

## Desafios das Inteligências Artificiais Generativas para as Práticas de Combate à Desinformação das Plataformas de *Fact-checking*<sup>1</sup>

Leonardo Pereira TAVARES<sup>2</sup>

Edinaldo Gabriel Alves BARRETO<sup>2</sup>

Fellipe Sá BRASILEIRO<sup>3</sup>

Universidade Federal da Paraíba, João Pessoa, PB

### RESUMO

Este estudo investiga os desafios das Inteligências Artificiais Generativas para as práticas de combate à desinformação das plataformas de *fact-checking*. Foi realizada uma análise de conteúdo das postagens de uma plataforma de *fact-checking*, relacionadas à desinformação, durante o período de seis meses após o lançamento do ChatGPT. Os resultados revelam a crescente utilização de IAG no processo de criação de conteúdos falsos, além dos desafios enfrentados pelas plataformas como: dificuldade de verificar informações geradas por IAG, falta de transparência dos dados e conteúdos visualmente convincentes.

**PALAVRAS-CHAVE:** Desinformação; Inteligência Artificial; *Fact-checking*.

### INTRODUÇÃO

O conjunto das emergentes tecnologias digitais, que inclui dispositivos e plataformas digitais, possibilitou um ecossistema de hiperconexões a reboque da sociedade em rede (Castells; Espanha, 2007). Segundo Ripoll e Matos (2020), a sociedade tem passado por um processo de migração dos diversos aspectos da vida humana, desde a sua forma de se comunicar até os novos formatos de trabalho, para os ecossistemas digitais, que, dentre os seus diferentes desafios, apresentam a “hiperinformação” caracterizada por Moretzsohn (2017) como o motor da alienação.

Nesse cenário, nos últimos anos, é notório uma desordem informacional agravada pelo rápido acesso às informações de origens distintas, que podem ser decorrentes dos fatos ou, problematicamente, criadas a fim de manipular e desinformar. Nesse contexto, em 2022, termos como Inteligência Artificial Generativa (IAG) alcançaram uma maior notoriedade nos meios de comunicação. Embora já se falasse sobre as IAG, e existisse algumas outras ferramentas, foi com a ascensão do ChatGPT (*Generative Pre-trained*

---

<sup>1</sup> Trabalho apresentado no Grupo de Trabalho (Desinformação, educação midiática e *fact-checking*), evento integrante da programação do 24º Congresso de Ciências da Comunicação na Região Nordeste, realizado de 8 a 10 de maio de 2024.

<sup>2</sup> Mestrando do Programa de Pós-graduação em Comunicação. Universidade Federal da Paraíba.

<sup>3</sup> Doutor em Ciência da Informação Professor do Programa de Pós-Graduação em Comunicação e do Programa de Pós-Graduação em Ciência da Informação. Universidade Federal da Paraíba.

*Transformer*),<sup>4</sup> lançado em 30 de novembro de 2022, com seus incríveis 100 milhões de usuários ativos em apenas um mês, que a sociedade conseguiu na prática entender o poder das ferramentas IAG (Donato; Escada; Villanueva, 2023).

Esse novo modelo de criar conteúdo através de IAG tem preocupado pesquisadores ao redor do globo (Essenfelder; Araújo; César, 2023; Khan *et al.*, 2023; Oliveira *et al.*, 2023), principalmente por abranger uma variedade de linguagens, desde a escrita de textos até criação de vídeos e áudios. A problemática que norteia este trabalho é sobre a desinformação e como ela pode ser intensificada através desses meios. Diante disso, este estudo busca investigar os desafios das IAGs para as práticas das plataformas de *fact-checking*, uma vez que, segundo Marques e Laipelt (2023), modelos como ChatGPT são capazes de desenvolver narrativas com credibilidade.

Portanto, este trabalho, por meio da análise de conteúdo das postagens de uma plataforma de *fact-checking*, pretende apresentar e refletir sobre a repercussão das IAGs durante os seis primeiros meses após o lançamento do ChatGPT. Ao examinar os alertas e as observações feitas pelas plataformas de checagem dos fatos, buscamos identificar os padrões e as tendências relacionadas à disseminação de desinformação por IAG.

