

O Combate à Desinformação nas Plataformas Digitais: Desafios e Oportunidades

Julio C. S. Reis¹, Philipe Melo², Márcio Silva³, Fabrício Benevenuto⁴

¹Departamento de Informática – Universidade Federal de Viçosa (UFV)

²Instituto de Ciências Exatas e Tecnológicas - Universidade Federal de Viçosa (UFV) – Florestal

³Faculdade de Ciência da Computação - Universidade Federal de Mato Grosso do Sul (UFMS)

⁴Departamento de Ciência da Computação - Universidade Federal de Minas Gerais (UFMG)

jreis@ufv.br, philipe.freitas@ufv.br,
marcio.inacio@ufms.br, fabricio@dcc.ufmg.br

Abstract. *Digital platforms have evolved into powerful ecosystems for the production and dissemination of information in the online environment. They have drastically changed the way we communicate and receive information, opening up unprecedented opportunities and creating complex challenges, e.g. in the fight against misinformation. This is one of the most serious and complex challenges of our time, with profound implications such as political polarization, attacks on democracy and risks to public health. This complexity requires solutions that span multiple fields of knowledge and the collaboration of different sectors of society. This article explores the challenges and opportunities that arise in combating misinformation on digital platforms.*

Resumo. *Plataformas digitais se tornaram poderosos ecossistemas de produção e disseminação de informação no ambiente online, transformando drasticamente o modo como nos comunicamos e nos informamos, abrindo oportunidades imprevistas e criando desafios complexos, como os relacionados ao combate à desinformação. Este é um dos desafios mais graves e complexos da atualidade, com impactos profundos, como a polarização política, ataques à democracia e riscos à saúde pública. Essa complexidade exige soluções que envolvam múltiplas áreas do conhecimento e a colaboração de diversos setores da sociedade. Este artigo explora os desafios e oportunidades que surgem no combate à desinformação em plataformas digitais.*

1. Introdução

Plataformas digitais deixaram de ser simples ferramentas de interação entre pessoas (e.g., amigos e familiares) para se tornarem poderosos ecossistemas de produção e distribuição de informação no ambiente online, transformando radicalmente o modo como nos comunicamos e nos informamos. A democratização da inclusão digital e acesso à informação, impulsionada com a proliferação de *smartphones*, trouxe diversos benefícios para a sociedade, mas também revelou vulnerabilidades deste formato de comunicação. O espaço digital não facilita somente a disseminação de informações verídicas e notícias oriundas de fontes confiáveis, mas também favorece a propagação de conteúdos enganosos, que

podem distorcer percepções e influenciar decisões críticas, como nas esferas da política e saúde pública [Reis et al., 2023]. Enquanto até pouco tempo atrás tínhamos um fluxo relativamente controlado de informações, mediado por fontes mais confiáveis, hoje temos cada vez mais um universo gigantesco de dados amplificados por algoritmos e inteligência artificial, onde a desinformação se espalha com uma eficiência, velocidade e escala, muitas vezes, até maior que o conhecimento verificado [Vosoughi et al., 2018].

Nesta direção, plataformas como Facebook, Instagram, Twitter/X, e aplicativos de mensagem instantânea como WhatsApp e Telegram, hoje acessados por bilhões de pessoas, desempenham um papel crucial, pois intensificam ainda mais esse problema ao priorizar o engajamento¹ sobre a precisão e confiança do conteúdo compartilhado, ou permitindo ainda conteúdo viral e anônimo circular por uma rede mais privada e particular protegida por camadas de criptografia [Benevenuto and Melo, 2024]. Parte do desafio é inerente à natureza das próprias plataformas: (i) muitas vezes é mais oportuno e menos dispendioso produzir e consumir informações (*i.e.*, notícias) nesses ambientes em comparação com meios noticiosos tradicionais, como jornais, rádio ou televisão; e (ii) é mais fácil compartilhar, comentar e discutir com amigos ou outros leitores em plataformas digitais, o que melhora e/ou impulsiona a comunicação e as interações entre os usuários nestes sistemas [Shu et al., 2017].

