

Um Estudo Exploratório sobre os Perigos à Democracia com o Uso da Inteligência Artificial

Domingos S. Ferreira de Melo¹, Carla I. M. Bezerra², Leonardo O. Moreira¹,
Emanuel F. Coutinho^{1,2}

¹Programa de Pós-Graduação em Tecnologia Educacional (PPGTE)
Universidade Federal do Ceará (UFC) – Fortaleza, CE – Brasil

²Universidade Federal do Ceará (UFC) – Campus de Quixadá, Quixadá, CE – Brasil

domingosmelo@alu.ufc.br, carlailane@ufc.br,
{leomoreira, emanuel.coutinho}@ufc.br

Abstract. *Artificial Intelligence (AI) is gaining relevance and impact in democracy, being applied in various areas of the democratic process, such as campaigns, elections, and political decision-making. While AI has potential benefits for democracy, it is also important to be aware of the dangers associated with its use. Furthermore, it is essential to understand not only the dangers but also the secondary impacts on democracy in the face of AI utilization, in order to mitigate risks and protect fundamental democratic values. The main objective of this work is to present a discussion on the perils of using AI in democracy through a literature review.*

Resumo. *A Inteligência Artificial (IA) está ganhando relevância e impacto na democracia, sendo aplicada em várias áreas do processo democrático, como campanhas, eleições e tomadas de decisões políticas. Embora a IA possua benefícios potenciais para a democracia, também é importante estar ciente dos perigos associados ao seu uso. Além disso, é fundamental conhecer não só os perigos, mas também os impactos secundários à democracia diante do uso da IA, para que seja possível mitigar os riscos e proteger os valores fundamentais democráticos. Este trabalho tem como objetivo principal apresentar uma discussão sobre os perigos do uso da IA na democracia por meio de uma revisão da literatura.*

1. Introdução

A democracia pode ser entendida como a participação e deliberação coletiva sobre a fixação de prioridades comuns da sociedade civil [Freitas e Teixeira 2019]. As redes sociais, principais meios preponderantes de comunicação e de articulação política, têm se transformado em um ambiente caracterizado pela polarização e pelo extremismo, onde a livre troca de informações é frequentemente prejudicada por interferências que, constantemente, dificultam a ocorrência de diálogos e compreensões efetivas [Carvalho 2020]. A Inteligência Artificial (IA) é uma área de pesquisa em computação que tem como objetivo proporcionar às máquinas comportamentos semelhantes aos humanos para a solução de problemas, raciocínio e aprendizado [Wilner 2018]. Atualmente, a IA está presente em

diversos cenários do nosso cotidiano, como entretenimento, saúde, finanças, educação, segurança, cultura, relações pessoais, tecnologia etc.

No âmbito da política, impactando positivamente na democracia, a IA tem sido utilizada como ferramenta para garantir transparência, detecção de desinformação, acesso a informação de forma igualitária, incentivo à participação cívica etc. Infelizmente, a IA também tem servido como ferramenta para impactar negativamente na democracia. Diversos trabalhos relatam e discutem os impactos negativos à democracia diante do uso da IA como estratégia de produção e divulgação de informações nos meios digitais de comunicação e interação [Wilner 2018, Carvalho 2020, Rodrigues et al. 2020, König e Wenzelburger 2020, Bernardi e Costa 2020, Dias et al. 2021, Valadares 2021]. Este trabalho tem como objetivo principal apresentar uma discussão sobre os perigos do uso da IA na democracia por meio de uma revisão da literatura.

A elaboração desta pesquisa tem natureza exploratória quanto ao seu objetivo. Pesquisas exploratórias visa o aprimoramento de ideias ou a descoberta de intuições, onde na maioria dos casos essas pesquisas envolvem: a) levantamento bibliográfico; b) entrevistas com sujeitos que tiveram experiências práticas com o problema pesquisado; e c) análise de exemplos que estimulem a compreensão [Gil 2002]. Em particular, esta pesquisa se serve, principalmente, do instrumento de levantamento bibliográfico por meio de artigos científicos publicados em periódicos e conferências para enriquecimento das definições e discussões.