## **METODOLOGIA**

O processo metodológico partiu de uma análise de conteúdo quali-quantitativa a fim de entender a repercussão das IAGs durante os seis primeiros meses após o lançamento do ChatGPT (30 de novembro de 2022). A primeira etapa foi selecionar a plataforma de *fact-checking*. A etapa seguinte foi realizar a coleta de dados através do mecanismo de busca da plataforma com o termo “Inteligência Artificial”; após a busca, foram selecionadas postagens durante o período de 30 de novembro de 2022 a 30 de maio de 2023, com o objetivo de compreender o período de destaque das IAGs. Todas as postagens da agência foram consideradas, incluindo entrevistas, checagens de fatos, acontecimentos, informativos e artigos. A etapa três focou na análise de dados a fim de identificar desafios enfrentados pelas plataformas de *fact-checking* relacionados às práticas de detecção de desinformação manipulada por IAG.

---

<sup>4</sup> Tradução livre para “Transformador pré-treinado generativo” o ChatGPT é uma ferramenta de conversação de texto entre o usuário e a IA, geralmente utilizado para criar textos.

## RESULTADOS E DISCUSSÃO

A primeira etapa analisou de modo exploratório três plataformas de *fact-checking*. Entretanto, por apresentar facilidades quanto ao uso de sua plataforma e, ainda, a possibilidade de delimitação do período pré-estabelecido para a pesquisa, a Agência Lupa foi selecionada para a extração dos dados. Além disso, a plataforma foi a primeira agência especializada em checagem do Brasil (Dourado; Alencar, 2020).

A segunda etapa analisou quinze postagens dentro do período de análise. Das quinze postagens, cinco não estavam diretamente relacionadas às IAGs ou continham conexões com postagens anteriores sobre o assunto, sendo consideradas irrelevantes para a análise. Com a coleta realizada, observou-se que nos meses de dezembro e janeiro não houveram postagens relacionadas às IAGs, conforme apresenta a Tabela 1.

Tabela 1 – Postagens Lupa

Nº	TÍTULO DAS POSTAGENS	MÊS
1	ChatGPT: Na volta às aulas, experimentação precisa ser o caminho, defendem especialistas	Fev
2	Golpe do cartão de crédito usa reportagens falsas para atrair vítimas no Facebook	Fev
3	No Facebook, ‘fakes’ usam conteúdos pagos para atrair vítimas para fraude	Fev
4	Em teste, ChatGPT falha em duas a cada cinco respostas sobre posts falsos	Mar
5	Vídeo de Bill Gates irritado com perguntas sobre vacinas é 'deep fake'	Mar
6	É falsa a foto do Papa Francisco usando um estiloso casaco branco	Mar
7	Imagem falsa do papa mostra como estamos despreparados para o fascínio do visual	Mar
8	Deepfakes são história: republicanos criam vídeo com IA para atacar Biden	Abr
9	É falsa imagem que mostra explosão ao lado do Pentágono, nos EUA	Mai
10	Falso curso promete porte de arma sem restrições a qualquer cidadão	Mai

Fonte: Agência Lupa (2024)

Para iniciar a análise, consideramos uma postagem de cada mês, trazendo os dados e discussões mais importantes. Ao final, fizemos uma consideração geral sobre as postagens. A primeira postagem do ano da Agência Lupa sobre IAG aconteceu em fevereiro, sendo um artigo que aborda a fala de especialistas sobre o assunto, afirmando que embora as plataformas IAGs (como o ChatGPT) tenham gerado preocupação, existem potencial tecnológico para se aprender com elas e entender um pouco mais, antes de querer demonizá-las. Também relata que o ChatGPT tem potencialidade de aumentar a produção de conteúdos, tanto para informar quanto desinformar. O professor Diego Cortiz ainda comenta “o ChatGPT dá uma resposta, mas não te fala a fonte[...] Para fazer a checagem de fatos nesse caso fica ainda mais difícil” (Terra, 2023).

