XVII SIMPÓSIO NACIONAL DA ABCIBER – Associação Brasileira de Pesquisadores em Cibercultura. Universidade do Estado de Santa Catarina. De 4 a 6 de desembro de 2024.

FátimaGPT do Aos Fatos: IA generativa no apoio ao combate à desinformação[1]

Francilene de Oliveira Silva[2]
Laura Rayssa de Andrade Cabral[3]
Rita de Cássia Romeiro Paulino[4]

Resumo expandido

Ao mesmo tempo em que a IA generativa tem potencial para a desinformação, seu uso também pode ser explorado para combatê-la. Jornalistas expressam preocupações sobre essa natureza dupla de equilibrar benefícios e vulnerabilidades (Peña-Fernández et al., 2023). Desde 2017, houve um ponto de virada nos usos e pesquisas da IA com o aperfeiçoamento dos modelos generativos, que operam por meio de Large Language Model e que, por modo probabilístico, podem gerar conteúdos novos a partir de uma imensa base de dados sob a qual foram treinados (Goldstein, 2023, p. 15).

Empresas que atuam com chatbots de IA generativa utilizam o conteúdo publicado na internet para treinamento de seus dispositivos. A Open AI (2023), criadora do ChatGPT, defende que utilizar materiais da internet disponíveis ao público para treinar inteligência artificial se enquadra na doutrina de uso justo, porém essa prática está sendo questionada na justiça por instituições jornalísticas e órgãos de defesa de direito autoral.

---

[1] Trabalho apresentado no Painel Temático Estratégias Comunicacionais em Eventos Climáticos Extremos do XVII Simpósio Nacional da ABCiber – Associação Brasileira de Pesquisadores em Cibercultura. Universidade do Estado de Santa Catarina - UDESC, realizado nos dias 4 a 06 de dezembro de 2024.

[2] Doutoranda no Programa de Pós-graduação em Jornalismo da Universidade Federal de Santa Catarina (oliveirafrancilene@gmail.com).

[3] Doutoranda no Programa de Pós-graduação em Jornalismo da Universidade Federal de Santa Catarina (laurandradec@gmail.com)

[4] Professora no Programa de Pós-graduação em Jornalismo da Universidade Federal de Santa Catarina (rcpauli@gmail.com).

Um agravante para o jornalismo é que estes modelos de linguagem natural podem gerar texto sem sentido ou infiel ao dado fornecido como fonte de entrada, ao que os pesquisadores começaram a se referir como "alucinação" (Ziwei Ji et al, 2023). Por consequência, seu uso pode propagar informações falsas inadvertidamente.

No entanto, desde o lançamento do ChatGPT, em novembro de 2022, a corrida tecnológica entre empresas como Google, Microsoft, Meta e OpenAI se acirrou e a tecnologia tem sido incorporada nos produtos destas empresas, muitas vezes, sem o debate devido com a sociedade e órgãos reguladores provocando mudanças nos ecossistemas de informação.

Para Pérez-Seijo e Vicente (2022), o jornalismo digital passou por mudanças imensuráveis na última década levando a uma reconfiguração dos processos de produção, distribuição e consumo de notícias com o jornalismo se tornando mais automatizado, personalizado e imersivo por meio do uso de dispositivos tecnológicos que trouxeram oportunidades, mas também muitos desafios. Para Ioscote et al (2024), uma organização de notícias que se recuse a usar a IA terá dificuldades para permanecer competitiva, mas é essencial considerar os valores jornalísticos como a transparência e a veracidade.

A IA generativa está sendo incorporada no Jornalismo com maiores impactos na fase de produção de notícias, pois pode produzir conteúdo textual, sonoro, audiovisual, gráficos e infográficos usando técnicas de deep learning. Uma das preocupações da área é o uso da tecnologia para propagar notícias falsas.

