XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 5 – Política e Economia da Informação

INTELIGÊNCIA ARTIFICIAL GENERATIVA E DESINFORMAÇÃO: DESAFIOS PARA A CIÊNCIA DA INFORMAÇÃO

GENERATIVE ARTIFICIAL INTELLIGENCE AND DISINFORMATION: CHALLENGES FOR INFORMATION SCIENCE

Pollyany Annenberg Nascimento Gomes – Universidade Federal de Alagoas (UFAL)
Maria Lívia Pachêco de Oliveira – Universidade Federal da Paraíba (UFPB), Universidade Federal de Alagoas (UFAL)

Modalidade: Trabalho Completo

Resumo: Este artigo investiga os impactos da inteligência artificial (IA) generativa na integridade da informação e explora como as ferramentas, como o ChatGPT e o DALL-E, desafiam a confiabilidade dos conteúdos produzidos por essa tecnologia. Por meio de pesquisa qualitativa e estudos bibliográficos com contribuições inerentes à ética e à autoria da IA, são analisadas as consequências da disseminação de textos manipulados e de deepfakes em decisões individuais e políticas. Conclui-se que os casos exemplificados no artigo revelam a dificuldade em distinguir entre desinformação e informações consideradas legítimas. Também foi possível perceber que a conscientização pública, a regulamentação eficaz e o uso ético da IA são essenciais para enfrentar continuamente os desafios da desinformação e proteger a integridade da informação.

Palavras-chave: desinformação; inteligência artificial generativa; lei da inteligência artificial no Brasil.

Abstract: This article investigates the impacts of generative artificial intelligence (AI) on information integrity and explores how tools like ChatGPT and Gemini challenge the reliability of content produced by this technology. Through qualitative research and bibliographic studies with inherent contributions to AI ethics and authorship, the consequences of disseminating manipulated texts and deepfakes in individual and policy decisions are analyzed. The cases exemplified in the article reveal the difficulty in distinguishing between disinformation and real information. It has also been observed that public awareness, effective regulation, and ethical use of AI are essential to continuously address the challenges of disinformation and safeguard information integrity.

Keywords: disinformation, generative artificial intelligence; artificial intelligence law in Brazil.

1 INTRODUÇÃO

A ciência da informação, tradicionalmente preocupada com o acesso e uso ético da informação (Araújo, 2020), confronta atualmente o desafio da desinformação, que cria, distorce e manipula informações com o objetivo de influenciar opiniões e comportamentos. Para as pesquisas em informação, é essencial compreender o impacto das tecnologias na geração e disseminação de informação para mitigar os efeitos prejudiciais desse fenômeno, já que, como considera Araújo (2020), essa atividade não se limita mais à compreensão técnica ou cognitiva da informação, mas busca entender seu papel e impacto na sociedade como um todo, considerando suas múltiplas dimensões e implicações.

Profissionais desse campo podem desempenhar um papel essencial na compreensão e mitigação da desinformação, considerando uma sociedade imersa em recursos digitais de informação e comunicação. Combater a disseminação de conteúdos falsos, fraudulentos e distorcidos é crucial para preservar a integridade da informação e fortalecer a confiança pública nas instituições, nas organizações e nas pessoas que desempenham atividades pautadas na ética. Exemplos recentes, como os estudos científicos sobre materiais falsos compartilhados durante a pandemia de COVID-19 no Brasil, ilustram como pesquisadores têm analisado o impacto da desinformação na confiabilidade da ciência e na comunicação pública.

Apresentada pela computação, a inteligência artificial (IA) reflete avanços significativos em diversos setores, da medicina à indústria. Neste artigo, a IA é compreendida, com base nas ideias de Kaufman (2019), como uma área do conhecimento dedicada ao desenvolvimento de sistemas e máquinas capazes de realizar tarefas que geralmente exigem inteligência humana. Entre as abordagens, destaca-se o modelo de IA generativa, capaz de criar conteúdos convincentes como textos, imagens e vídeos, o que levanta questões éticas e práticas em diversas áreas do conhecimento, incluindo a ciência da informação. Como observado por Westerlund (2019), a capacidade da IA generativa em produzir *deepfakes* — conteúdos falsos realistas — amplia os desafios relacionados à verificação e autoria de informações no ambiente digital, provocando discussões tanto em nível técnico - relacionadas às próprias tecnologias, quanto sobre aplicações que envolvem direitos fundamentais, segurança e privacidade.

Por isso, nos últimos anos, a disseminação de desinformação tornou-se um desafio para todos os setores da sociedade, exigindo respostas eficazes que vão desde a

conscientização pública até a regulação de mídias, plataformas e conteúdo. A introdução comercial da inteligência artificial generativa nos últimos anos – notadamente após o lançamento do DALL-E[1], em 2021, e do ChatGPT[2], em 2022 – ampliou essa complexidade ao permitir não apenas a criação, mas também a disseminação rápida de conteúdos manipulados, ampliando os impactos da desinformação na sociedade.

