ISBN: 978-65-83849-24-3

RISCOS DE DESINFORMAÇÃO GERADOS POR IA E ESTRATÉGIAS DE MITIGAÇÃO

https://doi.org/10.63330/aurumpub.021-010

Rodrigo Thomé de Moura
Pós-graduação em Dança Terapia
Prisma
E-mail: rodrigofsw@gmail.com

RESUMO
Este trabalho analisou os riscos de desinformação gerados pela Inteligência Artificial e apresentou estratégias de mitigação capazes de reduzir seus impactos sociais, políticos e institucionais. O estudo teve como objetivo investigar de que forma tecnologias de IA, especialmente modelos generativos, ampliaram a produção e a circulação de conteúdos falsos, enganosos e manipulados, além de avaliar consequências desse fenômeno para a confiança pública, a democracia, a ciência e a segurança nacional. A pesquisa adotou metodologia bibliográfica e qualitativa, baseada na revisão de artigos científicos, relatórios institucionais e obras especializadas que discutiram Inteligência Artificial, desinformação digital e integridade informacional. Os resultados mostraram que a IA generativa permitiu a criação de textos, imagens, vídeos e áudios sintéticos em grande velocidade e escala, tornando a desinformação mais sofisticada e difícil de detectar. Observou-se que o realismo crescente de deepfakes, a clonagem de voz e a automação em redes de bots ampliaram significativamente a capacidade de manipulação de percepções públicas, favorecendo campanhas coordenadas e interferências em processos democráticos. A análise identificou também que esse cenário contribuiu para a erosão da confiança nas instituições, para o descrédito em evidências científicas e para vulnerabilidades em áreas sensíveis, como saúde pública e segurança nacional. O estudo concluiu que a mitigação desses riscos depende da combinação de estratégias técnicas, políticas e educacionais, incluindo ferramentas de detecção de mídias sintéticas, regulamentações de transparência algorítmica, políticas de governança digital e programas de alfabetização midiática capazes de fortalecer a capacidade crítica dos cidadãos diante do ambiente informacional contemporâneo.

Palavras-chave: Inteligência Artificial; Desinformação; Deepfakes; Segurança informacional; Mitigação.

1 INTRODUÇÃO

A rápida expansão da Inteligência Artificial (IA), especialmente em suas aplicações generativas, transformou profundamente o cenário informacional contemporâneo, criando novas possibilidades tecnológicas, mas também desencadeando riscos significativos para a circulação de informações e para a estabilidade social. O avanço de modelos capazes de produzir textos, imagens, vídeos e áudios sintéticos com elevado grau de realismo inaugurou um período no qual fronteiras entre verdade e falsificação tornaram-se cada vez mais difíceis de identificar. Autores como Goodfellow et al. (2016), Wardle e Derakhshan (2017) e Zuboff (2019) têm discutido como esse fenômeno altera as dinâmicas de comunicação, influencia a opinião pública e expõe a sociedade a novas formas de manipulação. Nesse contexto, compreender a relação entre IA e desinformação torna- se essencial para avaliar seus impactos e pensar estratégias que preservem a integridade informacional.

O presente trabalho teve como objetivo analisar os riscos de desinformação gerados por tecnologias de Inteligência Artificial e discutir estratégias de mitigação capazes de reduzir seus efeitos sociais, políticos e institucionais. Como objetivos específicos, buscou-se: compreender os fundamentos da IA generativa; examinar o funcionamento da desinformação digital e seu ecossistema de circulação; investigar como a IA potencializa a criação e disseminação de conteúdos falsos; avaliar impactos sociais e políticos decorrentes desse processo; e apresentar possíveis caminhos para enfrentamento do problema. A hipótese que orientou a pesquisa considerou que a IA, apesar de seus benefícios, intensifica substancialmente o alcance e o realismo da desinformação, contribuindo para erosão da confiança pública, manipulação democrática e vulnerabilidades em áreas como saúde, ciência e segurança nacional. A justificativa para a realização deste estudo reside na centralidade que a informação ocupa nas sociedades contemporâneas. Em um ambiente marcado pela hiperconectividade, pela velocidade dos fluxos comunicacionais e pela crescente dependência de plataformas digitais, a desinformação não representa apenas um problema técnico, mas um fenômeno estrutural que ameaça processos democráticos, políticas de saúde pública, estabilidade institucional e relações sociais. O avanço da IA generativa intensifica esse cenário, exigindo análises críticas e multidisciplinares sobre seus riscos e implicações éticas.