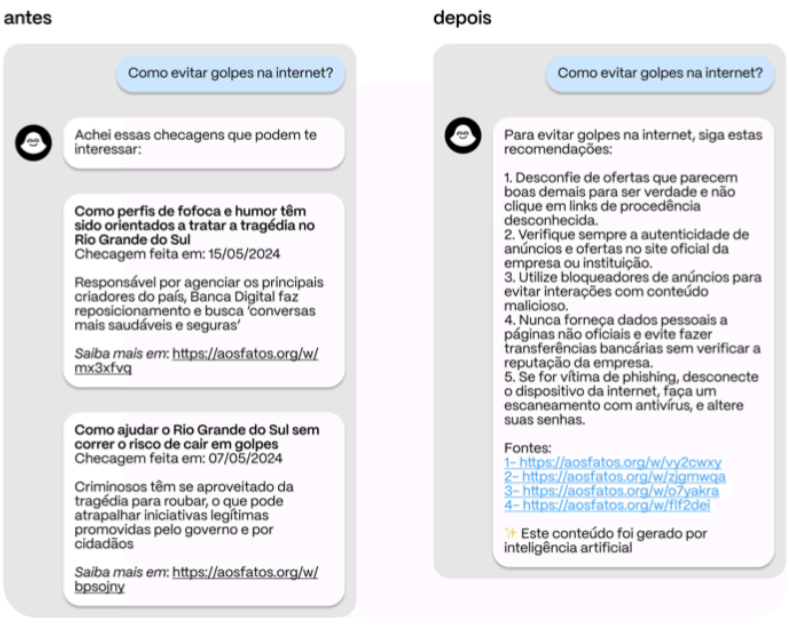
As fake news são "mensagens falsas que parecem verdadeiras, cuja distribuição segue uma intencionalidade específica de enganar e influenciar grupos específicos" (Oliveira, 2023, p. 52). Elas não são um problema recente, mas "o atual arranjo da cultura digital, marcada por plataformas, pela dataficação e por uma sociabilidade algorítmica potencializa a distribuição de conteúdos desinformativos" (Oliveira, 2023, p. 24) .

As notícias falsas são muito usadas como estratégias políticas. 2018, ano eleitoral no Brasil, foi marcado por desinformação nas plataformas de redes sociais. Neste mesmo ano, surgiram vários projetos para descobrir, investigar e desmascarar conteúdos suspeitos como o Comprova, criado pela Associação Brasileira de Jornalismo Investigativo (Abraji), o sistema

web Monitor de WhatsApp, o Estadão Verifica, do jornal O Estado de S. Paulo e o Fato ou Fake, do Grupo Globo.

As eleições brasileiras de 2022 proporcionaram um ambiente desinformativo ainda mais complexo do que em 2018. No entanto, a IA foi utilizada igualmente para detectar, comparar e verificar fake news, numa tentativa de equilibrar as forças e mostrando que os bots podem ser eficientes para capturar, verificar, comparar e informar com veracidade o conteúdo previamente manipulado (Welter; Canavilhas, 2023).

Ao passo em que o processo de educação para a mídia caminha vagarosamente, as agências de fact-checking ainda são as melhores opções no combate à desinformação (Welter; Canavilhas, 2023). Neste sentido, este trabalho tem como objeto de estudo a agência de checagem Aos Fatos, em especial, a chatbot FátimaGPT, criado em 2018, e que incorporou, em setembro de 2024, a tecnologia IA generativa para dar respostas mais personalizadas e naturais no WhatsApp, Telegram e no site do Aos Fatos, contribuindo no combate à desinformação. O objetivo é entender seu uso e potencial no combate à desinformação.



Fonte: Aos Fatos (2024b)

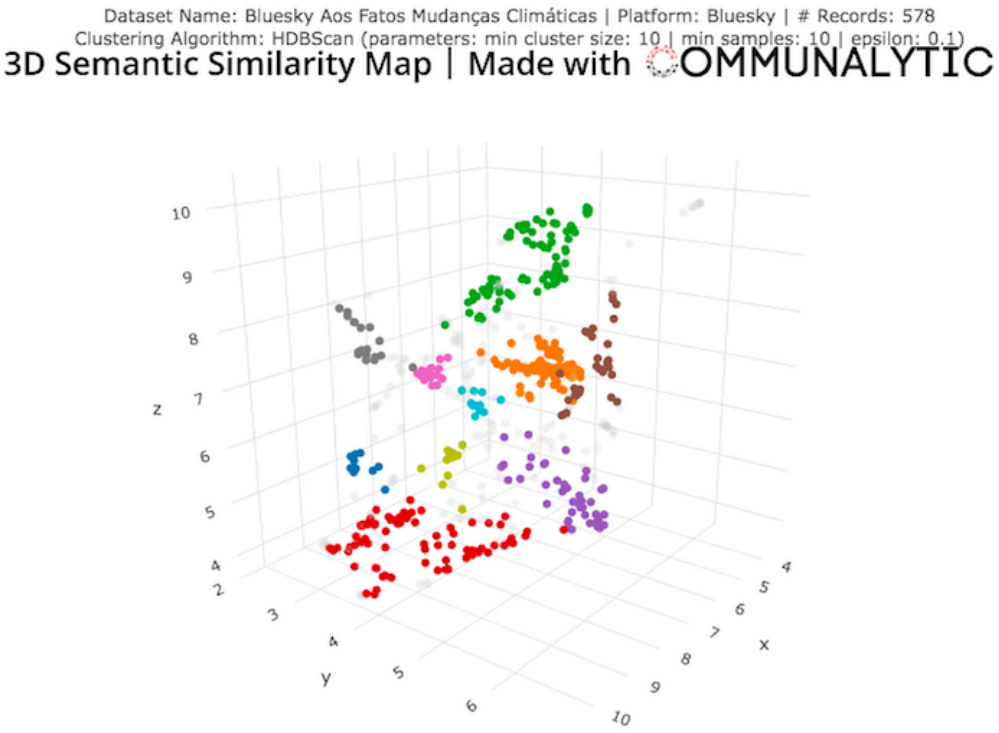
Para minimizar os efeitos das alucinações presentes neste modelo, a Fátima foi treinada apenas com conteúdo publicado no site (Aos Fatos, 2024a, 2024b). Um pouco antes do lançamento do bot, em julho de 2024, o Aos Fatos lançou sua política de uso de IA. Entre as orientações está que a empresa não usa a tecnologia para criar conteúdo original sem supervisão humana, mas que pode ser usada para adaptar textos escritos, editados e publicados por jornalistas para novos formatos ou linguagens, como resumir reportagens, fazer traduções e criar respostas para a Fátima (Aos Fatos, 2024a).