O Brasil enfrenta desafios particulares na regulamentação da inteligência artificial, como a coordenação de *startups* que produzem essa tecnologia, desenvolvimento de ações educativas sobre o uso de IA e atenção à classificação de riscos associados aos serviços que usam esses sistemas – até porque um *chatbot*[3] de atendimento não possui os mesmos riscos de um sistema que produz vídeos baseados em IA. Iniciativas como o Projeto de Lei (PL) 2.338/2023[4] e a recente aprovação das regras do Tribunal Superior Eleitoral (TSE) sobre o uso de IA nas eleições de 2024 refletem os esforços iniciais para controlar o uso de IA no país.

A partir desses desafios apresentados, este trabalho busca fazer uma análise inicial e investigar os impactos da inteligência artificial (IA) generativa na integridade da informação e explora como as ferramentas, como o ChatGPT e o DALL-E, desafiam a confiabilidade dos conteúdos produzidos por essa tecnologia. Por meio de uma análise de casos que envolvem figuras públicas e personagens fictícios em redes sociais e das implicações sobre a autonomia de conteúdos gerados por IA, o estudo busca fornecer caminhos para que políticas públicas considerem os riscos da desinformação gerada pelos materiais criados pela IA generativa.

Este estudo é uma pesquisa qualitativa e exploratória, realizada com base em artigos científicos brasileiros sobre desinformação e IA, além de artigos em inglês sobre IA publicados nos últimos cinco anos em que se destacaram as análises sobre os desafios trazidos por sua acessibilidade no formato generativo. A referência clássica do trabalho de Turing (1950) também foi considerada por sua relevância histórica no campo. Dessa forma, em um primeiro momento, esta pesquisa se concentra nas questões filosóficas sobre a autonomia de conteúdos e reflexões para a pesquisa em ciência da informação. Em seguida, este trabalho também foca na análise de um estudo de caso envolvendo personagens gerados por IA em

[1] DALL-E é um modelo de inteligência artificial desenvolvido pelo OpenAI capaz de gerar imagens a partir de descrições textuais.

[2] ChatGPT é um modelo de inteligência artificial desenvolvido pela OpenAI que gera textos em linguagem natural, com base em interações com os usuários

[3] Tecnologia que simula conversas humanas em dispositivos digitais, como um robô conversador.

[4] Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233

redes sociais e explora os efeitos da desinformação gerada por IA na avaliação crítica dos usuários.

2 A IA GENERATIVA E SEUS EFEITOS PARA A CIÊNCIA DA INFORMAÇÃO

Nos subtópicos a seguir, é explorado como os conteúdos gerados pela inteligência artificial em sua forma generativa desafiam a autoria e a confiabilidade da informação. Para começar a problematização do tema, este artigo apresenta as reflexões de décadas atrás e também algumas contemporâneas de cientistas e filósofos da computação, como Alan Turing (1950) e Mark Coeckelbergh (2023). Além das questões filosóficas, esta seção também explora alguns desafios tanto sobre responsabilidade quanto sobre questões práticas da IA generativa na ciência da informação e sua influência na criação e disseminação de conteúdos, abrindo caminho para debates nas questões corporativas e no campo regulatório nos tópicos em sequência.

2.1 Autoria das máquinas: reflexão inicial para a ciência da informação

A inteligência artificial está cada vez mais presente no cotidiano das pessoas, desde os algoritmos que priorizam conteúdos específicos para os usuários nas redes sociais até a direção de veículos autônomos. Os sistemas generativos de IA – uma subcategoria dessa tecnologia – representam um avanço significativo na computação, sendo capazes de criar materiais inéditos com base nos dados que são treinados. Por outro lado, esse modelo desafia princípios fundamentais da produção de informação, especialmente no que diz respeito à confiabilidade das fontes reais e à autoria de conteúdo.

A reflexão iniciada por Alan Turing (1950) – pioneiro na área da inteligência artificial – em seu famoso artigo "*Computing Machinery and Intelligence*" (1950, p. 433) permanece em questão até os dias de hoje: "Podem as máquinas pensar?". Embora o artigo de Turing não aborde diretamente as questões éticas, sua provocação sugere não apenas uma avaliação dos limites de máquinas avançadas, mas também prevê debates sobre quem poderia, atualmente, responsabilizar-se pelos conteúdos gerados por sistemas de IA, como os materiais complexos e convincentes em textos, imagens, vídeos e áudio.