Para desenvolver essa investigação, o trabalho foi estruturado em quatro partes principais. Após esta introdução, a seção de metodologia descreve o percurso adotado, baseado em pesquisa bibliográfica e abordagem qualitativa. Em seguida, o desenvolvimento é organizado em quatro subseções: a primeira apresenta os conceitos fundamentais da Inteligência Artificial Generativa; a segunda discute os diferentes tipos de desinformação digital e sua dinâmica nas plataformas; a terceira analisa a IA como vetor de desinformação, destacando escalabilidade, realismo e automação; e a quarta examina os impactos sociais e políticos decorrentes desse fenômeno. Por fim, a conclusão

retoma os principais achados, reafirma a relevância do tema e aponta para a necessidade de políticas e estratégias integradas de mitigação.

Assim, esta pesquisa busca contribuir para o debate contemporâneo sobre os desafios informacionais na era da Inteligência Artificial, oferecendo uma compreensão ampla e crítica dos riscos envolvidos e reforçando a importância de respostas éticas, tecnológicas e educacionais que assegurem a preservação da verdade e da confiança social.

2 METODOLOGIA

A presente pesquisa foi desenvolvida por meio de uma abordagem qualitativa e de caráter essencialmente bibliográfico, fundamentada na análise de obras, artigos científicos, relatórios institucionais e documentos especializados que discutem Inteligência Artificial, desinformação digital e segurança informacional. A escolha por esse tipo de metodologia se justificou pela natureza conceitual e interpretativa do tema, que exige compreensão ampla das transformações tecnológicas recentes e de seus impactos sociais e políticos. Assim, a investigação buscou reunir e interpretar contribuições de autores como Goodfellow, Wardle, Derakhshan, Zuboff, Floridi, Chesney e Citron, entre outros pesquisadores que se destacam no debate contemporâneo sobre algoritmos, mídias sintéticas e integridade informacional.

O percurso metodológico consistiu em quatro etapas principais. A primeira envolveu revisão sistemática da literatura nacional e internacional, com o objetivo de identificar conceitos, definições e fundamentos teóricos relevantes sobre IA generativa, ecossistemas de desinformação e impactos sociopolíticos da manipulação informacional. Em seguida, foram selecionados estudos de caso, relatórios de organizações de pesquisa e documentos de referência produzidos por entidades como o Council of Europe, que abordam fenômenos como deepfakes, campanhas coordenadas e interferência digital em processos democráticos. A terceira etapa concentrou-se na análise crítica desse material, buscando relacionar os achados teóricos com as transformações observadas na sociedade, especialmente no que diz respeito ao avanço de modelos generativos, à circulação de conteúdos enganosos e aos mecanismos algorítmicos que amplificam a desinformação.

Por fim, os resultados obtidos foram organizados de forma temática nas seções do desenvolvimento, permitindo uma discussão integrada entre fundamentos técnicos e impactos sociais. Essa estrutura metodológica possibilitou compreender o fenômeno de maneira ampla, articulando perspectivas tecnológicas, sociológicas e políticas, e oferecendo subsídios para reflexão sobre estratégias de mitigação que possam contribuir para o enfrentamento dos riscos associados à Inteligência Artificial. Dessa forma, a metodologia adotada não apenas sustentou teoricamente o estudo, mas também permitiu construir uma análise consistente, crítica e alinhada às necessidades contemporâneas de proteção da

integridade informacional.