A desinformação é um dos desafios mais graves da atualidade²³, com impactos profundos, como a polarização política, ataques à democracia e riscos à saúde pública. Ela se propaga por qualquer plataforma digital com grande audiência e permeia todos os tipos de conteúdo como imagens, texto, áudio e vídeo. Essa desinformação acaba tendo um impacto devastador na ciência, em particular em áreas como saúde pública [Suarez-Lledo and Alvarez-Galvez, 2021] e crise climática [Treen et al., 2020], ao distorcer fatos e minar a confiança no conhecimento científico. Essa complexidade exige soluções que envolvam múltiplas áreas do conhecimento e a colaboração de diversos setores da sociedade. Do ponto de vista da computação, a produção de tecnologias que possam identificar, monitorar e conter a disseminação de campanhas de desinformação em plataformas digitais é essencial, mas insuficiente se não for acompanhada de uma coordenação entre pesquisadores, tecnólogos, legisladores, educadores e a sociedade civil para reforçar a resiliência cívica e proteger a integridade das informações e dos processos democráticos. Este artigo explora os desafios e oportunidades que surgem no combate à desinformação em plataformas digitais, analisando as estratégias tecnológicas emergentes e as abordagens regulatórias necessárias para restaurar a confiança pública e proteger a integridade dos processos informacionais e democráticos.

2. O Desafio do Combate à Desinformação

Combater a desinformação é uma típica luta adversária. A cada evento crítico, as campanhas de desinformação exploram novas estratégias de manipular a informação/opinião, enquanto novos mecanismos de defesa são criados visando mitigar tais campanhas de desinformação e conter seu avanço [Wu et al., 2024]. Em contextos como saúde, por exemplo, campanhas de desinformação podem ter motivações completamente diferentes [Swire-Thompson et al., 2020], o que torna o problema ainda mais desafiador.

¹<https://www.facebook.com/business/help/718033381901819>

²<https://tinyurl.com/4jn5tzvz>

³<https://tinyurl.com/2efvfvpa>

Aqui, são elencadas algumas questões que consideramos críticas de serem abordadas e que comitadamente configuram oportunidades de pesquisa nesta temática. Uma visão geral dessas questões é apresentada na Figura 1.

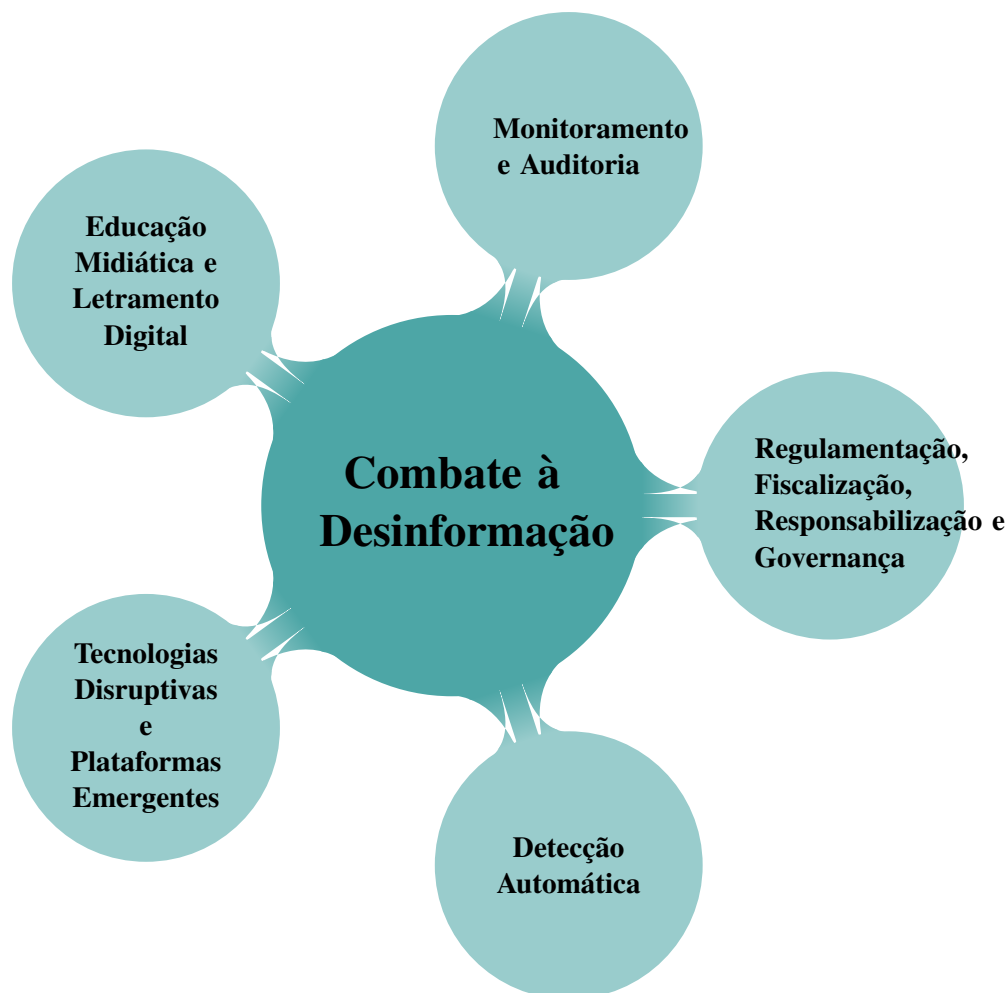


Figura 1. Dimensões que abarcam desafios e oportunidades de pesquisa no combate à desinformação em plataformas digitais.