É válido ressaltar que este trabalho se propõe a realizar um estudo exploratório por meio de uma revisão da literatura, e não uma revisão sistemática da literatura. A principal distinção entre a revisão da literatura e a revisão sistemática da literatura reside na abordagem metodológica e no âmbito. Enquanto a revisão da literatura proporciona uma perspectiva mais abrangente sobre um determinado tópico, a revisão sistemática é caracterizada por sua estruturação, objetividade e foco em responder de maneira sistemática e transparente a uma pergunta de pesquisa específica. A seleção entre esses dois métodos está condicionada aos objetivos e às exigências particulares do estudo em questão [Brizola e Fantin 2017].

2. Revisão da Literatura

Esta seção tem como objetivo fornecer o contexto teórico relacionado a pesquisas realizadas sobre o IA e suas implicações na democracia, destacando as ameaças. É essencial esclarecer os pressupostos teóricos que sustentam a pesquisa e destacar as contribuições proporcionadas por investigações anteriores. Essa revisão não deve se limitar a referências ou resumos dos estudos existentes, mas deve incluir uma discussão crítica sobre o estado atual da questão [Gil 2002].

Quanto à etapa de planejamento e seleção de fontes foram adotadas as bases de dados acadêmicas: IEEE Xplore, ACM Digital Library, Scopus, Web of Science e Google Acadêmico. Foram selecionados trabalhos nacionais e internacionais publicados em periódicos e conferências durante o período de 2018 a 2023. Trabalhos que não discutiam assuntos relacionados à IA ou à democracia no âmbito digital foram excluídos da revisão. Por fim, como critério de qualidade dos trabalhos e aderência ao objetivo central desta revisão, foram excluídos trabalhos que não apresentavam, de alguma forma, ameaças à democracia por meio do uso da IA.

2.1. Perigos da Inteligência Artificial na Democracia

Bolhas de filtro e polarização. A crítica fundamental às bolhas de filtro está relacionada à homogeneização das relações sociais, pois elas tendem a manter os indivíduos em círculos sociais fechados e a perpetuar a igualdade de pensamentos [Kaufman e Santaella 2020]. As capacidades do aprendizado de máquina em processar informações sobre as preferências e comportamentos individuais permitem obter mensagens personalizadas e oferecer informações que se alinham de forma ideal com as predisposições de cada pessoa. Dessa forma, os cidadãos podem ficar cada vez mais confinados dentro dos limites do que é conhecido como “bolhas de filtro” [König e Wenzelburger 2020]. A Figura 1 ilustra um cenário no qual o utilizador procura e seleciona ativamente informações de apoio (*Pull*). Posteriormente, o utilizador recebe passivamente informações de apoio de serviços da Internet que anteciparam as suas preferências futuras (*Push*) [Dahlgren 2021].

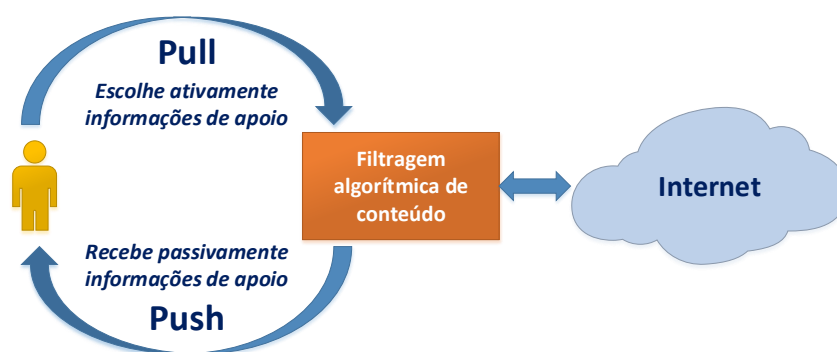


Figura 1. Mais do mesmo: A espiral de auto-reforço de como uma bolha de filtro surge. Fonte: [Dahlgren 2021].

Desinformação e fake news. A influência da IA, quando utilizada como meio de propagar desinformação nas redes sociais, tem um impacto significativo no exercício da cidadania, ao mesmo tempo em que influencia os processos de tomada de decisão dos cidadãos [Valadares 2021]. Além disso, a IA realiza uma classificação e filtragem algorítmica com o objetivo principal de capturar a atenção do público e envolver os usuários, em vez de transmitir uma imagem precisa e informativa da política [König e Wenzelburger 2020]. Diversos trabalhos investigam o uso de *fake news* nas campanhas/governos [Bernardi e Costa 2020]. Já outros trabalhos apresentam o impacto da desinformação em democracias, como no caso do escândalo Facebook-Cambridge Analytica [Rodrigues et al. 2020].