3 DESENVOLVIMENTO

3.1 INTELIGÊNCIA ARTIFICIAL GENERATIVA

A Inteligência Artificial Generativa tornou-se um dos campos mais inovadores e transformadores da tecnologia contemporânea, marcada pela capacidade de criar conteúdos novos e originais a partir do aprendizado de grandes volumes de dados. Em termos conceituais, ela se refere a sistemas capazes de gerar textos, imagens, sons, vídeos e outros formatos de mídia sem que esses conteúdos tenham sido previamente armazenados, mas sim produzidos a partir de padrões aprendidos durante o treinamento. Segundo Goodfellow et al. (2016), esse tipo de IA baseia-se em modelos probabilísticos capazes de compreender estruturas complexas da linguagem e da percepção humana, reproduzindo-as de forma autônoma e surpreendentemente natural. Essa característica coloca a IA generativa como um marco na história da computação, ampliando seu uso em educação, saúde, design, comunicação e entretenimento, ao mesmo tempo em que abre discussões fundamentais sobre ética e responsabilidade.

Dentro dessa categoria, os modelos de linguagem de grande escala, conhecidos como LLMs (*Large Language Models*), tornaram-se os mais populares. Eles são capazes de gerar textos coerentes, traduzir conteúdos, responder perguntas, produzir resumos e até simular estilos de escrita específicos. Esses modelos, como GPT, Gemini e LLaMA, foram treinados em bilhões de palavras, desenvolvendo a habilidade de prever a próxima palavra em uma sequência e, assim, construir narrativas completas. Conforme Kaplan et al. (2020) apontam, quanto maior o volume de dados e parâmetros de um modelo, maior tende a ser sua capacidade de gerar respostas sofisticadas. Paralelamente, avanços semelhantes ocorreram no campo da geração de imagens e vídeos, com modelos como DALL·E, Midjourney e Stable Diffusion, que são capazes de criar ilustrações realistas, fotografias sintéticas e cenas complexas baseadas em descrições textuais. Essas ferramentas expandiram a criatividade digital, permitindo que pessoas sem formação técnica produzissem conteúdos antes restritos a profissionais altamente especializados.

Além dos textos e imagens, a IA generativa também revolucionou o audiovisual por meio da síntese de voz e da manipulação de vídeos. As tecnologias de clonagem vocal permitem replicar com alta fidelidade o timbre e a entonação de qualquer pessoa, gerando discursos inteiros que parecem autênticos. Da mesma forma, os *deepfakes* — vídeos manipulados por redes neurais profundas — ganharam notoriedade por sua capacidade de substituir rostos, sincronizar falas e alterar expressões com extremo realismo. Estudos como os de Chesney e Citron (2019) destacam que os *deepfakes* representam uma das formas mais preocupantes de conteúdo sintético, devido ao seu potencial para

disseminar desinformação, comprometer reputações, manipular processos políticos e criar evidências visuais falsas. A sofisticação desses métodos demonstra como a IA generativa pode tanto ampliar ferramentas criativas quanto gerar riscos significativos para a integridade informacional e para a confiança pública.

Nesse cenário, torna-se essencial compreender que a Inteligência Artificial Generativa não consiste apenas em um avanço técnico, mas em um fenômeno cultural, social e ético. À medida que a produção de conteúdo sintético se torna indistinguível da produção humana, questões relacionadas à autenticidade, autoria, privacidade e veracidade ganham centralidade no debate contemporâneo. Assim, estudar seus conceitos básicos, seus modelos e suas aplicações — inclusive as potencialmente nocivas — é fundamental para garantir que sua utilização ocorra de forma responsável, transparente e alinhada aos princípios éticos que regem a sociedade digital.