Mark Coeckelbergh (2023), filósofo da ética aplicada à IA, traz essa responsabilidade ao incorporar uma perspectiva ética onde a IA deve ser desenvolvida e utilizada para beneficiar a humanidade. Ao argumentar que essas ferramentas não apenas desafiam as pessoas a pensar sobre responsabilidades, Coeckelbergh sugere a reflexão sobre como a humanidade cria histórias e significados no mundo moderno. Para o filósofo, os humanos são os principais criadores de significado e contadores de histórias, "no entanto, se as direções mais pós-humanistas na hermenêutica da tecnologia estiverem corretas, isso sempre terá que ser feito em 'co-autoria' com a IA e outras tecnologias de nosso tempo" (Coeckelbergh, 2023).

Embora de diferentes épocas e campos de estudo, as visões de Turing (1950) e Coeckelbergh (2023) contribuem para uma visão holística da IA, que engloba tanto a capacidade técnica quanto a responsabilidade ética no seu uso e implementação. Nesse sentido, é possível levar essas questões filosóficas aprofundadas para implicações práticas no campo da ciência da informação para compreender como as tecnologias influenciam a criação e disseminação de informação.

Ao retomar o tópico sobre as principais preocupações da ciência da informação apresentadas na introdução deste artigo, pode-se dizer que os cientistas da área têm um papel importante na denúncia dos efeitos negativos da desinformação e na proposição de alternativas para enfrentá-la. Isso seria um "resgate de valores com os quais o campo historicamente se comprometeu e que se encontram ameaçados, como a democracia, a inclusão, a diversidade, a sustentabilidade e a promoção de uma cultura da paz" (Araújo, 2020). Ou seja, trata-se da importância em garantir a integridade e a autenticidade da informação em um mundo cada vez mais dominado por tecnologias que desafiam esses conceitos tradicionais.

No contexto mais atual, os efeitos da disseminação de textos e *deepfakes* gerados pela IA generativa de forma irresponsável intensificam os desafios enfrentados pela ciência da informação em garantir a integridade informacional e combater a desinformação. Para este artigo, entende-se *deepfakes* a partir da revisão de Westerlund (2019) sobre conteúdos em imagens, vídeos e áudios manipulados que miram as mídias sociais, onde boatos e conteúdos falsos se espalham facilmente e que serão abordados a seguir.

2.2 *Deepfakes*: efeitos de casos recentes no Brasil

O caso das meninas no Rio de Janeiro (Nascimento; Correia, 2023), vítimas de compartilhamento de falsos nudes criados por IA generativa em 2023, evidencia impactos preocupantes da tecnologia *deepfake* nos campos individual e social. A disseminação de conteúdos manipulados não apenas compromete a integridade pessoal das vítimas ao expô-las indevidamente, mas também mina a confiança pública na veracidade das informações, o que abre caminho para a necessidade de mitigar esses abusos.

O uso irresponsável da IA generativa também levanta questões de confiabilidade nas comunicações dos governos. Em 2023, um vídeo *deepfake* com a simulação do presidente Luiz Inácio Lula da Silva indicando um site suspeito com promessa de resgate de valores circulou amplamente nas redes sociais (Brasil, 2023). A distribuição do conteúdo também foi amplificada por *chatbots* e sites, demonstrando o poder dessa nova era da desinformação: a união de um ecossistema de plataformas online com a inteligência artificial na propagação de conteúdos falsos com o objetivo de causar danos — nesse caso, também financeiros – aos cidadãos.

As duas situações de uso de *deepfakes* mostram que a capacidade de criar desinformação de forma convincente coloca em risco a reputação e a segurança das pessoas. Em tempos de conflitos geopolíticos, como a guerra entre a Rússia e a Ucrânia iniciada em fevereiro de 2022, os *deepfakes* são desinformação transformada em arma, com o objetivo de interferir em processos eleitorais e provocar confrontos civis (Westerlund, 2019).

No Brasil, a falta de regulamentação adequada amplificaria esses riscos, permitindo que essas tecnologias sejam exploradas para manipulação e exploração de conteúdos com fins prejudiciais, incluindo difamação de pessoas, disseminação de desinformação e até influência em decisões políticas. O desenvolvimento e a aprovação de legislações podem ser garantias para a proteção dos indivíduos contra o uso indevido da IA generativa, já que promoveriam a responsabilidade dos desenvolvedores e dos usuários na criação e disseminação desses conteúdos.

Atualmente, os *deepfakes* não são especificamente abordados por uma lei de forma geral no Brasil, embora o país tenha aprovado a regulamentação do Tribunal Superior Eleitoral

(TSE) sobre o uso dessa tecnologia para as eleições de 2024 e o Projeto de Lei (PL) 2.338/2023[5] em andamento. Nesse sentido, sugere-se que a rápida evolução da IA generativa e a introdução comercial apressem a aprovação de um marco regulamentar.

