poderia enfraquecer a confiança pública nas eleições e a aceitação dos resultados eleitorais. Portanto, é importante manter a incerteza organizada viva diante da IA, não a enfraquecer por meio de especulações irresponsáveis e fantásticas.

A tecnologia subjacente tem utilizações benignas, desde aplicações frívolas que permitem trocar rostos com celebridades até algoritmos de aprendizagem profunda significativos (a tecnologia que sustenta as falsificações profundas) que têm sido utilizados para sintetizar novos compostos farmacêuticos e proteger a vida selvagem dos caçadores furtivos.

No entanto, o acesso imediato a tecnologias falsas profundas também permite que cibercriminosos, ativistas políticos e estados-nação criem rapidamente falsificações baratas e realistas. Esta tecnologia reduz os custos de envolvimento na guerra de informação em grande escala e alarga o leque de intervenientes capazes de participar nela. As falsificações profundas representarão o maior risco quando combinadas com outras tecnologias e tendências sociais: aumentarão os ataques cibernéticos, acelerarão a propagação da propaganda e da desinformação online e exacerbarão o declínio da confiança nas instituições democráticas.

Qualquer tecnologia que possa ser usada para gerar conteúdo falso ou enganoso, desde fotocopiadoras e software Photoshop até falsificações profundas, pode ser transformada em arma. Este artigo argumenta que os decisores políticos enfrentam uma janela de oportunidade cada vez mais estreita para minimizar as consequências das falsificações profundas transformadas em armas. Qualquer resposta deve incluir medidas em três linhas de esforço: investimento e implantação de tecnologias de detecção de falsificações profundas que alterem o comportamento online, nomeadamente através de medidas políticas que capacitem o público digital a interagir de forma crítica com o conteúdo e que reforcem canais de comunicação confiáveis criação e aplicação de padrões de autenticação digital.

Notícias falsas são uma das formas de criar mitos hoje. Notícias falsas não são apenas ausência de verdade, mas também presença de mentiras misturadas com informações precisas. Portanto, as notícias podem ser completamente falsas ou parcialmente verdadeiras [2]. Notícias falsas têm se espalhado desde o início da humanidade; na verdade, é tão antigo quanto a própria humanidade [3]. Em termos leigos, as notícias falsas, que estão na moda hoje em dia, referem-se à disseminação intencional ou não intencional de informações falsas destinadas a enganar ou enganar, bem como a criar confusão, tensão e descrença [4]. A contaminação de conteúdos irrelevantes e imprecisos com essas informações tornou-se um problema de nível global, conhecido como poluição informacional ou "infolution" [5]. Não é um conceito novo [6]; no entanto, na era digital, as notícias falsas ganharam um novo impulso e tornaram-se um problema mundial a resolver [7].

Na era digital de hoje, a Internet tornou-se um reservatório de informações para todos. Porém, é preciso saber que nem tudo que está disponível na internet é confiável. Somente o conhecimento pode iluminar nosso caminho para o reconhecimento do conteúdo real; caso contrário, seremos levados a situações indesejáveis [8], pois podem levar a uma falta generalizada de confiança naquilo que vemos e ouvimos online [9]. O surgimento das plataformas sociais online veio alimentar esta chama, uma vez que a utilização das redes sociais passou a fazer parte da rotina de todos para interagirem entre si, bem como para obterem notícias. As informações são espalhadas por causa da rede em nuvem em colaboração com a IT [10]. Plataformas de redes sociais [11,12,13,14], como Facebook, WhatsApp [11] e muitas outras [12], permitem que as pessoas criem grupos de amigos ou familiares para compartilhar conteúdo; posteriormente, quando recebem notícias novas ou emocionantes, nem sempre reservam tempo para verificar os factos antes de partilhar ou comentar. Devido à velocidade com que a informação se espalha online, os utilizadores podem influenciar várias comunidades com informações verdadeiras e falsas [13]. Isto leva à disseminação inconsciente de notícias forjadas. É fato que a maior parte do conteúdo falso chama mais a atenção do que as notícias reais e, portanto, se espalha mais rapidamente [14].

Um estudo recente do Instituto de Massachusetts of Technology (MIT) descobriu que notícias falsas no Twitter se espalham seis vezes mais rapidamente do que notícias reais e têm 70% mais probabilidade de serem retuitadas [15]. A desinformação, as mentiras e as teorias da conspiração predominam em muitas plataformas de redes sociais. As consequências no mundo real deste ambiente online tóxico vão desde o genocídio até à intromissão eleitoral e ameaças à saúde pública global. Segundo estudos, durante as eleições presidenciais de 2016 nos Estados Unidos, as redes sociais foram inundadas com notícias falsas que foram amplamente partilhadas e favoreceram fortemente Donald Trump [16]. A maioria de nós testemunhou notícias falsas e desinformação a serem espalhadas em grande número durante estes tempos de COVID-19 [17], aumentando a confusão numa situação já difícil. Isto consiste em uma abundância de informações sobre a propagação de doenças, algumas das quais são precisas e outras não [18].