3.2 DESINFORMAÇÃO DIGITAL

A desinformação digital tornou-se um dos problemas mais urgentes da sociedade contemporânea, intensificada pela velocidade e pela escala das interações mediadas por tecnologias digitais. No ambiente online, a desinformação assume diferentes formas, cada uma com características específicas e graus variados de intencionalidade. As chamadas *fake news* são conteúdos totalmente falsos, produzidos deliberadamente para enganar, manipular ou gerar repercussão. Há também conteúdos manipulados, que se baseiam em fatos parcialmente verdadeiros, mas alterados ou retirados de contexto para induzir interpretações equivocadas. Além disso, as informações enganosas — conhecidas como *misinformation* — são aquelas compartilhadas sem a intenção explícita de causar dano, mas que acabam contribuindo para a circulação de narrativas falsas ou distorcidas. Wardle e Derakhshan (2017) classificam esses fenômenos dentro de um espectro que vai do erro não intencional à manipulação estratégica, destacando que cada categoria demanda formas específicas de identificação e combate.

Esse tipo de conteúdo se fortalece em um ecossistema digital marcado pela hiperconectividade, pela lógica de viralização e pela fragmentação das fontes de informação. Nas redes sociais, a desinformação circula de forma acelerada porque encontra terreno fértil em interações rápidas, pouco verificadas e emocionalmente carregadas. Plataformas como Facebook, X/Twitter, WhatsApp, TikTok e Instagram funcionam como ambientes onde conteúdos atraentes têm maior chance de serem compartilhados, independentemente de sua veracidade. Segundo Vosoughi, Roy e Aral (2018), notícias falsas se espalham mais rapidamente que notícias verdadeiras justamente porque tendem a despertar surpresa, indignação e forte engajamento emocional. Isso significa que o design das plataformas — estruturado para maximizar atenção e engajamento — contribui para perpetuar ciclos

de desinformação, muitas vezes tornando difícil para o usuário médio distinguir entre informação verificada e manipulação intencional.

O papel dos algoritmos e da automação nesse processo é igualmente central. Ferramentas de recomendação baseadas em inteligência artificial selecionam conteúdos de acordo com padrões de interesse do usuário, criando bolhas informacionais que reforçam crenças pré-existentes e dificultam o acesso a perspectivas diversas. Esses algoritmos, ao privilegiarem engajamento, acabam amplificando mensagens sensacionalistas ou polarizadoras, tornando a desinformação mais visível que conteúdos verificáveis. Paralelamente, sistemas automatizados — como *bots* e contas coordenadas — são frequentemente utilizados para aumentar artificialmente o alcance de determinados conteúdos ou campanhas. Zannettou et al. (2019) mostram que grupos organizados utilizam redes de bots para impulsionar narrativas falsas, manipular debates públicos e influenciar processos políticos, transformando a desinformação em um fenômeno sistêmico e difícil de conter.

Nesse cenário, compreender a complexidade da desinformação digital exige analisar simultaneamente seus tipos, suas formas de circulação e os mecanismos tecnológicos que amplificam seu impacto. A interação entre usuários, plataformas e algoritmos cria um ambiente no qual conteúdos falsos ou manipulados encontram elevada capacidade de difusão, com efeitos diretos sobre a opinião pública, a confiança social e o funcionamento das instituições democráticas. Assim, o estudo desse fenômeno torna-se fundamental para o desenvolvimento de estratégias eficazes de mitigação e de políticas públicas que fortaleçam a integridade informacional na era digital.

3.3 IA COMO VETOR DE DESINFORMAÇÃO

A Inteligência Artificial tem se consolidado como um vetor central na expansão da desinformação digital, principalmente por sua capacidade de produzir, replicar e amplificar conteúdos falsos em velocidade e escala sem precedentes. Diferentemente dos processos tradicionais de manipulação informacional, que exigiam tempo, recursos e conhecimentos técnicos avançados, a IA tornou possível criar textos, imagens, vídeos e áudios falsificados com poucos comandos, democratizando a capacidade de gerar conteúdos enganosos. Ferramentas generativas permitem que um único indivíduo produza centenas ou milhares de mensagens falsas em minutos, o que intensifica o volume de desinformação circulante e sobrecarrega os mecanismos de verificação. Segundo Floridi (2021), a automação transformou a desinformação em um fenômeno industrializado, capaz de se espalhar rapidamente e de forma altamente adaptável às dinâmicas das redes sociais.