Monitoramento e Auditoria de Plataformas Digitais

As plataformas digitais passam por constantes mudanças que impactam diretamente como as informações são disseminadas. O WhatsApp, por exemplo, implementou recentemente o recurso de comunidades, que potencializa o disparo massivo de mensagens, enquanto também colocou rótulos que auxiliam o usuário a identificar uma mensagem compartilhada várias vezes dentro da plataforma [Melo et al., 2024]. Além disso, em vários casos, as plataformas digitais são utilizadas apenas como veículos (ou vitrine) para disseminação de conteúdos produzidos por *websites* externos dedicados exclusivamente à produção e disseminação de desinformação [Couto et al., 2024]. Neste cenário, *viabilizar o acesso e a obtenção dos dados para monitoramento e/ou auditoria desses sistemas é algo essencial*. O monitoramento e auditoria de plataformas digitais consistem em medidas para trazer maior transparência ao espaço midiático. Ademais, elas viabilizam uma avaliação crítica do ecossistema online, com potencial para colocar em evidência a (des)informação

que tem circulado de forma viral nestes ambientes, além de revelar como os algoritmos operam, investigando quais critérios são utilizados na disseminação e potenciais estratégias de priorização de conteúdo. Estas estratégias contribuem para o aumento da confiança do público nas plataformas, e também estimulam uma cultura de responsabilidade entre as empresas de tecnologia, incentivando a colaboração entre diferentes atores, como governos, ONGs e pesquisadores, na criação de soluções eficazes para o combate à desinformação.

Regulamentação, Fiscalização, Responsabilização e Governança

Ter regulamentações efetivas é fundamental no combate à desinformação, pois estabelece diretrizes claras para o funcionamento das plataformas digitais e responsabiliza os devidos atores pela disseminação de informações errôneas de forma deliberada e intencional. Essas regulamentações podem incluir exigências para a transparência na publicidade política, a rotulação de conteúdo falso e a proibição de práticas fraudulentas, como a criação de perfis falsos. Com regras bem definidas, é possível limitar o alcance de campanhas de desinformação e garantir que os usuários tenham acesso a informações mais precisas e verificáveis. Além disso, regulamentações robustas podem incentivar as plataformas a adotarem tecnologias e práticas que ajudem a detectar e remover conteúdos enganosos de forma mais eficaz. Neste cenário, protocolos como C2PA⁴ que se baseiam em criptografia podem ser utilizados para codificar detalhes sobre as origens de um conteúdo em uma plataforma digital. Tais medidas podem ser utilizadas no combate à desinformação, bem como para fiscalização e responsabilização das plataformas. Ainda, para a criação de marcos regulatórios eficazes, é necessária a colaboração entre governos, especialistas de diferentes áreas do conhecimento e a sociedade civil. Isso garante que as regras sejam atualizadas e adaptáveis às rápidas mudanças no ambiente digital. Finalmente, a implementação de mecanismos de fiscalização e sanções para aqueles que desrespeitam as normas é igualmente essencial ao assegurar que as plataformas levem a sério a responsabilidade de monitorar e gerenciar o conteúdo. Essas ações, juntamente com estratégias de governança [Almeida et al., 2024], são fundamentais para a criação de um ambiente mais responsável, confiável, monitorável, e por fim, propício ao combate à desinformação.

Deteção Automática de Desinformação

O desenvolvimento de abordagens de detecção automática de desinformação representa um desafio complexo e multifacetado. À medida que as estratégias para disseminação de desinformação se tornam mais sofisticadas e crescem em velocidade e escala, é essencial que as ferramentas de detecção automática evoluam simultaneamente para acompanhar esse ritmo. Neste contexto, técnicas de aprendizado de máquina e inteligência artificial têm sido empregadas para analisar características de linguagem, padrões de propagação, identificar fontes suspeitas e classificar conteúdos [Reis et al., 2019]. No entanto, a eficácia dessas abordagens é dependente da qualidade dos dados utilizados, bem como da capacidade de adaptação às nuances culturais e contextuais que influenciam a percepção da verdade. Além disso, a detecção automática de desinformação enfrenta desafios éticos e práticos. Questões como a privacidade dos usuários, a liberdade de expressão e a possibilidade de viés algorítmico devem ser cuidadosamente consideradas durante o projeto

