

Ferramentas de IA para identificação de Fake News: desinformação dos conteúdos publicados sobre eventos climáticos extremos¹

Aline de Oliveira Silva²

Dirceu Matheus Junior³

Resumo expandido

No intervalo de uma década, o Brasil foi impactado por 36.522 eventos extremos da natureza (2013 a 2023), divididos nas seguintes categorias: geológica, hidrológica, meteorológica e climatológica. As informações estão disponíveis no painel do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT), da Fiocruz.

Ao realizar uma breve análise nos números disponibilizados na plataforma da instituição, é possível verificar o crescimento contínuo das ocorrências de mudanças climáticas, que culminaram na tragédia vivenciada pelos moradores do Rio Grande do Sul, entre os meses de abril e maio de 2024.

De acordo com a nota técnica nº 02 (CGDTI) divulgada pelo Instituto de Pesquisas Aplicadas (IPEA), entre os meses de abril e maio, a sequência de chuvas torrenciais provocou uma série de enchentes e deslizamentos de terra ou lama, em 418 municípios do estado, localizado na região Sul do país.

O estudo acrescentou que, pelo menos 876,2 mil pessoas, ou seja, 8,8% população foram diretamente impactadas com a perda de familiares, moradia e todos os pertences materiais de subsistência. A situação foi ainda mais grave para 9,7% de famílias (138,8 mil), em situação de vulnerabilidade socioeconômica.

¹ Trabalho apresentado no C4 - Jornalismo de dados, ética da informação, fake news e crise dos pontos de vista centrais do XVII Simpósio Nacional da ABCiber – Associação Brasileira de Pesquisadores em Ciberultura. Universidade do Estado de Santa Catarina - UDESC, realizado nos dias 04 a 06 de dezembro de 2024.

² Mestra em Comunicação pela Universidade Federal de Mato Grosso do Sul (UFMS), professora convidada da Universidade Presbiteriana Mackenzie. 4898489@mackenzie.terceiros.br.

³ Mestre em Engenharia Elétrica e Computação e Ciência da Computação, Universidade Presbiteriana Mackenzie, dirceu.matheus@mackenzie.br.

Diante do cenário apresentado, o presente artigo tem o objetivo de propor a reflexão sobre as seguintes questões: qual a motivação dos usuários para produzirem ou compartilharem materiais falsos nas redes sociais, demonstrando nenhum respeito e empatia com as vítimas? A quem interessa provocar reações, como: medo, insegurança e incitação de discursos de ódio? Quais as ferramentas de Inteligência Artificial Generativa mais usadas na produção de: imagens, vídeos ou réplicas de portais oficiais de empresas e instituições públicas?

A fim de contextualizar a gravidade e a imoralidade nas produções de conteúdos falsos, popularizados pelo termo inglês “Fake News”, serão analisadas 31 matérias do jornal O Estado de São Paulo. A iniciativa chamada de “Estadão Verifica” está em atividade desde o início de 2019 e segue o código de princípios estabelecidos pela IFCN (International Fact Checking Network ou Aliança Internacional de Checagem de Fatos).

Com esse entendimento, o veículo de comunicação atende os compromissos firmados pelo grupo internacional de mídia, trabalhando no sentido de manter as seguintes condutas: apartidarismo, imparcialidade, transparência em relação as fontes, financiamento e tecnologia, além de estabelecer uma política de correções justa e honesta.

Do mesmo modo, serão analisadas iniciativas de combate à desinformação, desenvolvidas em projetos das seguintes organizações: redes sociais (Meta), OSCs, universidades e pesquisadores. O grupo citado trabalha há pelo menos cinco anos (desde 2019), com intuito de elaborar tecnologias que deem conta de: desmascarar, coibir e educar digitalmente a população, sobre a importância de não compartilhar materiais de origem duvidosa.

Vale acrescentar que a ferramenta digital empregada para apresentar os conteúdos falsos e os esclarecimentos do jornal é um recurso do Google chamado *Fact Check Tools*. A plataforma utiliza algoritmos treinados por Inteligência Artificial (IA), de modo a realizar as seguintes funções: localizar, identificar e publicizar para a imprensa nacional, conteúdos falsos, produzidos com intuito de prejudicar grupos, governos ou pessoas.

A fim de ampliar o olhar sobre as tecnologias baseadas em IA - capazes de auxiliar a sociedade a conhecer e evitar práticas com características de desinformação - serão detalhadas algumas iniciativas em funcionamento nos continentes europeu, asiático e norte-americanos. Os

sistemas variam de dispositivos simples até estruturas mais sofisticadas, todas idealizadas com o mesmo objetivo: combater o compartilhamento de fake news.

Os referenciais bibliográficos utilizados para embasar o artigo começam pela Teoria da Informação, que tiveram como idealizadores, Claude Shannon e Warren Weaver, em 1949. Na análise de Araújo (2015), os autores afirmam que a comunicação é entendida como o processo de transmissão de uma mensagem, efetuada por uma fonte de informação, que se utiliza de um canal para levar o conteúdo até o destinatário.