Num esforço para combater a infodemia da COVID-19, a OMS e o Reino Unido trabalharam juntos num estudo, revelando que, nos primeiros 3 meses de 2020, a desinformação sobre o coronavírus levou à hospitalização de aproximadamente 6.000 pessoas em todo o mundo [19]. Durante este período, os pesquisadores também suspeitaram que matou pelo menos 800 pessoas [20]. Além disso, em meados de maio de 2020, no meio de uma epidemia mundial, 28% dos americanos pensavam que Bill Gates pretendia utilizar a COVID-19 para implementar uma vacina obrigatória equipada com microchips de rastreio. O público em geral ainda mantém esta afirmação absurda como verdadeira [21].

Conclusão

Embora muitas aplicações de IA ainda estejam no futuro, já começamos a ver o impacto da IA na democracia. É verdade que muitos dos usos e efeitos futuros da IA permanecem incertos. No entanto, é importante que a ciência social se envolva desde o início com a IA e ajude a observar, avaliar e orientar sua implementação. Isso inclui os usos da IA na política, no governo e sua regulamentação e governança. A qualidade da análise dos efeitos da IA na democracia depende da especificidade em relação ao tipo de IA, como ela funciona, as condições para sua implantação bem-sucedida e o(s) aspecto(s) da democracia que ela toca. Narrativas sobre uma AGI

superpoderosa e não especificada e seu suposto impacto na sociedade podem ser uma leitura estimulante, mas oferecem pouco para a análise dos efeitos reais na sociedade ou na democracia. Na verdade, as partes interessadas podem usar a discussão sobre AGI e supostos perigos de eventos de nível de extinção como uma cortina de fumaça, distraindo a atenção pública e regulatória de questões mais mundanas, mas cruciais, de governança de IA, regulamentação e distribuição social de ganhos e riscos impulsionados por IA. Embora a IA seja frequentemente discutida como um perigo ou ameaça, ela também pode fornecer oportunidades para compensar alguns dos desafios contemporâneos à democracia. Pensar abertamente sobre a aplicação da IA na democracia pode fornecer algum alívio desses desafios. Escolhas conscientes de design e auditorias transparentes podem ajudar a amenizar disfunções e revelar vieses. Em geral, o impacto da IA depende da implementação e supervisão do público e reguladores.

Para isso, empresas, reguladores e a sociedade precisam ser explícitos e transparentes sobre quais objetivos econômicos, políticos ou sociais eles querem alcançar usando IA e como seu funcionamento específico pode impulsionar ou inibir essa busca. Por natureza, essa discussão combina argumentos e considerações normativas, mecanicistas e tecnológicas. É importante não se desviar por visões grandiosas, mas em última análise imaginárias, de uma AGI, mas, em vez disso, focar em instâncias específicas de IA restrita, seu funcionamento interno, usos em áreas específicas de interesse e efeitos. Isso inclui a discussão de efeitos potencialmente positivos e negativos. É improvável que a IA tenha impacto direto em muitos aspectos da democracia.

No entanto, o discurso público provavelmente continuará a se concentrar em ameaças, manipulação e mudanças de poder esperadas. Esse discurso e essas expectativas têm o potencial de moldar as atitudes públicas em relação à IA e seu impacto na democracia fortemente, independentemente de sua base factual. Efeitos percebidos podem importar mais fortemente do que efeitos reais. Pesquisadores têm a responsabilidade de não atirar as chamas do discurso com peculato, mas, em vez disso, permanecer focados no funcionamento e efeitos reais da IA. A ampla aplicação da IA na sociedade e na política está apenas começando; desenvolvimentos futuros são, é claro, desconhecidos, e, portanto, a forma específica que a tecnologia de IA tomará, a aplicação da IA na sociedade e seus efeitos subsequentes na democracia devem ser monitorados continuamente. Este artigo identifica áreas de contato entre IA e democracia que permitem rastrear e monitorar esse processo. Mais especificamente, há uma série de oportunidades para trabalho futuro. Em um nível muito básico, há uma necessidade de um trabalho mais sistemático sobre como a IA é empregada na política e pelos governos (Engstrom & Haim, 2023). Existem alguns estudos iniciais — em, por exemplo, a área de policiamento preditivo e o sistema legal (Brayne, 2021; Ferguson, 2017) — mas eles são comparativamente poucos e predominantemente focados nos Estados Unidos. Existem oportunidades para estudos de outros países e para trabalho comparativo. Além disso, muito do que achamos que sabemos sobre IA na sociedade é baseado em relatos jornalísticos, mas embora eles possam ser profundamente instrutivos, há uma necessidade consistente de relatos científicos desses desenvolvimentos em diferentes setores e sociedades. Há também um fluxo constante de estudos examinando os vieses de sistemas específicos de IA em contextos particulares, como auditorias de análises de textos e imagens orientadas por IA e a implementação de IA em processos sociais, como policiamento, sentenças judiciais, decisões de crédito ou medicina (Bolukbasi et al., 2016; Buolamwini & Geburu, 2018; Caliskan et al., 2017; Mayson, 2019; Mehrabi et al.,

2022; S. Mitchell et al., 2021; Obermeyer et al., 2019). No entanto, a IA também pode ser usada para estudar vieses históricos na sociedade examinando conjuntos de dados em larga escala de textos e imagens históricas para identificar a distribuição e mudanças na representação de grupos sociais específicos — por exemplo, ao longo de linhas de gênero ou raciais (Jürgens et al., 2022) — ou, mais geralmente, identificar para análise posterior casos promissores nos quais dados empíricos contradizem expectativas baseadas em modelos (Munk et al., 2022; Rettberg, 2022). Com base nisso, também é necessário um trabalho sistemático sobre a regulamentação da IA em diferentes países e contextos (Veale et al., 2023). Embora estes sejam os primeiros dias para a regulamentação oficial da IA, já existem atividades interessantes — notavelmente, suporte governamental para pesquisa e desenvolvimento focados em IA.