3 CONTEÚDO CONVINCENTE E CAPACIDADE CRÍTICA

Os próximos subtópicos trazem duas análises de casos com o objetivo de aprofundar a discussão proposta neste artigo sobre os efeitos da IA generativa na desinformação e no desenvolvimento do pensamento crítico dos usuários. Primeiro, apresenta-se uma pesquisa sobre a percepção dos usuários na validação de *deepfakes* e de textos gerados por IA. Na sequência, o perfil do Instagram "Will Baiano"[6] é apresentado como exemplo prático de como a tecnologia *deepfake* pode ser utilizada para criar identidades convincentes.

3.1 A pesquisa de Zurique – a IA desinforma melhor

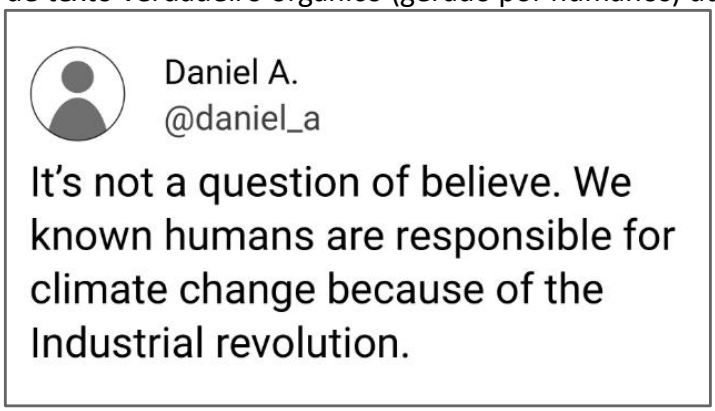
Um estudo de 2022 de cientistas da Universidade de Zurique[7], na Suíça, e publicado na revista *Science Advances*, revelou que a IA generativa tem capacidade de criar desinformação convincente. Para chegarem a essa conclusão, os pesquisadores solicitaram ao GPT-3, um modelo de linguagem generativa baseado desenvolvido pela OpenAI, que escrevesse textos para a rede social X (antigo Twitter*)* informativos e desinformativos com temas diversos sobre vacinas, COVID-19, teoria da evolução e outros assuntos comumente sujeitos à desinformação. Ao mesmo tempo, eles exploraram a rede social X em busca de informações verdadeiras e desinformação sobre os mesmos tópicos. Os participantes conseguiram identificar os conteúdos falsos orgânicos (textos gerados por humanos) com maior facilidade do que os textos falsos sintéticos (criados pelo GPT). São resultados que sugerem um desafio crescente na detecção de desinformação gerada por sistemas inteligentes por parte dos usuários (Spitale; Biller-Andorno; German, 2023).

Em outro momento da pesquisa, os resultados também revelaram que a capacidade dos participantes de distinguir entre textos verdadeiros, tanto orgânicos (Figura 1) quanto

[5] Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233
[6] Disponível em: https://www.instagram.com/willbaiano/
[7] Disponível em: *https://www.science.org/doi/10.1126/sciadv.adh1850*

sintéticos (Figura 2), foi muito semelhante. Ou seja, houve dificuldade em determinar a distinção entre informações reais, independentemente de sua origem.

Figura 1 - Exemplo de texto verdadeiro orgânico (gerado por humanos) utilizado na pesquisa



Fonte: Extraído do repositório do estudo de Spitale, Biller-Andorno e German (2023)

Como ilustrado na Figura 1, o texto criado por humanos não se distingue pela forma de escrita do conteúdo gerado por IA, como ilustrado abaixo, na Figura 2. Assim, é essencial aprimorar as técnicas de avaliação de informações e fortalecer a capacidade crítica dos usuários para enfrentar a proliferação de conteúdos gerados por IA.

Figura 2 - Exemplo de texto verdadeiro sintético (gerado pelo GPT) utilizado na pesquisa

Fonte: Extraído do repositório do estudo de Spitale, Biller-Andorno e German (2023)

Além da necessidade de regulamentações para mitigar a desinformação gerada por IAs generativas, a pesquisa de Zurique também sugere ações para o desenvolvimento de campanhas educativas entre usuários de internet e a importância do tema para as pesquisas em informação. Essa proposta é interessante para a ciência da informação, até porque, como defende Araújo (2020), o desafio central para os cientistas desse campo é promover uma cultura de busca pela verdade. Dessa forma, a ciência da informação é desafiada a