Em março, a primeira postagem considerou a discussão atual sobre IAG. A agência conduziu um experimento que utilizou o ChatGPT para avaliar se as respostas geradas pela IAG contribuem para a propagação de desinformação. Os resultados mostraram que 40% das respostas em relação às postagens enganosas careciam de esclarecimentos, eram imprecisas ou incorretas. Também notaram que em alguns casos, informações desatualizadas foram apresentadas como sendo atuais (Soares, 2023).

Em março também houve uma repercussão das IAGs incluindo a imagem do Papa Francisco vestido com um casaco branco de grife. Tal desinformação afirmava que o Papa teria contratado um dos melhores estilistas para criação de suas novas roupas. Nessa questão, Paulo Knauss, historiador e pesquisador, diz que é necessário que “seja preparado um olhar preparado para analisar as imagens”. Seguindo o mesmo raciocínio, a pesquisadora Ana Mauad diz que “não importa se a imagem mente, mas, sim, o porquê e como se deu tal mentira” (Kapa, 2023). Peter Reichel reitera a discussão ao dizer que a imagem causa fascínio:

Algumas imagens são tão impactantes que criam uma negociação de sentido com quem as vê. Em outras palavras, queremos acreditar naquilo que está ali, já que o que vemos nos faz sentido e, claro, se é uma imagem, temos a orientação de que provavelmente é verdadeira (Kapa, 2023).

Problemáticas como essas ressaltam a necessidade de conscientização sobre os perigos da desinformação visual propagada por meio de avançadas técnicas de IAGs. As postagens seguintes que finalizam a presente análise compartilham semelhanças notáveis em relação ao uso das IAGs para manipular informações e criar conteúdo enganoso. Em abril, a Agência Lupa fez uma verificação de veracidade em uma postagem em que republicanos<sup>5</sup> produziram um vídeo falso com IAG para atacar Joe Biden, presidente dos Estados Unidos (EUA), destacando a crescente utilização dessa tecnologia para influenciar campanhas políticas e desinformar o público.

Da mesma forma, uma postagem em maio faz a verificação de uma imagem falsa que retrata uma explosão ao lado do Pentágono nos EUA. Entretanto, foi identificada como sendo gerada por IAG, evidenciando a capacidade dessa tecnologia de criar

---

<sup>5</sup> Uma das duas principais forças políticas dos EUA, promovendo o conservadorismo fiscal, a livre iniciativa, valores tradicionais e uma política externa firme, em contraposição ao Partido Democrata.

conteúdos visualmente convincentes para propagação de desinformação. Esses exemplos ilustram como a IAG está sendo cada vez mais empregada para manipular e disseminar desinformação, representando um desafio significativo para as plataformas de *fact-checking* e para a sociedade como um todo.

## CONCLUSÃO

A variedade de tópicos abordados sobre as IAGs e o seu potencial de criar desinformação destacou a preocupação de especialistas que combatem a desinformação, como as plataformas de *fact-checking*. Principalmente, diante da dificuldade de verificar informações geradas por IAGs, desse modo, a falta de transparência das fontes e, muitas das vezes, a manipulação dos dados acaba sendo uma das questões mais desafiadoras.

Outro ponto é que, embora o trabalho tenha proposto uma análise a partir da repercussão do ChatGPT, é sabido que diversas outras ferramentas de IAGs têm auxiliado pessoas mal intencionadas na propagação de desinformação, principalmente as que utilizam de IAGs para a produção de imagens. Além disso, é notório que figuras públicas continuam sendo um dos alvos principais das máquinas de disseminação de desinformação, como o caso do presidente dos EUA, Joe Biden, e do Papa Francisco.

Diante desse cenário, é essencial continuar investigando e desenvolvendo estratégias eficazes para mitigar os impactos negativos da desinformação gerada por IAG a fim de proteger a integridade da informação online. Como relatado por Kapa (2023), o uso das plataformas de IAGs está destinado a crescer em diversas aplicações, entretanto, manter uma regulamentação e um letramento para os aspectos desinformativos é essencial à coexistência humana frente às novas modalidades de criação de conteúdo.

## AGRADECIMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

## REFERÊNCIAS

AGÊNCIA LUPA. **Verificações**. 2024. Disponível em: <https://lupa.uol.com.br/>. Acesso em: 20 mar. 2024.

CASTELLS, Manuel; ESPANHA, Rita. **A era da informação: economia, sociedade e cultura**. [S. l.]: Fundação Calouste Gulbenkian. Serviço de Educação e Bolsas, 2007.

DONATO, Helena; ESCADA, Pedro; VILLANUEVA, Tiago. A transparência da ciência com o ChatGPT e as ferramentas emergentes de inteligência artificial: como se devem posicionar as revistas científicas médicas. **The Transparency of Science with ChatGpt and the Emerging Artificial Intelligence Language Models: Where Should Medical Journals Stand**, [s. l.], 2023.

DOURADO, Jacqueline Lima; ALENCAR, Marta Thaís. Agência Lupa: fact-checking como modelo de negócio na Internet. **Comunicação & Inovação**, [s. l.], v. 21, n. 46, 2020.

ESSENFELDER, Renato; ARAÚJO, Daniel Ladeira de; CÉSAR, Timóteo Camargo. Da forma ética à ação ética: riscos na adoção da inteligência artificial em sala de aula. **Educação e tecnologias: formação docente, inteligência artificial e humanismo em tempos de desafios**, [s. l.], n. 1, p. 113–130, 2023.

KAPA, Raphael. **Imagem falsa do papa mostra como estamos despreparados para o fascínio do visual**. [S. l.], 2023. Disponível em: <https://lupa.uol.com.br/educacao/2023/03/29/imagem-falsa-do-papa-mostra-como-estamos-despreparados-para-o-fascinio-do-visual>. Acesso em: 21 mar. 2024.

KHAN, Bangul *et al.* Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. **Biomedical Materials & Devices**, [s. l.], v. 1, n. 2, p. 731–738, 2023.

MARQUES, Simone Dias; LAIPELT, Rita do Carmo Ferreira. Pós-realidade e Teoria da Desinformação: inquietações sobre o uso massivo de IA Generativa. **Fórum de Estudos em Informação, Sociedade e Ciência (5.: 2023: Porto Alegre). Anais. Porto Alegre: UFRGS, PPGCIN**, [s. l.], 2023.

MORETZSOHN, Sylvia Debossan. “Uma legião de imbecis”: hiperinformação, alienação e o fetichismo da tecnologia libertária|“A legion of imbeciles”: hyperinformation, alienation, and the fetishism of libertarian technology. **Liinc em revista**, [s. l.], v. 13, n. 2, 2017.

OLIVEIRA, Laize Almeida de *et al.* Inteligência artificial na educação: uma revisão integrativa da literatura. **Peer Review**, [s. l.], v. 5, n. 24, p. 248–268, 2023.

RIPOLL, Leonardo; MATOS, José Claudio. O contexto informacional contemporâneo: o crescimento da desinformação e suas manifestações no ambiente digital. **Informação@Profissões**, [s. l.], v. 9, n. 1, p. 87–107, 2020.

SOARES, Gabriela. **Em teste, ChatGPT falha em duas a cada cinco respostas sobre posts falsos**. [S. l.], 2023. Disponível em: <https://lupa.uol.com.br/jornalismo/2023/03/06/em-teste-chatgpt-falha-em-duas-a-cada-cinco-respostas-sobre-posts-falsos>. Acesso em: 21 mar. 2024.

TERRA, Victor. **ChatGPT: Na volta às aulas, experimentação precisa ser o caminho, defendem especialistas**. [S. l.], 2023. Disponível em: <https://lupa.uol.com.br/educacao/2023/02/08/chatgpt-na-volta-as-aulas-experimentacao-precisa-ser-o-caminho-defendem-especialistas>. Acesso em: 21 mar. 2024.