Manipulação e interferência eleitoral. A IA pode criar e disseminar desinformação em larga escala para influenciar a opinião pública e o resultado das eleições [Rodrigues et al. 2020]. Algoritmos de IA analisam o comportamento *online* e adaptam mensagens políticas enganosas (*fake news*) para atingir objetivos específicos. Além disso, plataformas de mídia social e mecanismos de busca utilizam algoritmos de recomendação suscetíveis a manipulações por meio da IA, favorecendo candidatos e intensificando a polarização. *Bots* e perfis falsos são usados para disseminar informações falsas, distorcer debates políticos e amplificar opiniões específicas [Dias et al. 2021], interagindo com usuários e criando uma aparência de apoio ou rejeição a candidatos e questões políticas.

A Figura 2 ilustra um fluxo básico de criação e disseminação de *fake news* em redes sociais, explicando como ocorrem tanto os perigos relacionados à **desinformação e fake news**, quanto à **manipulação e interferência eleitoral**. Por meio da Figura 2 (seta 1), é possível observar uma interação entre o usuário propagador de *fake news* e os *bots* de *fake news*. Essa interação pode envolver o desenvolvimento de *bots*, a implementação de novos algoritmos de IA, ou até mesmo a emissão de comandos para que os *bots* executem ações desejadas pelos usuários propagadores de *fake news*. Os *bots* de *fake news* são programas de computador que realizam ações, tais como compartilhar, curtir, retuitar, comentar e seguir usuários em redes sociais.

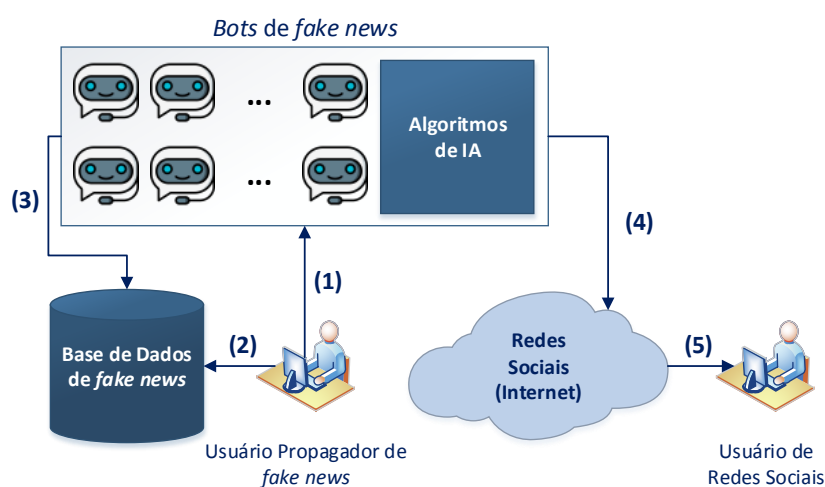


Figura 2. Fluxo básico de criação e disseminação de fake news em redes sociais.
Fonte: Elaborado pelos autores.

A seta 2 da Figura 2 descreve uma interação na qual o usuário propagador de *fake news* alimenta um banco de dados de *fake news* que, no futuro, será utilizado pelos *bots* de *fake news* com o propósito de uma disseminação inteligente em massa de tais *fake news*, fazendo uso dos algoritmos de IA. A seta 3 da Figura 2 representa o processo de acesso aos dados armazenados na base de dados de *fake news*, que são subsequentemente processados pelos *bots* de *fake news*. Ao obter acesso às *fake news*, os *bots* têm a capacidade de compartilhá-las automaticamente, criando a impressão de que várias pessoas diferentes estão apoiando a mesma informação.

A automação permite uma rápida propagação das *fake news* nas redes sociais (seta 4 da Figura 2), atingindo um público considerável. Essa rápida propagação, por meio dos algoritmos de IA, pode incluir o compartilhamento em grupos de discussão, páginas de fãs, perfis pessoais e até mesmo comentários em publicações de outros utilizadores. Em alguns casos, a disseminação é tão intensa que as *fake news* podem se tornar tendência nas redes sociais. Finalmente, a seta 5 da Figura 2 representa o objetivo final, que consiste em influenciar a opinião dos utilizadores das redes sociais, disseminar desconfiança, criar divisões, afetar processos eleitorais, prejudicar reputações e incitar conflitos.