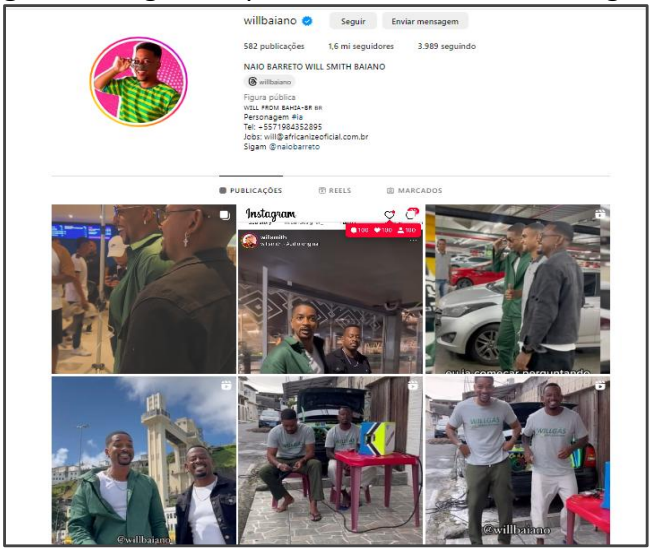
acompanhar e responder às evoluções rápidas na tecnologia digital, garantindo que a sociedade possa navegar de maneira informada e responsável no ambiente online.

3.2 O caso “Will baiano”

A introdução comercial da inteligência artificial em sua forma generativa ampliou a acessibilidade das pessoas a essa tecnologia e assumiu uma nova dimensão para a desinformação: os usuários não apenas consomem os conteúdos gerados, mas também se tornam produtores ativos. Isso não apenas facilita a disseminação dos materiais como também aumenta o alcance e a influência das narrativas que podem não ser verificadas.

No Brasil, um exemplo do impacto da acessibilidade facilitada à IA generativa é o “Will baiano”, um perfil fictício no Instagram criado pelo influenciador Naio Barreto (Figura 3). Em 2024, esse personagem ganhou notoriedade ao ser apresentado como o suposto gêmeo brasileiro do ator americano Will Smith por meio do uso de IA para reproduzir traços faciais semelhantes aos da celebridade real. Naio conseguiu criar uma transformação visual extremamente convincente em vídeos que se tornaram virais[8] e acumular mais de 580 mil seguidores na rede social. Na plataforma, comentários como “Gente é ou não é inteligência artificial? Tô confusa”, da usuária @rebecawalessabeltrao e “Ele parece mais o Will do que o próprio Will”, publicado pela usuária @vitorialudugerios2_ refletem a impactante reação do público diante da aplicação de IA generativa.

Figura 3 - Página do perfil “Will baiano” no Instagram



Fonte: Reprodução do Instagram/Perfil “Will baiano” (2024)

[8] Vídeo com mais de 1 milhão de visualizações produzido por “Will baiano” e republicado pelo ator Will Smith em junho de 2024: https://www.instagram.com/reel/C74vZXQpWsP/

Muitos *deepfakes* são inspirados em celebridades e políticos porque a internet está repleta de fotos e vídeos dessas personalidades, a partir dos quais é possível construir materiais para treinar a IA generativa para entretenimento. Por outro lado, exemplos de *deepfakes* mal intencionados também estão surgindo cada vez mais, já que usuários podem criar falsos discursos políticos, além do fato de pessoas comuns também estarem submetidas a conteúdos manipulados a partir de fotos pessoais nas redes sociais (Westerlund, 2019).

O caso do "Will baiano" exemplifica a propagação de *deepfakes* gerados por IA nas redes sociais de forma convincente e voltada para o entretenimento, mas levanta preocupações sobre a percepção e crença pública em informações. A elevada qualidade das transformações visuais proporcionadas pela IA cria um cenário onde a distinção entre o real e o falso se torna cada vez mais confusa, facilitando potenciais mal-entendidos e a disseminação de desinformação, mesmo que o autor evidencie a informação que se trata de um personagem fictício criado por IA, como no caso do "Will baiano".

A sociedade digital convive com as *fake news* – informações deliberadamente fabricadas e falsas em formato jornalístico – e a desinformação de uma forma ainda mais arriscada pela forma que a IA generativa facilita a criação de conteúdo enganoso e pela rapidez com que a informação se espalha por dispositivos móveis. Desenvolver um olhar crítico para esses materiais, ou, em melhores palavras, desenvolver competência em informação, pode ser considerada uma ferramenta essencial na luta contra esses fenômenos, já que "tal competência prepara as pessoas para analisar criticamente as informações e permite-lhes usá-las para produzir novos conhecimentos de forma criativa e contextualizada" (Brisola; Bezerra, 2018). Essa habilidade também contribui no questionamento de como as plataformas digitais priorizam o lucro em detrimento da verdade e da ética, possibilitando o entendimento de um ecossistema informacional complexo e direcionado aos moldes capitalistas.

Para o debate sobre a disseminação da desinformação, vale acrescentar que esse fenômeno pode ser potencializado pela redução de custos de produção e pela conexão de outros dispositivos – como as ferramentas de automatização nas redes sociais. Ou seja, comunicações de massa que antes exigiam equipes inteiras de especialistas, como programadores, redatores, designers e editores de vídeo, agora podem ser realizadas por uma única pessoa ou pequenos grupos de profissionais por meio dessas tecnologias.