Desse modo, a problemática da desinformação está diretamente ligada a duas questões identificadas na comunicação, a simplificação em discordância com a complexidade das mensagens. Como bem exemplifica Araújo (2015, p.121): “a noção de informação ligada à incerteza, probabilidade, grau de liberdade e escolha das mensagens”.

Vale destacar que a existência da desinformação no contexto empírico e científico não se trata de um fenômeno recente. No entanto, após o surgimento e popularização da internet (a partir da década de 1990), as tecnologias criadas foram se aperfeiçoando e impulsionando boatos mentirosos e prejudiciais em diversos aspectos da sociedade.

Outro termo que caminha junto com a desinformação é o termo fake news (Vilela et al, 2022), apontado para se referir a um material (textos, vídeos, áudios e imagens), alterado do modelo original, com objetivo de enganar, falsificar ou causar confusão sobre determinada situação ou condição. Vale acrescentar que se trata de um formato específico de desinformação, assim como o boato e o spam, empregados com intuito de prejudicar gestões públicas, privadas e ainda, personalidades formadoras de opinião.

Partindo para o enfoque dos processos de digitalização, surgidos nas últimas décadas, Felice, Torres e Yanase (2012) argumentam que o ambiente virtual pode ser analisado como um conjunto de sistemas, dotados de particularidades que necessitam de uma conexão interligada para funcionarem. Nesse caso, por intermédio da ação proativa dos usuários/internautas que interagem com os materiais postados e compartilhados nas redes.

É a partir desse entendimento, que se pode explicar o fenômeno da viralização de conteúdos, de modo a tornar extremamente difícil, o processo de monitoramento sobre o que é postado, comentado e compartilhado nas redes sociais.

“Tais propriedades tendem a influenciar o funcionamento de todos os elementos de um sistema, mas são imprevisíveis, dado que não dependem do observador, somente da interação entre os elementos constituintes do todo (...). A impossibilidade de obter uma representação total e visual dos conjuntos de rede impõem um limite à sua compreensão e ao seu estudo, questionando-nos sobre as técnicas apropriadas para a realização de uma abordagem compreensiva que não seja nem externa e nem frontal (Felice, Torres e Yanase, 2015, p.76-77)”.

De modo a explicar como as ferramentas dotadas de IA podem contribuir no combate à desinformação, principalmente em situações de desastres climáticos, foi realizada uma pesquisa com a ferramenta *Fact Check Tools*, utilizando duas palavras-chaves: “enchentes no Rio Grande do Sul” e “Vítimas das enchentes no Rio Grande do Sul”. O sistema criado pelo Google identificou um total de 31 matérias envolvendo fake news nas redes sociais: Facebook, Instagram e Tik Tok.

É importante esclarecer que a Big Tech tem uma parceria firmada com o jornal O Estado de São Paulo, que recebe os materiais, averigua e produz matérias jornalísticas, desmentindo produtos, como: vídeos, fotos-montagens e imagens produzidas a partir de ferramentas de IA. No entanto, o projeto se estende a todos veículos interessados em combater situações que envolvam desinformação.

A plataforma é altamente intuitiva e tem como princípio de funcionamento, as seguintes etapas: uso de palavras-chaves na realização da pesquisa, listagem dos materiais identificados (dos mais recentes para os mais antigos), identificação da organização responsável pela verificação com hiperlink original e ainda, atualização constante dos textos armazenados no sistema.

É importante ressaltar uma informação apontada no início do texto: que algumas fake news são produzidas a partir de ferramentas dotadas de tecnologias da Inteligência Artificial. Essa situação causa insegurança e desconfiança por parte da sociedade, já que por outro lado, a IA também demonstra ser uma forte aliada na identificação de conteúdo duvidoso.

Por esse motivo, outro objetivo do trabalho será pesquisar com mais detalhes e trazer à público, ferramentas de IA que contribuem na descoberta de materiais elaborados para promover a desinformação entre a sociedade. A precisão e agilidade dos dispositivos demonstram que o alinhamento da IA com o Aprendizado de Máquina (*Machine Learning*), são fundamentais para coibir a proliferação de conteúdos, assim que forem identificados e comprovados como falsos.

É fato que a rapidez na elaboração de tecnologias provenientes da IA dificulta o acompanhamento e avaliação sobre os efeitos provocados pelas fake news. Contudo, algumas iniciativas, provenientes de outros países, podem ser aplicadas, por meio de parcerias empresariais ou institucionais, no sentido de atenderem regiões com mais dificuldade de lidar com a viralização de informações falsas.

A título de exemplificação, pode-se citar o projeto C2PA (*Coalition for Content Provenance and Authenticity*), que em tradução livre significa Coalizão para Proveniência e Autenticidade de Conteúdo. A iniciativa é formada pela parceria das seguintes empresas: Adobe, Arm, Intel, Microsoft e TruePic, que desenvolveram um guia de padrões técnicos para certificar a fonte, histórico, procedência e conteúdo da mídia avaliada. Na página principal do portal é informado que o C2PA é: “um padrão técnico aberto, que tem o objetivo de oferecer aos editores, criadores e consumidores, a capacidade de rastrear a origem de diferentes tipos de mídia”.