Temas predominantes na checagem do Aos fatos

Para realizar este estudo, sentiu-se a necessidade de entender os principais temas desinformativos checados pelo Aos Fatos. Adotou-se, portanto, uma abordagem metodológica que combina técnicas de métodos digitais e análise com base nos gráficos gerados. A base de dados utilizada foi a da Bluesky no período de 05 de julho de 2023, o primeiro registro da organização na rede social, até 26 de outubro de 2024. No total, foram 578 registros, sendo que oito postagens foram excluídas porque eram muito curtas (menos de três palavras) para capturar seu significado semântico ou eram duplicadas. Uma informação importante em relação aos dados é que entre outubro de 2023 a agosto de 2024, o Aos Fatos não postou no Bluesky. O retorno aconteceu em setembro de 2024, após a suspensão do Twitter pelo Supremo Tribunal Federal (STF) entre 31 de agosto a 07 de outubro.

Para a extração e análise dos temas checados mais recorrentes na empresa jornalística usamos a ferramenta Communalytic, desenvolvida pelo Social Media Lab, da Toronto Metropolitan University. Na ferramenta foi utilizado o módulo Análise de Tópicos que transforma dados textuais legíveis por humanos (como postagens de mídia social) em embeddings (vetores de números legíveis por computador). Uma vez transformados, os posts são agrupados com base em sua similaridade semântica e visualizados por meio de um mapa 3D interativo. A partir do mapa, pode-se descobrir tópicos e temas latentes em um conjunto de dados. A ferramenta classificou os registros em 10 clusters nomeados pelas autoras e os

"Outliers" tópicos que não se agruparam em clusters. Nem todos os assuntos estão totalmente relacionados. Os temas mais relacionais estão mais próximos. Quanto maior a distância, maior a probabilidade de, apesar de pertencer ao tema, versar sobre um tópico diferente. Os cluster identificados foram:

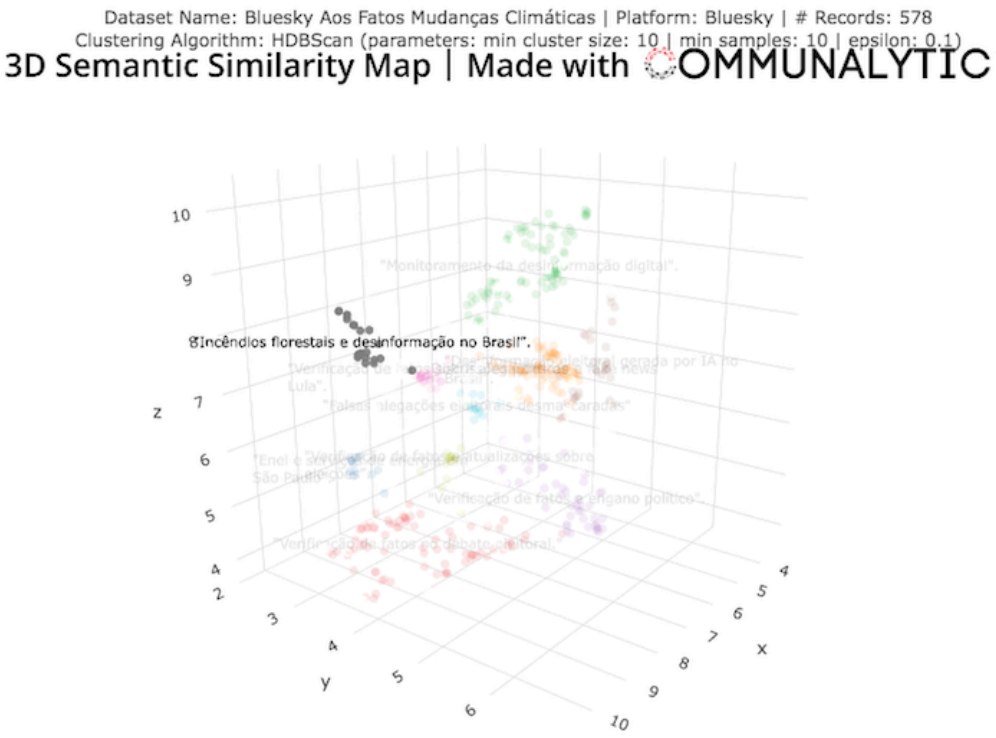


Clusters	**Registros**
1- Discussões políticas e fake news	83
2- Monitoramento da desinformação digital	81