Com a queda de custos de produção e agilidade na disseminação, o volume de conteúdo enganoso online mostra ter potencialidade para, rapidamente, causar danos para a

sociedade. Essa realidade desafia a verificação de fatos, pois a quantidade de conteúdo a ser verificado pode superar a capacidade humana de combater a desinformação de forma eficaz.

Westerlund (2019) traz a questão da disseminação de *deepfakes* como uma ameaça para a confiança no jornalismo e nos governos já que, ao contrário das *fake news* convencionais, que podem ser detectadas e desmascaradas com investigação, os *deepfakes* se assemelham aos humanos de forma extremamente convincente. Ao mesmo tempo, a pesquisadora também argumenta que as pessoas podem até considerar imagens verdadeiras como falsas simplesmente porque elas estão em contato constante com a desinformação e concluem que tudo está distorcido – o que também foi apontado no estudo de Zurique.

A partir dessas análises, é colocada em questão a urgente necessidade de mais aprimoramento das ferramentas de detecção de desinformação e de iniciativas de educação do público — tanto por governos quanto por empresas privadas – sobre os riscos dos textos e *deepfakes* gerados por IA generativa. Para a desinformação, além das questões regulatórias, promover um pensamento crítico é essencial para que a sociedade questione a responsabilidade desses conteúdos, pois até pessoas comuns podem estar sujeitas à manipulação ao disponibilizarem suas fotos pessoais nas redes sociais. O uso estratégico de tecnologias disponíveis, aliado a um plano de ação bem elaborado com fins a gerar desinformação, pode manipular a opinião pública de maneira eficaz, causando danos aos processos sociais, tanto de maneira individual como coletiva. Devido à desconfiança generalizada e à facilidade de acesso a ferramentas tecnológicas poderosas, é mais fácil influenciar e moldar as percepções e opiniões das pessoas, o que pode ser perigoso para a integridade da informação e da democracia (Boarini; Ferrari, 2021).

4 IA NO CAMPO REGULATÓRIO

Os casos exemplificados neste artigo sobre o uso de *deepfakes* e os resultados da pesquisa de Zurique sobre a confiabilidade em textos gerados por humanos e GPT sugerem a criação de um nível de supervisão para atribuir responsabilidades de maneira clara. Para monitorar o desempenho dos sistemas, corrigir danos e atribuir responsabilidades é essencial enfrentar os efeitos negativos da IA desde o início da sua implementação e educar os usuários sobre os princípios e funcionamento da tecnologia (Kaufman, 2021).

A ideia de regular a inteligência artificial no Brasil não é nova, com iniciativas como a regulamentação do TSE e o Projeto de Lei (PL) 2.338/2023[9] em andamento. No entanto, diversos desafios precisam ser superados, como a coordenação entre diferentes setores da tecnologia e governo, as questões éticas dos conteúdos produzidos por sistemas de IA e, principalmente, garantir que a regulamentação acompanhe o ritmo acelerado da IA generativa e suas novas implicações no contexto da desinformação.

Geralmente operadas por empresas privadas, as redes sociais, as plataformas de mensagem e os serviços que utilizam IA generativa atuam na formação da opinião pública, influenciando não apenas o consumo de bens e serviços, mas servindo como palco de discussões além das interações privadas (Archegas; Estarque, 2021). Por serem atividades que impactam o interesse público, como os debates sobre saúde pública e compartilhamento de dados pessoais, a regulação da IA pelos governos torna-se essencial.

Para a Organização das Nações Unidas (ONU), os países devem moldar o desenvolvimento da IA e, por isso, em março de 2024, a Assembleia Geral da ONU adotou a primeira resolução global sobre a tecnologia. Em relação à legislação, a União Europeia (UE) está liderando o caminho da regulamentação, cujo princípio é estabelecer, principalmente, que a IA tenha classificação de riscos com a supervisão de pessoas, em vez de ser automatizada, para evitar resultados prejudiciais, como a falta de transparência e discriminação (Conselho da União Europeia, 2024).

No Brasil, ainda há um longo caminho de aprendizados a percorrer para acompanhar os avanços da lei europeia e equilibrar essa trajetória com as peculiaridades do país, como a diversidade cultural, o ecossistema de pequenas *startups* e o acesso à educação e à tecnologia. Por exemplo, um olhar para o incentivo à educação tecnológica e ao estímulo ao pensamento crítico em relação à desinformação gerada por IAs generativas ou ainda prover suporte às *startups* do setor para garantir que a IA seja uma ferramenta de crescimento, além de lançar luz aos processos de treinamento dessas inteligências para evitar produções discriminatórias.