⁴<https://c2pa.org/>

e desenvolvimento de tais ferramentas. Neste contexto, é necessário criar soluções eficazes que não apenas protejam a integridade da informação, mas também respeitem os direitos individuais das pessoas. Ainda neste sentido, também é crucial que os sistemas de detecção não sirvam apenas para identificar a desinformação, mas também ofereçam transparência sobre como essas determinações são feitas, o que pode ser útil para substantiar e/ou suportar vereditos fornecidos por agências de checagem de fatos, por exemplo. É fundamental ressaltar que a desinformação não aparece apenas por meio de textos. Atualmente, ela também se espalha nas plataformas digitais num formato multimídia, como áudio, imagem e vídeos (como *deepfakes*). Assim, é imprescindível que as abordagens automáticas de detecção sejam projetadas, adaptadas ou desenvolvidas para lidar com diferentes tipos de mídia.

Tecnologias Disruptivas e Plataformas Emergentes

O ambiente online muda constantemente, surgindo novas plataformas e tecnologias. Neste contexto, campanhas de desinformação se beneficiam do surgimento de “novidades”, especialmente aquelas que apresentam menos controle e regulamentação. Logo, *estratégias de combate à desinformação precisam ser propostas, revistas e atualizadas constantemente para acompanhar e contemplar evoluções e/ou mudanças tecnológicas*. Em resposta a esses desafios, tecnologias emergentes, como modelos de linguagem avançados, estão sendo desenvolvidas para auxiliar no combate à disseminação de informações falsas [Liu et al., 2024]. No entanto, a própria sofisticação dessas ferramentas também introduz novos riscos. Modelos de linguagem, como o ChatGPT⁵ e DeepSeek⁶, têm o potencial de gerar conteúdo incorreto, inclusive muitas vezes de forma a parecer confiável [Kreps et al., 2022]. Além disso, a utilização maliciosa dessas ferramentas por atores com intenções de manipulação aumenta o risco de danos sociais e políticos. Modelos generativos que, devido ao acesso massivo e à facilidade no uso sem critérios nem regras bem definidos, têm apresentado potencial para serem explorados como recurso para a propagação de desinformação. Neste contexto, é fundamental uma análise constante de como essas tecnologias e plataformas emergentes estão/podem ser utilizadas para a disseminação da desinformação no ambiente online, provendo uma avaliação diagnóstica que compreenda uma investigação, inclusive, de questões éticas relacionadas, que forneça insumos para criação de novas e/ou atualização de medidas de combate.

Educação Midiática e o Letramento Digital

A educação midiática e o letramento digital são indispensáveis no combate à desinformação. Elas capacitam os indivíduos a analisar criticamente as informações que consomem e compartilham [Jones-Jang et al., 2021]. Ao aprender a identificar fontes confiáveis, entender a diferença entre fatos e opiniões e reconhecer técnicas de manipulação, os cidadãos se tornam mais resistentes à conteúdos enganosos. Essa formação melhora a capacidade de discernimento em relação às informações compartilhadas em plataformas digitais e também promove uma cultura de verificação e responsabilidade no consumo de informação, essencial em um mundo saturado de dados. Além disso, outro desafio neste contexto está relacionado a como promover uma educação midiática mais inclusiva e acessível. Programas educacionais que ensinam habilidades de

⁵<https://chatgpt.com/>

⁶<https://www.deepseek.com/>

navegação na Internet, análise crítica de conteúdos e compreensão das dinâmicas desses sistemas podem ter um impacto significativo na forma como as pessoas interagem com a informação.

3. Considerações Finais

A desinformação é uma das questões mais desafiadoras ultimamente, com consequências graves, incluindo polarização política, ataques à democracia e riscos para a saúde pública, bem como para diversas outras áreas. A desinformação manifesta-se em qualquer plataforma com uma grande base de usuários, incluindo redes sociais online, aplicações de mensagens e em última instância a Internet. Logo, medir o impacto causado pela propagação deste tipo de conteúdo nestes ambientes é algo extremamente desafiador. É difícil aferir o quanto a desinformação influencia decisões políticas ou votos, por exemplo. Neste cenário, propor e/ou explorar métricas mais palpáveis que podem ser acessadas por pesquisas de opiniões no sentido de identificar, por exemplo, se as pessoas acreditam em teorias conspiratórias, na ciência, nas vacinas, nas mudanças climáticas, etc., pode ser uma maneira de mensurar seu impacto e gerar insumos para a proposição de estratégias efetivas contra o problema.