Do mesmo modo, serão apresentadas outras tecnologias baseadas em IA, desenvolvidas para identificar diferentes tipos de desinformação, entre os quais: *GPTZero* (identifica plágio e conteúdos criados por chatbots), *Factive AI Editor* (projetado para analisar se uma declaração é verdadeira ou não); *Connectively* (antigo *Haro*, promove a conexão entre jornalistas e fontes especializadas) e *TrollWal* (ferramenta alimentada por IA para proteger perfis de mídias sociais empresariais contra ataques de ódio).

Com esse entendimento, o artigo propõe-se a demonstrar, por intermédio de um estudo de caso, como é possível empregar uma plataforma, cujo funcionamento é dotado de tecnologia de IA para combater materiais produzidos com outros dispositivos provenientes da mesma base tecnológica. Além da reflexão sobre o modo como são trabalhados os dados na produção de notícias jornalísticas, o estudo deseja comprovar que a área de comunicação tem muito a ganhar, ao se debruçar sobre os estudos de tecnologia no combate à desinformação científica e empírica.

Palavras-chave

Inteligência Artificial; Desinformação; Fake News; Eventos Climáticos; Tecnologias Emergentes.

Referências bibliográficas

Araújo, Carlos Alberto. **Teorias da Comunicação: conceitos, escolas e tendências**. 15ª ed. Petrópolis, RJ: Vozes, 2015. ISBN: 978-85-326-2615-8

Associação para Sistemas de Informação (AIS). Exploring Potential of Generative AI in **Business Applications**. Disponível em: <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1002&context=isla2022>. Acesso em: 22 out. 2024.

BALBE, Antônio Luís; LOOSE, Fernando J. Jornalismo, medo e alterações climáticas: análise dos efeitos da comunicação de risco em Portugal. **Repositório da Universidade do Minho**, 2020. Disponível em: https://repositorium.sdum.uminho.pt/bitstream/1822/75628/1/2020_Balbe_Loose_Jornalismo%2c%20me%20e%20alterac%cc%a7o%cc%83es%20clima%cc%81ticas.pdf. Acesso em: 22 out. 2024.

COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO (CETESB). **Eventos climáticos extremos no Brasil: tendências e projeções**. CETESB, 2014. Disponível em: https://cetesb.sp.gov.br/proclima/wp-content/uploads/sites/36/2014/05/mc_eventos_extremos_brasil.pdf. Acesso em: 22 out. 2024.

DI FELICE, M.; TORRES, J.C. e YANASE, L.K.H. **Redes Digitais e sustentabilidade: as interações com o meio ambiente na era da informação**. São Paulo, Anablume, 2012. ISBN: 978-85-391-0312-6

DIGITAL FUTURE SOCIETY. Desinformação e fake news: Impactos sociais e implicações para a governança. **Digital Future Society**, 2024. Disponível em: <https://digitalfuturesociety.com/disinformation-fakenews/> Acesso em: 22 out. 2024.

FUNDAÇÃO OSWALDO CRUZ (Fiocruz). **Monitoramento de extremos climáticos e saúde pública**. Mapas Clima e Saúde, 2024. Disponível em: <https://mapas.climaesaude.icict.fiocruz.br/extremos> Acesso em: 22 out. 2024.

Google News Initiative. **Google Fact Check Tools**. Disponível em: <https://newsinitiative.withgoogle.com/pt-br/resources/trainings/google-fact-check-tools/> Acesso em: 22 out. 2024.

GONDIM, Gilberto Cardoso. Mudanças climáticas e seus impactos na saúde pública. **Repositório do Instituto de Pesquisa Econômica Aplicada - IPEA**, 2023. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/14337> Acesso em: 22 out. 2024.

International Journalists' Network (IJNet). 4 AI tools to help newsrooms avoid spreading harmful content. Disponível em: <https://ijnet.org/en/story/4-ai-tools-help-newsrooms-avoid-spreading-harmful-content>. Acesso em: 22 out. 2024.

MATHEUS, R. F. Rafael Capurro e a filosofia da informação: abordagens, conceitos e metodologias de pesquisa para a Ciência da informação. In: **Perspect.ciênc.inf.**, Belo Horizonte (MG), v.10 n.2, p. 140-165, jul./dez. 2005. Disponível em: <https://cip.brapci.inf.br/download/45886> Acesso em: 22 de outubro de 2024.

NAÇÕES UNIDAS. Número de eventos climáticos extremos dobra em 20 anos e aumenta pressão sobre adaptação. **ONU News**, 20 maio 2024. Disponível em: <https://news.un.org/pt/story/2024/05/1831366>. Acesso em: 22 out. 2024.

XU, B.; ZHANG, Y.; WU, J.; WANG, X.; YAO, Y. "Impact of Extreme Weather on IoT Networks and Climate Data Collection: A Review". **Electronics**, v. 12, n. 24, p. 5041, 2024. Disponível em: <https://www.mdpi.com/2079-9292/12/24/5041#B25-electronics-12-05041>. Acesso em: 22 out. 2024.