Em relação às pesquisas em informação que focam o potencial de manipulação de materiais falsos, elas não só evidenciam os riscos associados à desinformação, mas também reforçam a necessidade dessa abordagem colaborativa entre academia, governo e grandes

[9] Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233

empresas de tecnologia no campo regulatório. Um caso que reforça essa proatividade no combate à desinformação é o trabalho do Netlab, laboratório de pesquisa da Universidade Federal do Rio de Janeiro (UFRJ), ao expor a negligência na moderação de anúncios na plataforma Meta[10] que imitavam programas sociais e circulavam pelas redes sociais em 2023 (Coutinho, 2024). A resistência demonstrada pelas grandes empresas em assumir responsabilidades – como aconteceu no caso do Netlab e da Meta – evidencia a necessidade urgente de proteção aos direitos individuais acima dos interesses lucrativos das corporações. Como destacado por Brizola e Bezerra (2018), esse cenário propicia o surgimento dos conteúdos falsos e manipulados para fins prejudiciais à população, já que as plataformas reprodutoras desses materiais não se preocupam com a credibilidade a longo prazo, mas sim com lucros imediatos.

5 CONSIDERAÇÕES

As notícias relacionadas ao uso de *deepfakes* e os estudos apresentados neste artigo evidenciam que os materiais gerados pela IA generativa apresentam desafios significativos para a ciência da informação e os riscos da desinformação para as pessoas, a sociedade e a democracia. Enquanto este estudo trouxe os problemas para a integridade individual e nos direitos humanos causados pela disseminação de materiais manipulados por IA, foi apresentado como esses avanços tecnológicos desafiam conceitos tradicionais da ciência da informação e levantam questões éticas sobre a responsabilidade e regulamentação necessárias para mitigar esses riscos.

As questões filosóficas de Turing sobre “máquinas pensantes”, as implicações éticas de Coeckelbergh (2023) e a as ideias de Araújo (2020) sobre as preocupações em ciência da informação trazem reflexões práticas para a pesquisa deste campo, como a necessidade de desenvolver estratégias educacionais que promovam o pensamento crítico, além da urgência de uma legislação brasileira que considere os riscos no uso de IA no campo da desinformação. Essa é uma questão inicial proposta neste artigo como caminhos essenciais à resistência à manipulação digital na era da informação.

[10] A Meta é controladora das plataformas Facebook, Instagram e WhatsApp e é considerada uma das principais empresas de tecnologia globais.

Para abordar as implicações futuras no processo eleitoral – o evento mais importante para a democracia no Brasil – é importante considerar os riscos da IA generativa mostrados pelos exemplos de notícias com *deepfake* neste artigo. A regulamentação aprovada com urgência pelo TSE para controlar o uso de IA nas eleições 2024 é um avanço, mas não suficiente, pois, como sugerido pelas análises das notícias, da pesquisa de Zurique e do caso "Will baiano" neste artigo, é necessária uma abordagem abrangente que inclua a cooperação entre governos e empresas, a conscientização pública e a responsabilização das plataformas tecnológicas. Isso porque, antes, a criação de vídeos e textos manipulados exigia habilidades especializadas (designers, editores de vídeo e comunicadores), mas agora, com modelos de IA generativa como o ChatGPT, essa capacidade está acessível a pessoas sem conhecimentos técnicos avançados, aumentando o risco de uso indevido da tecnologia (Kaufman; Santaella, 2024).

Neste estudo, também se destaca a necessidade de proteger a integridade da informação publicizada por órgãos públicos e garantir a segurança das pessoas diante do potencial abuso dessa tecnologia, como foi mostrado nas notícias exemplificadas com o uso de *deepfakes* para propagar conteúdo falso supostamente divulgado pela presidência da república e dos falsos nudes de meninas em uma escola no Rio de Janeiro.

Para pesquisas futuras, sugere-se uma investigação aprofundada de mecanismos de disseminação de desinformação por meio da IA generativa em diferentes contextos culturais e sociais. Além disso, indica-se a exploração de métodos avançados de detecção e verificação de conteúdos gerados por IA, bem como a aprimoração de modelos de regulamentação que se adaptem à rápida evolução dessas tecnologias no Brasil.

O aprofundamento neste tema tem por objetivo contribuir para o entendimento crítico dos desafios emergentes na interseção entre inteligência artificial generativa, desinformação e responsabilidade. As conclusões iniciais apontam para a necessidade de ações coordenadas e proativas para mitigar seus impactos negativos enquanto se promove um desenvolvimento tecnológico responsável e inclusivo.

REFERÊNCIAS

ARAÚJO, C. A. A. A pós-verdade como desafio central para a ciência da informação contemporânea. **Em Questão**, Porto Alegre, v. 27, n. 1, p. 13–29, 2020. Disponível em: https://seer.ufrgs.br/index.php/EmQuestao/article/view/101666. Acesso em: 28 mai. 2024.