Um dos aspectos mais preocupantes dessa transformação é o realismo crescente das mídias sintéticas geradas por IA. Modelos avançados produzem imagens extremamente detalhadas, vídeos que simulam expressões faciais com precisão e áudios que imitam perfeitamente a voz humana. Os

deepfakes tornaram- se uma das expressões mais visíveis dessa capacidade, permitindo a criação de vídeos nos quais pessoas aparentam dizer ou fazer coisas que nunca ocorreram. Conforme destacam Chesney e Citron (2019), esses conteúdos têm potencial para comprometer reputações, manipular decisões políticas, gerar pânico em situações de crise e minar a confiança em registros audiovisuais — um elemento historicamente central na validação da verdade. Apesar de inicialmente detectáveis por falhas visuais, os *deepfakes* evoluíram a ponto de se tornarem indistinguíveis a olho nu, o que aumenta significativamente seu impacto.

Além da produção direta de conteúdos falsos, a IA amplia a desinformação por meio de bots automatizados e campanhas coordenadas. Esses sistemas são programados para simular comportamentos humanos, participar de debates, impulsionar hashtags, comentar publicações e reforçar narrativas específicas. A automação permite que um pequeno grupo crie a impressão de consenso social ou engajamento orgânico em torno de temas sensíveis. Zannettou et al. (2019) demonstram que redes de bots têm sido amplamente utilizadas por grupos políticos, econômicos e ideológicos para manipular percepções públicas, interferir em debates eleitorais e amplificar mensagens polarizadoras. Em muitos casos, esses bots funcionam de forma articulada com modelos de IA generativa, criando uma cadeia de produção e distribuição de desinformação altamente eficiente.

Os impactos desse fenômeno são visíveis em diferentes contextos recentes. Em processos eleitorais, conteúdos falsificados por IA foram utilizados para difamar candidatos, manipular narrativas e influenciar a opinião pública, como ocorreu em eleições nos Estados Unidos, Índia e diversos países europeus. Durante crises sanitárias, como a pandemia de COVID-19, vídeos e textos sintéticos foram usados para espalhar informações falsas sobre vacinas, tratamentos e medidas de prevenção, contribuindo para comportamentos de risco e desconfiança institucional. Em áreas como economia e segurança, mídias manipuladas já provocaram quedas temporárias em mercados financeiros e espalharam alarmes sobre ameaças inexistentes, demonstrando a capacidade da IA de gerar impactos concretos e imediatos na vida social.

Assim, a Inteligência Artificial, embora represente um avanço tecnológico significativo, também se tornou um poderoso catalisador da desinformação. Sua capacidade de produzir conteúdos altamente realistas, operar de forma automatizada e interagir com as dinâmicas algorítmicas das plataformas digitais faz com que se torne um elemento central na compreensão dos desafios contemporâneos ligados à integridade informacional. Reconhecer esse papel é fundamental para o desenvolvimento de estratégias de mitigação que sejam capazes de responder à velocidade, sofisticação e escala do problema.