Discriminação algorítmica. Algoritmos de IA usados em processos políticos e elei-

torais podem apresentar viés discriminatório, resultando em desigualdades e injustiças sistemáticas. A discriminação algorítmica no contexto eleitoral pode afetar o acesso igualitário à informação, o direito ao voto e a representatividade política. Além disso, algoritmos de recomendação podem amplificar visões políticas extremas, aprofundando a polarização e prejudicando o debate saudável. Algoritmos treinados em dados históricos enviesados podem perpetuar desigualdades, favorecendo certos candidatos ou excluindo grupos de eleitores. Essa discriminação algorítmica mina a confiança na democracia, causa exclusão política e intensifica desequilíbrios de poder.

A Figura 3 apresenta um cenário e um conjunto de etapas que podem resultar em uma situação de discriminação algorítmica. Inicialmente, são coletados dados de diversas fontes (Fontes de Dados) relacionadas a um determinado contexto ou domínio de interesse. Esses dados desempenham um papel fundamental na aquisição de conhecimento, que será posteriormente utilizado para o aprendizado e a geração de recomendações. É relevante destacar que esses dados podem abranger informações de milhares de usuários e suas interações em redes sociais, por exemplo.

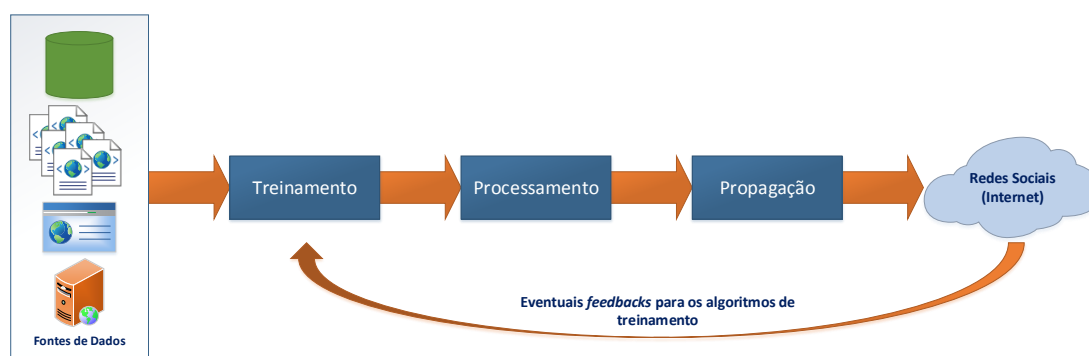


Figura 3. Fluxo básico de ocorrência da discriminação algorítmica. Fonte: Elaborado pelos autores.

Na etapa de treinamento, os dados adquiridos na fase inicial (Fonte de Dados) são utilizados para treinar algoritmos, os quais podem conter vieses históricos, políticos e sociais. Nesse contexto, se parte dos dados contiver traços de discriminação e preconceito, os algoritmos podem aprender e reproduzir esses vieses. A etapa de treinamento em algoritmos de IA é fundamental para instruir o modelo sobre como realizar tarefas específicas com base nas fontes de dados. Com base no aprendizado obtido a partir dos dados de treinamento, emerge a etapa de processamento, na qual os algoritmos fazem recomendações, classificam conteúdo ou tomam decisões, como por exemplo, recomendar conteúdo de mídia ou candidatos políticos.

Se os algoritmos foram treinados com dados enviesados, eles podem gerar recomendações que refletem ou ampliam esse viés (Processamento). Por exemplo, um sistema de recomendação de vídeos pode automaticamente sugerir vídeos com conteúdo discriminatório para um público suscetível. Por fim, quando os usuários interagem com as recomendações por meio das redes sociais (Propagação), se os algoritmos tiverem acesso aos conteúdos que foram acessados pelos usuários, então os algoritmos podem receber *feedbacks*. Se os usuários acessarem conteúdos enviesados, os algoritmos podem interpretar isso como uma preferência e continuar a oferecer conteúdo semelhante. Dessa forma, esse

feedback em *loop* (repetitivamente) pode amplificar o viés ao longo do tempo, criando um ciclo que reforça preconceitos e discriminação.