3- Verificação de fatos no debate eleitoral	79
4- Verificação de fatos e engano político	44
5- Desinformação eleitoral gerada por IA no Brasil	32
6- Verificação de fatos sobre alegações de Lula	25
7- Incêndios florestais e desinformação no Brasil	23
8- Verificação de fatos e atualizações sobre eleições	15
9- Falsas alegações eleitorais desmascaradas	12
10- Enel e serviços de energia em São Paulo	11
Outliers	173

Tabela: As autoras (2024)

A etapa final de resultados envolveu a interpretação dos gráficos gerados e a discussão dos insights obtidos. Percebe-se o foco dos registros do Aos Fatos em assuntos relacionados às desinformações sobre eleições ou discursos políticos. Do total de 578 registros, apenas 23 (3,9%) formam um cluster relacionado a eventos climáticos versando sobre as queimadas no Brasil. Outro insight é que do total, 29,9% são "outliers" ou tópicos sobre assuntos diversos, que não se agrupam em clusters.



Cluster sobre queimadas no Brasil em relação a outros clusters

Todo o conteúdo checado faz parte da base de treinamento da FátimaGPT para dar respostas melhores aos usuários. Percebe-se que a maior parte das checagens estão relacionadas às desinformações eleitorais e enganos políticos, mas que assuntos relacionados às mudanças climáticas também estão presentes sob forma de clusters.

Palavras-chave

Inteligência Artificial Generativa; desinformação; jornalismo; eventos climáticos; tecnologia

Referências

AOS FATOS. Política de uso de inteligência artificial do Aos Fatos. 7 de jul. 2024a. Disponível em: https://www.aosfatos.org/politica-ia/.

AOS FATOS. Aos Fatos lança Fátima 3.0, expansão do chatbot com IA generativa. 5 de set. 2024b. Disponível em: https://www.aosfatos.org/noticias/aos-fatos-lanca-fatima-30-expansao-do-chatbot-com-ia-generativa/.

GRUZD, A., & MAI, P. . Communalytic: A computational social science research tool for studying online communities and discourse, 2024. Disponível em: https://Communalytic.org.

IOSCOTE, F.; GONÇALVES, A.; QUADROS, C. Artificial Intelligence in Journalism: A Ten-Year Retrospective of Scientific Articles (2014–2023). Journal. Media 2024, 5, 873-891, 2024. DOI: https://doi.org/10.3390/journalmedia5030056.

OLIVEIRA, Frederico. As fake news e a produção jornalística de referências. 2023. 382f. Tese (Doutorado em Comunicação e Cultura Contemporâneas) – Universidade Federal da Bahia, Salvador, 2023. Disponível em: https://repositorio.ufba.br/handle/ri/37610

GOLDSTEIN, Josh A, et al. Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations, arxiv.org, 2023. Disponível em: https://arxiv.org/abs/2301.04246

PEÑA-FERNÁNDEZ, Simón et al . El discurso de los periodistas sobre el impacto de la inteligencia artificial generativa en la desinformación. Estudios sobre el Mensaje Periodístico, 29(4), 833-841, 2023. DOI: https://doi.org/10.5209/esmp.88673

PÉREZ-SEIJO, Sara; VICENTE, Paulo Nuno. After the hype: How hi-tech is reshaping journalism. In Total Journalism: Models, Techniques and Challenges. Cham: Springer International Publishing, pp. 41–52, 2022. DOI: https://doi.org/10.1007/978-3-030-88028-6_4

OPEN AI. OpenAI and journalism. 8 de jan. 2024. Disponível em: https://openai.com/index/openai-and-journalism/.

WELTER, Lahis; CANAVILHAS, João. La inteligencia artificial en la lucha contra la desinformación en las presidenciales brasileñas 2022: estudio de caso con Lupa e o Aos Fatos. , en Miguel Hernández Communication Journal, Vol. 14 (2), pp. 409 a 426. Universidad Miguel Hernández, UMH (Elche-Alicante), 2023. DOI: https://10.21134/mhjournal.v14i.1984

ZIWEI JI ET AL. Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*. 55 (12), 1-38, 2023. DOI: https://doi.org/10.1145/3571730