3.4 IMPACTOS SOCIAIS E POLÍTICOS

Os impactos sociais e políticos da desinformação amplificada por tecnologias de Inteligência Artificial constituem um dos desafios mais graves enfrentados pelas sociedades contemporâneas. A circulação massiva de conteúdos falsos, manipulados ou enganosos contribuiu significativamente para a erosão da confiança pública nas instituições, nos meios de comunicação e até mesmo na própria noção de verdade compartilhada. Quando informações sintéticas geradas por IA se tornam indistinguíveis de conteúdos reais, cidadãos passam a questionar a credibilidade de evidências visuais, declarações oficiais e fatos comprovados. Zuboff (2019) destaca que esse cenário de incerteza constante enfraquece a coesão social e cria terreno fértil para discursos polarizadores, tornando mais difícil a construção de consensos mínimos necessários para o funcionamento democrático. Assim, a confiança — base de qualquer sociedade estruturada — torna-se fragilizada diante de um ambiente no qual a autenticidade das informações se encontra permanentemente sob suspeita.

A desinformação apoiada por IA também desempenha papel central na manipulação de opinião pública e na interferência em processos democráticos. Em épocas eleitorais, conteúdos falsificados podem alterar percepções, moldar narrativas e influenciar a decisão de milhares de eleitores em curtos intervalos de tempo. Deepfakes, textos fabricados por modelos de linguagem e redes de bots coordenadas ampliam o alcance de campanhas manipulativas, criando impressões artificiais de apoio popular ou disseminando falsas acusações contra adversários políticos. Wardle e Derakhshan (2017) afirmam que esses mecanismos desafiam os princípios fundamentais da deliberação democrática ao distorcer o debate público, dificultar o acesso a informações confiáveis e manipular emoções de maneira estratégica. A interferência digital em eleições já foi documentada em diversos países, evidenciando que a integridade eleitoral pode ser significativamente comprometida por tecnologias de automação informacional.

Além do campo político, os impactos estendem-se à ciência, à saúde pública e à segurança nacional. A disseminação de informações falsas sobre tratamentos médicos, diagnóstico de doenças ou campanhas de vacinação — amplamente observada durante a pandemia de COVID-19 — demonstrou como a desinformação pode gerar comportamentos de risco, reduzir adesão a medidas sanitárias e agravar crises epidemiológicas. No campo científico, a propagação de teorias conspiratórias e conteúdos anticientíficos desacredita pesquisadores, dificulta políticas baseadas em evidências e reduz a capacidade da população de diferenciar conhecimento validado de especulação. Já na segurança nacional, deepfakes e manipulações audiovisuais podem ser utilizados para provocar instabilidade, simular ataques, criar falsas declarações de autoridades ou gerar pânico coletivo. Chesney e Citron (2019) alertam que a sofisticação das falsificações produzidas por IA pode inviabilizar a distinção entre ameaças reais e fabricadas, dificultando a ação rápida de instituições responsáveis por proteção social

e estatal.

Diante desse cenário, torna-se evidente que a desinformação potencializada pela IA não representa apenas um problema técnico, mas um fenômeno social e político de grande magnitude. Seus efeitos atravessam instituições, processos decisórios, sistemas de saúde e relações sociais, comprometendo pilares fundamentais da convivência democrática. Compreender esses impactos é um passo essencial para formular políticas eficazes de mitigação, fortalecer a alfabetização digital da população e garantir que as tecnologias emergentes sejam utilizadas de forma ética e responsável.

4 CONCLUSÃO

O presente trabalho permitiu compreender que a Inteligência Artificial, especialmente em suas aplicações generativas, desempenhou papel decisivo na transformação do ecossistema informacional contemporâneo. A capacidade dessas tecnologias de produzir textos, imagens, vídeos e áudios sintéticos com elevado grau de realismo alterou profundamente as dinâmicas de circulação de informações, ampliando o potencial de criação e disseminação de conteúdos falsos, manipulados ou enganosos. A análise revelou que a IA não apenas acelerou esses processos, mas também os tornou mais complexos, eficientes e difíceis de detectar, configurando um cenário de desinformação mais poderoso e sofisticado do que qualquer outro já observado em períodos anteriores.