Privacidade e vigilância. A IA coleta e analisa grandes volumes de dados pessoais e comportamentais, invadindo a privacidade das pessoas ao extrair informações sensíveis sem consentimento. Além disso, possibilita o monitoramento em tempo real das atividades individuais, como comunicações, interações nas redes sociais e localização geográfica, comprometendo a sensação de liberdade e privacidade [Kaufman e Santaella 2020]. A criação de perfis detalhados com base em dados pessoais e atividades *online* é outra aplicação da IA, sendo utilizada para manipulação comportamental por meio de anúncios políticos personalizados, visando influenciar opiniões e decisões dos cidadãos [König e Wenzelburger 2020].

A Figura 4 ilustra um processo que descreve as etapas de privacidade e vigilância em redes sociais. Na primeira etapa, ocorre a coleta de dados pessoais dos usuários, incluindo informações como idade, gênero, localização geográfica, histórico de postagens, curtidas, compartilhamentos, comentários e conexões. Além disso, as redes sociais monitoram as atividades dos usuários, como visualização de postagens, reações a conteúdo e interações com outros usuários. Na etapa de análise de dados, os dados coletados são submetidos à análise por meio de algoritmos de IA, com o intuito de criar perfis minuciosos dos usuários. Essa análise abarca a elaboração de perfis que englobam interesses, preferências políticas, comportamentais, demográficos, entre outros aspectos relevantes.

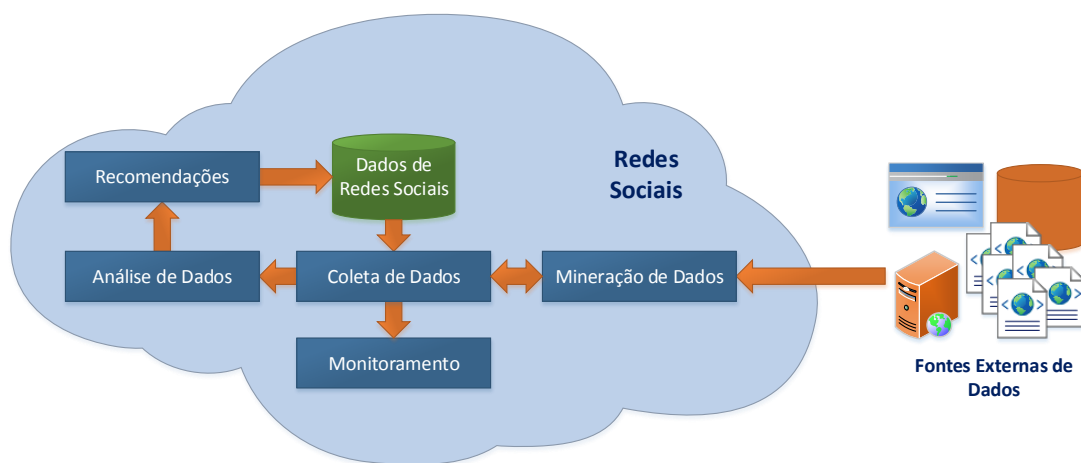


Figura 4. Fluxo básico do funcionamento das ações de privacidade e vigilância em redes sociais. Fonte: Elaborado pelos autores.

Com base nos perfis criados, na etapa de recomendações, é possível direcionar conteúdo altamente segmentado aos usuários, recomendando conteúdos que estejam alinhados com seus interesses e características individuais. A etapa de monitoramento, em determinadas situações, monitora as atividades dos usuários não apenas com fins de recomendações, mas também para fins de vigilância, incluindo a observação de comunicações privadas, como mensagens diretas, e a análise de comportamentos suspeitos. Por fim, a etapa de mineração de dados pode ser utilizada pelas redes sociais para adquirir informações adicionais, tais como dados demográficos, informações públicas de

perfil e postagens públicas.

2.2. Resultados e Discussão

Aqui são discutidas estratégias para combater os perigos mencionados na subseção anterior. Uma forma de combater as bolhas de filtro, polarização, desinformação e *fake news* é promover a diversidade de fontes de informação e opiniões nas plataformas de mídia social e mecanismos de busca. Certamente, a diversidade de fontes de informação e opiniões é crucial para promover uma visão mais ampla e equilibrada do mundo. Além de promover conteúdos variados, as plataformas de mídia social e os mecanismos de busca podem adotar práticas que favoreçam a exposição a diferentes perspectivas. Isso inclui a criação de espaços dedicados a visões contrastantes, a promoção de conteúdos que representem uma ampla gama de opiniões e a implementação de ferramentas que incentivem a descoberta de informações fora das bolhas de filtro.