Finalmente, a desinformação permeia todas as formas de mídia e conteúdo, incluindo imagens, texto, áudio e vídeo. Tudo isso torna a desinformação um problema complexo e desafiador, que requer o envolvimento de diferentes áreas do conhecimento e segmentos da sociedade. Nesse artigo, apresentamos uma série de questões que consideramos críticas de serem abordadas pela sociedade brasileira para mitigar o problema da desinformação nos próximos anos.

4. Currículo Resumido dos Autores

Os autores deste artigo possuem doutorado em Ciência da Computação, são professores e pesquisadores com diversas publicações científicas de impacto no tema desinformação. Todos os autores atuaram no desenvolvimento de sistemas reais para o combate à desinformação através do projeto “Eleições sem Fake”⁷, vinculado ao Programa Permanente de Enfrentamento à Desinformação do Tribunal Superior Eleitoral (TSE). O Prof. Julio C. S. Reis fez sua tese de doutorado sobre detecção automática de desinformação em plataformas digitais (vencedora do CTD-SBSI e premiada no CTD-SBC). A tese de doutorado do Prof. Márcio Silva envolveu a construção de um sistema para trazer transparência para as propagandas realizadas em plataformas digitais, resultando em premiações relevantes internacionais, como o concorrido prêmio de proteção à privacidade dado pelo CNIL e INRIA de 2020. O Prof. Philipe Melo manteve no ar um sistema de monitoramento de grupos públicos de WhatsApp e Telegram por 4 anos, oferecendo material a dezenas de reportagens jornalísticas e para checagem de fatos, com destaque para duas reportagens do Fantástico. Finalmente, o Prof. Fabrício Benevenuto, é pesquisador nível 1D do CNPq e coordenador do projeto “Eleições sem Fake”. Ele possui dezenas de artigos científicos relacionado ao tema, incluindo o artigo vencedor do *test-of-time award* do ICWSM 2020. Fabrício também é membro do conselho consultivo de segurança do TikTok, dentre outras atividades ligadas ao combate à desinformação.

⁷<http://www.eleicoesemfake.dcc.ufmg.br/>

Agradecimentos

Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Referências

- Almeida, V., Almeida, J. M., and Meira, W. (2024). The role of computer science in responsible ai governance. *IEEE Internet Computing*, 28(3):55–58.
- Benevenuto, F. and Melo, P. (2024). Misinformation campaigns through whatsapp and telegram in presidential elections in brazil. *Communications of the ACM*, 67(8):72–77.
- Couto, J. M., Reis, J. C., and Benevenuto, F. (2024). Can computer network attributes be useful for identifying low-credibility websites? a case study in brazil. *Social Network Analysis and Mining (SNAM)*, 14(1):153.
- Jones-Jang, S. M., Mortensen, T., and Liu, J. (2021). Does media literacy help identification of fake news? information literacy helps, but other literacies don't. *American Behavioral Scientist*, 65(2):371–388.
- Kreps, S., McCain, R. M., and Brundage, M. (2022). All the news that's fit to fabricate: Ai-generated text as a tool of media misinformation. *Journal of experimental political science*, 9(1):104–117.
- Liu, A., Sheng, Q., and Hu, X. (2024). Preventing and detecting misinformation generated by large language models. In *Proc. of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3001–3004.
- Melo, P. F., Hoseini, M., Zannettou, S., and Benevenuto, F. (2024). Don't break the chain: Measuring message forwarding on whatsapp. In *Proc. of the Int'l AAI Conference on Weblogs and Social Media (ICWSM)*, pages 1054–1067.
- Reis, J. C., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, 34(2):76–81.
- Reis, J. C., Melo, P., Silva, M., and Benevenuto, F. (2023). Desinformação em plataformas digitais: Conceitos, abordagens tecnológicas e desafios. *Congresso da Sociedade Brasileira de Computação (CSBC). Jornada de Atualização em Informática (JAI)*.
- Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36.
- Suarez-Lledo, V. and Alvarez-Galvez, J. (2021). Prevalence of health misinformation on social media: systematic review. *Journal of medical Internet research*, 23(1):e17187.
- Swire-Thompson, B., Lazer, D., et al. (2020). Public health and online misinformation: challenges and recommendations. *Annu Rev Public Health*, 41(1):433–451.
- Treen, K. M. d., Williams, H. T., and O'Neill, S. J. (2020). Online misinformation about climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 11(5):e665.
- Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.
- Wu, J., Guo, J., and Hooi, B. (2024). Fake news in sheep's clothing: Robust fake news detection against llm-empowered style attacks. In *Proc. of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 3367–3378.