Os resultados indicaram que a desinformação gerada ou impulsionada por IA contribuiu diretamente para a erosão da confiança pública, fenômeno que afeta instituições democráticas, sistemas de comunicação, práticas científicas e relações sociais. A presença crescente de deepfakes, discursos sintéticos, imagens manipuladas e campanhas automatizadas dificultou a distinção entre conteúdo verdadeiro e falso, ampliando a sensação de incerteza e vulnerabilidade informacional. Além disso, destacou-se que essas tecnologias facilitaram a manipulação de opinião pública e interferências políticas, criando condições para campanhas coordenadas que afetam processos eleitorais, agendas governamentais e debates públicos de forma significativa. Evidenciou-se também que a desinformação apoiada por IA ocasionou prejuízos à ciência, à saúde e à segurança nacional, especialmente em contextos de crise, como pandemias e instabilidades geopolíticas. Diante desse cenário, o estudo demonstrou que os riscos associados à IA não residem apenas em sua capacidade técnica, mas na forma como tais tecnologias são incorporadas às dinâmicas sociais e políticas. Assim, a mitigação desses riscos depende de ações integradas e multidimensionais. As estratégias analisadas indicaram a necessidade de combinar soluções técnicas, como ferramentas de detecção de deepfakes, marcação de conteúdos sintéticos e sistemas de autenticação, com políticas regulatórias que estabeleçam transparência algorítmica, responsabilização de plataformas digitais e mecanismos de governança que protejam a integridade informacional. Da mesma forma, destacou-se a importância da educação

midiática como estratégia fundamental para fortalecer a capacidade crítica dos cidadãos, permitindo que reconheçam sinais de manipulação e desenvolvam maior autonomia diante de conteúdos enganosos.

Conclui-se, portanto, que enfrentar a desinformação gerada por IA exige um esforço conjunto entre governos, pesquisadores, empresas tecnológicas, instituições educativas e sociedade civil. Apenas por meio dessa cooperação será possível garantir que a Inteligência Artificial seja empregada de maneira ética, responsável e alinhada aos valores democráticos, preservando a confiança social e a segurança informacional em um contexto digital cada vez mais complexo e desafiador.

REFERÊNCIAS BIBLIOGRÁFICAS

CHESNEY, Robert; CITRON, Danielle. *Deep fakes: A looming challenge for privacy, democracy, and national security*. **California Law Review**, v. 107, p. 1753–1819, 2019. Disponível em: https://www.californialawreview.org/print/deep-fakes-a-looming- challenge-for-privacy-democracy-and-national-security/. Acesso em: 14 nov. 2025.

FLORIDI, Luciano. **The ethics of artificial intelligence**. Oxford: Oxford University Press, 2021.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. Cambridge: MIT Press, 2016.

KAPLAN, Jared et al. *Scaling laws for neural language models*. **arXiv**, 2020. Disponível em: https://arxiv.org/abs/2001.08361. Acesso em: 14 nov. 2025.

VOSOUGHI, Soroush; ROY, Deb; ARAL, Sinan. *The spread of true and false news online*. **Science**, v. 359, n. 6380, p. 1146–1151, 2018. Disponível em: https://www.science.org/doi/10.1126/science.aap9559. Acesso em: 14 nov. 2025.

WARDLE, Claire; DERAKHSHAN, Hossein**.** ***Information disorder:*** *Toward an interdisciplinary framework*. Strasbourg: Council of Europe, 2017. Disponível em: https://firstdraftnews.org/wp-content/uploads/2017/11/PREMS-162317-GBR-2018- Report-de%CC%81sinformation-1.pdf. Acesso em: 14 nov. 2025.

ZANNETTOU, Savvas et al. *Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web*. In: **Companion Proceedings of the World Wide Web Conference (WWW)**, 2019. Disponível em: https://arxiv.org/pdf/1801.09288. Acesso em: 14 nov. 2025.

ZUBOFF, Shoshana. ***The age of surveillance capitalism: the fight for a human future at the new frontier of power***. New York: PublicAffairs, 2019.