Além disso, as ferramentas que se utilizam da IA devem ser transparentes sobre como os algoritmos selecionam e apresentam conteúdos aos usuários. Compreender como os algoritmos funcionam para selecionar e apresentar conteúdos aos usuários é fundamental para desvendar o funcionamento das bolhas de filtro. Por meio da divulgação clara dos critérios de seleção e classificação de conteúdos, os usuários podem compreender melhor por que certas informações são priorizadas, reduzindo assim a opacidade e o impacto das bolhas de filtro. Além disso, a transparência nos algoritmos não apenas ajuda a mitigar o impacto das bolhas de filtro, mas também desempenha um papel fundamental na minimização da manipulação e do viés na apresentação da informação.

Compreender a forma como a informação é apresentada aos usuários não apenas proporciona consciência a respeito de influências externas, mas também fortalece a habilidade de discernir entre informações tendenciosas e imparciais. Esse entendimento contribui para uma sociedade mais capacitada a reconhecer e lidar com vieses, destacando a importância da promoção ativa da diversidade de perspectivas e da transparência nos algoritmos. Tais medidas não só representam passos cruciais, mas essenciais na busca por uma disseminação mais equitativa e justa da informação, reduzindo as manipulações e polarizações, e cultivando uma sociedade mais bem-informada e engajada.

As plataformas de mídia social desempenham um papel crucial na disseminação de desinformação. Elas devem implementar políticas rigorosas contra a propagação de *fake news*, investindo em algoritmos de IA para identificar e remover conteúdos falsos, além de promover fontes confiáveis de informação. Além disso, é preciso promover a educação e a conscientização sobre a desinformação, ensinando às pessoas habilidades críticas de avaliação e verificação de informações. Isso pode ser feito através da educação, ensinando habilidades de avaliação de fontes e verificação de fatos. Essas abordagens combinadas, tanto por parte das plataformas de mídia social quanto das iniciativas educacionais, são essenciais para mitigar o impacto da desinformação, promovendo um ambiente online mais seguro, informado e resiliente contra a propagação de *fake news*.

Com relação à discriminação algorítmica, é importante garantir que os algoritmos de IA sejam treinados com conjuntos de dados abrangentes e representativos é apenas o primeiro passo. É fundamental abordar a representatividade nos conjuntos de dados, incorporando não apenas diversidade étnica, mas também variáveis socioeconômicas, culturais e geográficas para evitar viés nos sistemas. Além da representatividade nos

dados, a transparência dos algoritmos é um pilar crucial na busca por sistemas justos e não discriminatórios. Isso implica em tornar os algoritmos compreensíveis, fornecendo informações claras sobre seu funcionamento e as decisões que tomam.

Assim, a transparência dos algoritmos desempenha um papel fundamental não apenas na detecção de possíveis discriminações, mas também no fortalecimento da capacidade de intervenção e correção de tais problemas. A colaboração interinstitucional, envolvendo governos, entidades e especialistas, torna-se crucial para compartilhar as melhores práticas e desenvolver abordagens éticas no emprego da IA em sistemas democráticos. Além disso, a implementação de regulamentações adequadas e uma supervisão efetiva por parte dos órgãos reguladores tornam-se elementos indispensáveis para prevenir a discriminação algorítmica e assegurar a equidade nos processos democráticos.

As recomendações supracitadas combatem indiretamente a manipulação e interferência eleitoral. Mas com relação à privacidade e vigilância, medidas de segurança sólidas devem ser implementadas para proteger a privacidade e evitar vigilância indesejada. Isso inclui o uso de criptografia, práticas adequadas de controle de acesso e monitoramento contínuo da segurança dos sistemas. As organizações que empregam sistemas de IA precisam assumir a responsabilidade pela proteção da privacidade, assegurando uma governança de dados sólida que esteja em conformidade com as regulamentações de privacidade vigentes. A transparência e a conformidade com leis e regulamentos de privacidade são pilares essenciais para estabelecer a confiança dos cidadãos no uso responsável da IA.