ARCHEGAS, J. V.; ESTARQUE, M. Redes Sociais e Moderação de Conteúdo: Criando regras para o debate público a partir da esfera privada. **Relatório do Instituto de Tecnologia e Sociedade**. Rio de Janeiro, 2021. Disponível em: https://itsrio.org/publicacoes/redes-sociais-e-moderacao-de-conteudo. Acesso em: 20 mai. 2024.

BOARINI, M.; FERRARI, P. A desinformação é o parasita do século XXI. **Organicom**, São Paulo, Brasil, v. 17, n. 34, p. 37–47, 2021. Disponível em: https://www.revistas.usp.br/organicom/article/view/17054. Acesso em: 24 jun. 2024.

BRISOLA, A.; BEZERRA, A. C. Desinformação e circulação de "fake news": distinções, diagnóstico e reação. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 19., 22-26 out. 2018, Londrina. **Anais** [...]. Londrina: UEL, 2018. Disponível em: http://hdl.handle.net/20.500.11959/brapci/102819. Acesso em: 27 mai. 2024.

COECKELBERGH, M. Narrative responsibility and artificial intelligence. **AI & Society**, v. 38, p. 2437–2450, 2023. Disponível em: https://doi.org/10.1007/s00146-021-01375-x. Acesso em: 21 maio. 2024.

COUTINHO, S. Artigo de reitor da UFRJ repudia censura à ciência. **Universidade Federal do Rio de Janeiro**, Rio de Janeiro, 12 jun. 2024. Disponível em: https://ufrj.br/2024/06/artigo-de-reitor-da-ufrj-repudia-censura-a-ciencia/. Acesso em: 24 jun. 2024.

BRASIL. Secretaria de Comunicação Social. Estelionatários usam imagem do presidente da República em *deep fake*. [Brasília]: Secretaria de Comunicação Social, 11 nov. 2023. Disponível em: https://www.gov.br/secom/pt-br/fatos/brasil-contra-fake/noticias/2023/12/estelionatarios-usam-imagem-do-presidente-da-republica-em-deep-fake. Acesso em: 24 jun. 2024.

KAUFMAN, D. **A inteligência artificial irá suplantar a inteligência humana?** São Paulo: Estação das Letras e Cores, 2019.

KAUFMAN, D. Inteligência artificial e os desafios éticos: a restrita aplicabilidade dos princípios gerais para nortear o ecossistema de IA. Paulus – **Revista de Comunicação da FAPCOM**, São Paulo, v. 5. N. 9, 2021. Disponível em: https://fapcom.edu.br/revista/index.php/revista-paulus/article/view/453/427 . Acesso em: 24 jun. 2024.

NASCIMENTO, R.; CORREIA, B. H. Alunos de colégio na Barra são suspeitos de usar inteligência artificial para fazer montagens de colegas nuas e compartilhar. **G1**, Rio de Janeiro, 01 nov. 2023. Disponível em: https://g1.globo.com/rj/rio-de-janeiro/noticia/2023/11/01/alunos-de-colegio-na-barra-sao-suspeitos-de-usar-inteligencia-

artificial-para-fazer-montagens-de-colegas-nuas-e-compartilhar.ghtml. Acesso em: 24 jun. 2024.

Regulamento Inteligência Artificial (IA): Conselho dá luz verde final às primeiras regras do mundo em matéria de IA. **Conselho da União Europeia**, Bruxelas, 21 mai. 2024. Disponível em: https://www.consilium.europa.eu/pt/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/. Acesso em: 24 jun. 2024.

SANTAELLA, L.; KAUFMAN, D. A Inteligência artificial generativa como quarta ferida narcísica do humano. **MATRIZes**, São Paulo, Brasil, v. 18, n. 1, p. 37–53, 2024. DOI: 10.11606/issn.1982-8160.v18i1p37-53. Disponível em: https://www.revistas.usp.br/matrizes/article/view/210834. Acesso em: 24 jun. 2024.

SPITALE, G.; BILLER-ANDORNO, N.; GERMANI, F. AI model GPT-3 (dis)informs us better than humans. **Science Advances**, v. 9, n. 26, 2023. Disponível em: https://www.science.org/doi/10.1126/sciadv.adh1850. Acesso em: 20 maio. 2024.

TURING, A. Computing machinery and intelligence. **Mind**, v. 59, n. 236, p. 433–460, 1950. Disponível em: https://doi.org/10.1093/mind/LIX.236.433. Acesso em: 20 maio 2024.

WESTERLUND, M. The emergence of deepfake technology: a review. **Technology Innovation Management Review**, v. 9, n. 11, p. 40–53, 2019. Disponível em: http://doi.org/10.22215/timreview/1282. Acesso em: 20 maio 2024.

WILL BAIANO. **Perfil do Instagram**. Salvador, 24 junho 2024. Instagram: @willbaiano. Disponível em: https://www.instagram.com/willbaiano. Acesso em: 24 jun. 2024.