A União Europeia e os governos dos EUA e da China identificaram a IA como uma área de competição geopolítica e econômica estratégica. Examinar iniciativas políticas concorrentes visando impulsionar a inovação é um passo inicial promissor para examinar as relações dos governos com a IA e políticas e regulamentações relacionadas. Estudos comparativos anteriores da regulamentação de privacidade e modelos de negócios de plataforma podem oferecer modelos para trabalho focado em IA (Bradford, 2023; Cohen, 2019; Farrell & Newman, 2019; Thelen, 2018).

Finalmente, o desenvolvimento da IA como um campo científico e comercial é um assunto promissor de pesquisa em si. A natureza altamente interdisciplinar e as mudanças na vanguarda do desenvolvimento da IA entre disciplinas científicas, academia e negócios e áreas geográficas tornam este um caso muito interessante do nexos contemporâneo de ciência, comércio e sociedade (Lee, 2018; McCorduck, 2004; Metz, 2021). Indo além disso, a IA é cada vez mais usada nas ciências sociais como um método de descoberta. As mudanças associadas no fluxo de trabalho científico, perguntas feitas e abordagens para teorizar são uma área fértil para reflexão futura (Grimmer et al., 2021; Mökander & Schroeder, 2022). Existem muitos caminhos promissores para o futuro trabalho científico sobre o impacto da IA na democracia. Aqui, é importante combinar insights de diferentes campos. Relatos puramente tecnológicos correm o risco de superestimar o impacto da IA nos sistemas sociais, devido à sua limitação e ao papel das estruturas sociais.

Relatos vindos puramente das ciências sociais correm o risco de deturpar o funcionamento real da IA existente e, assim, atribuir erroneamente suas consequências. A estrutura conceitual apresentada aqui oferece uma abordagem para combinar essas perspectivas interdisciplinares de forma produtiva. O impacto da IA na democracia já está progredindo. Seu exame e discussão sistemáticos e interdisciplinares também precisam prosseguir.

Referências

APPLEBAUM, Anne. **Twilight of democracy**: The seductive lure of authoritarianism. New York: Penguin Random House, 2020.

LEVITSKY, Steven e ZIBLATT, Daniel. **Como as democracias morrem**. Rio de Janeiro: Zahar, 2018.

MIRZA, R. How AI deepfakes threaten the 2024 elections. **The Journalist Resource**, v. 2(6): February, 2024. Disponível em: <https://journalistsresource.org/home/how-ai-deepfakes-threaten-the-2024-elections/>. Acesso em: out./2024.

MOUNK, Yascha. **The People vs. Democracy**: Why Our Freedom Is in Danger and How to Save It. Boston: Harvard University Press, 2018.

REPUCCI, Sarah e SLIPOWITZ, Amy. **Democracy under Siege**. New York: Freedom House, 2024.

RUNCIMAN, David. **Como a democracia chega ao fim**. São Paulo: Todavia, 2018.

SNYDER, Timothy. **Na contramão da liberdade**: a guinada autoritária nas democracias liberais. São Paulo: Cia. das Letras, 2019.

SODRÉ, Muniz. **Antropológica do espelho: uma teoria da comunicação linear e em rede**. Petrópolis, RJ; Vozes; 2009. 268 p.

PARIS, Britt e JOAN, Donovan. **Deepfakes and Cheapfakes**: The Manipulation of Audio and Visual Evidence. <https://datasociety.net/library/deepfakes-and-cheap-fakes/>, 2019.

PÉRES-CURIEL, C. e BAPTISTA, J. B. **Lying on social media**: Disinformation strategies of Iberian populist radical right. *The Routledge Handbook of Political Communication in Ibero-America*. London: Routledge, 2024.

SANTOS, E. O que os nudes falsos em Itararé ensinam sobre o aumento de casos de deepfakes pornográficos em escolas. G1, 2024. Disponível em : <https://g1.globo.com/educacao/noticia/2024/10/11/o-que-os-nudes-falsos-em-itarare-ensinam-sobre-o-aumento-de-casos-de-deepfakes-pornograficos-em-escolas.ghtml>. Acesso em out./2024.

VERMA, P. e ZAKRZEWSKI, C. Deepfakes threaten to upend global elections. No one can stop them. **The Washington Post**, 23/abril/2024. Disponível em: AI deepfakes are already hitting elections. We have little protection. - The Washington Post. Acesso em: 09/11/2024.