Além das responsabilidades das organizações, a educação e a conscientização dos cidadãos desempenham um papel crucial. Informar os indivíduos sobre seus direitos de privacidade, explicando as práticas de coleta de dados e os potenciais riscos envolvidos, capacita as pessoas a tomar decisões informadas sobre como interagir com os sistemas de IA. Capacitar os cidadãos com conhecimento sobre suas próprias informações e direitos é um componente vital para garantir que as pessoas possam fazer escolhas conscientes e informadas sobre seu envolvimento com a tecnologia. Isso contribui para uma interação mais consciente e responsável com os sistemas de IA, preservando a privacidade individual em ambientes eleitorais e em todas as esferas da sociedade.

3. Conclusão

Este trabalho apresentou uma discussão sobre os perigos do uso da Inteligência Artificial na democracia por meio de uma revisão da literatura. Para isso, foi elaborada e planejada uma revisão da literatura com o intuito de listar trabalhos recentes que abordassem assuntos relacionados à IA ou à democracia. Dos trabalhos recuperados, foram excluídos aqueles que, de alguma forma, não apresentavam ameaças à democracia por meio do uso da IA. Foram apresentadas e detalhadas algumas ameaças à democracia. Além disso, discutiram-se estratégias para combater as ameaças apresentadas.

Como trabalhos futuros, pretende-se: a) explorar ameaças não abordadas neste trabalho; b) apresentar ferramentas computacionais utilizadas para combater as ameaças mencionadas; c) expandir a revisão da literatura para incluir outros tipos de documentos científicos e outras bases de dados acadêmicas; e d) conduzir uma revisão sistemática da literatura com o objetivo de sintetizar, analisar e resumir todas as evidências disponíveis sobre a temática para obter uma visão abrangente e baseada em dados existentes.

Referências

- Bernardi, A. J. B. e Costa, A. L. V. (2020). Populismo e fake news na era da pós-verdade: comparações entre estados unidos, hungria e brasil. *Cadernos de Campo: Revista de Ciências Sociais*, 1(28):385–412.
- Brizola, J. e Fantin, N. (2017). Revisão da literatura e revisão sistemática da literatura. *Revista de Educação do Vale do Arinos (RELVA)*, 3(2):23–39.
- Carvalho, L. B. (2020). A democracia frustrada: fake news, política e liberdade de expressão nas redes sociais. *Internet & Sociedade*, 1(1):172–199.
- Dahlgren, P. M. (2021). A critical review of filter bubbles and a comparison with selective exposure. *Nordicom Review*, 42(1):15–33.
- Dias, J. A., Doca, H. H., e Silva, F. F. (2021). Bots, fake news, fake faces e deepfakes: a automação, sob o viés da dromologia, como forma sofisticada de biopoder para influenciar o processo eleitoral democrático. *Pensar: Revista de Ciências Jurídicas*, 26(3):1–14.
- Freitas, J. e Teixeira, A. V. (2019). Democracia digital e avaliação continuada de políticas públicas. *Revista Brasileira de Estudos Políticos*, 119:227–252.
- Gil, A. C. (2002). *Como Elaborar Projetos de Pesquisa*. Atlas, São Paulo, 4a edição.
- Kaufman, D. e Santaella, L. (2020). O papel dos algoritmos de inteligência artificial nas redes sociais. *Revista FAMECOS*, 27(1):1–10.
- König, P. D. e Wenzelburger, G. (2020). Opportunity for renewal or disruptive force? how artificial intelligence alters democratic politics. *Government Information Quarterly*, 37(3):101489.
- Rodrigues, T. M., Bonone, L., e Mielli, R. (2020). Desinformação e crise da democracia no brasil: é possível regular fake news? *Confluências — Revista Interdisciplinar de Sociologia e Direito*, 22(3):30–52.
- Valadares, H. C. F. (2021). Fake news e (des) informação: reflexões sobre o potencial da inteligência artificial e das novas tecnologias de acelerar a erosão da democracia. *Teoria Jurídica Contemporânea*, 6(1):1–29.
- Wilner, A. S. (2018). Cybersecurity and its discontents: Artificial intelligence, the internet of things, and digital misinformation. *International Journal*, 73(2):308–316.