

Essa informação é verdadeira ou não, Fátima? Uma análise da checagem automatizada da Aos Fatos sobre as enchentes no Rio Grande do Sul¹

Paulo Pessôa Neto²

Ivan Bomfim³

Universidade Estadual de Ponta Grossa - UEPG, Ponta Grossa, PR

RESUMO

O trabalho propõe uma testagem e análise do projeto de checagem automatizada da agência de fact-checking Aos Fatos em um momento de alta desordem informacional, comoção pública e rapidez na mudança de informações oficiais, como está sendo a situação das enchentes no Rio Grande do Sul em 2024. A robô verificadora da Aos Fatos, denominada FátimaGPT, está em período de testes com a implementação de tecnologia de inteligência artificial generativa. Os dados serão debatidos à luz dos conceitos sobre combate à desinformação, jornalismo automatizado e inteligência artificial.

PALAVRAS-CHAVE

Inteligência artificial generativa; jornalismo automatizado; fact-checking; combate à desinformação; enchentes no Rio Grande do Sul.

CONTEXTUALIZAÇÃO

Em dezembro de 2022, a empresa OpenAI lançou a versão 3.5 do chatbot ChatGPT (Santaella, 2023a), que passava a contar a tecnologia de inteligência artificial generativa (IAG). Desde então, várias áreas do conhecimento e do universo mercadológico estão em alta comoção para promover e lançar produtos que utilizam inteligência artificial (IA). Esse fenômeno, que Simon (2022) nomeia como Corrida de Ouro da IA, teve como característica de primeiro estágio o uso desregulamentado, criminal e de bastante fabricação de conteúdo enganoso, diz Santaella (2023a).

A conjuntura fez com que o Future of Life Institute (2023) formulasse uma carta pública pedindo a pausa nas testagens e comercialização com IA para que mais estudos pudessem ser realizados e, assim, fosse possível apontar melhores caminhos para a

¹ Trabalho apresentado no GP Tecnologias e Culturas Digitais, XIX Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 47º Congresso Brasileiro de Ciências da Comunicação.

² Mestrando Bolsista Capes do Programa de Pós-Graduação em Jornalismo da UEPG. Colaborador do Grupo de Estudo e Pesquisa em Mídias Digitais (GEMIDI), email: ivanbp17@gmail.com.

³ Professor Doutor do Programa de Pós-Graduação em Jornalismo da UEPG. Coordenador do Grupo de Estudo e Pesquisa em Mídias Digitais (GEMIDI), email: paulo.pterceiro@gmail.com.

utilização da ferramenta. A preocupação foi compartilhada por diversas áreas do conhecimento acadêmico e mercado profissional (Santaella, 2023b, p. 1). Para este trabalho, interessa a utilização da IA nas produções jornalísticas e os debates em torno da automatização nas redações de veículos de imprensa. Este processo, como indicam Pessoa e Bomfim (2023), foi acelerado devido a pressões de sindicatos de jornalistas ainda em 2023. Segundo Beckett e Yaseen (2023), 75% (de um total de 105) das empresas de mídia e veículos de imprensa ao redor do mundo já utilizam a IA em pelo menos uma das etapas de produção de notícias. De acordo com os autores, uma das quatro áreas que mais sofrerá impacto em muitas redações será a do fact-checking e verificação de desinformações.

É neste cenário que se encontra FátimaGPT, versão de teste da robô checadora da agência de *fact-checking* Aos Fatos. Lançada em 2020, em meio à desordem informacional que circulava sobre a pandemia da Covid-19, Fátima surge como projeto de jornalismo automatizado. A robô materializa a aposta apontada por Beckett e Yaseen (2023) para os próximos anos de que o *fact-checking* automatizado deve procurar projetos de chatbots que possam checar, validar ou refutar informações com muito mais velocidade, ao mesmo tempo em que coletam dados das interações com usuários para ajudar na verificação de desinformações.

FÁTIMAGPT E A VERIFICAÇÃO AUTOMÁTICA

A velocidade de verificação e interrupção de compartilhamento de desinformação é um problema apontado pelo próprio chefe de reportagem da Aos Fatos, Leonardo Cazes (2021). Para ele, o *fact-checking* sempre perderá numa corrida contra a desinformação, pois o trabalho necessário para produção com qualidade exige tempo e o produto final não alcançará o mesmo público atingido pela desinformação em primeiro movimento de circulação. Fátima surge para auxiliar nesse combate à desordem informacional, contribuindo com literacia midiática e busca de fontes de informação confiáveis. Em 2020, quando a ferramenta foi lançada, a grande novidade era que o usuário poderia conversar com a Fátima pelo aplicativo de mensagens WhatsApp para pedir a verificação de uma informação. Atualmente, Fátima pode ser acessada também por um navegador de internet e pelo aplicativo de mensagem Telegram.

Em novembro de 2023, a ferramenta incorpora a IAG ChatGPT, começando um período de testes e tendo o nome alterado para FátimaGPT. A proposta é que a integração permita à robô checadora aprender com perguntas dos usuários para entregar melhores respostas e fontes para as checagens requisitadas. Atualmente, FátimaGPT também participa de um programa de prevenção e busca de links que possam levar à desinformação nas redes sociais digitais X (antigo Twitter) e Facebook, além de possuir banco de checagens no site da Aos Fatos e permitir “conversa” para verificação de afirmativas suspeitas dentro das plataformas de mídias digitais nas quais se encontra.

Além da interface atual, quem entender que as afirmações de FátimaGPT não correspondem aos fatos verdadeiros pode contribuir para seu treinamento enviando feedback, podendo contestar erros tais como alucinações da IA. Para Pessoa e Bomfim (2023), estas criações são respostas geradas pelo chatbot que não podem ser explicadas pelos dados de treinamento ou afirmações independentes sem fundamentação geradas pela ferramenta. Essas alucinações podem contribuir para aumentar a desordem informacional, produzindo mesinformação ou confundindo ainda mais o usuário.

Apesar de FátimaGPT ter sido ativada durante período de comoção pública - momento que, como aponta Empoli (2020), é oportuno para “engenheiros do caos” se aproveitarem e lucrarem com um mercado de desinformação -, a robô checadora não havia sido testada em sua versão mais recente em momento de alta produção de desordem informacional. Instantes de agitação social no Brasil desde a criação de Fátima são anteriores à sua versão atual, podendo agora sua eficácia ser testada na situação das enchentes no Rio Grande do Sul.

Apesar de as fortes chuvas no estado sul-rio-grandense terem se iniciado em 27 de abril de 2024, foi em 29 de abril que o Instituto Nacional de Meteorologia (Inmet) emitiu o primeiro alerta para elevado volume pluviométrico. Desde essa data, o estado tem sofrido com a forte intensidade de água causando enchentes, destruição de patrimônio público e privado, deixando ao menos 81 mil pessoas desabrigadas, 132 desaparecidas e 145 mortas. O mercado de desinformação encontrou nesse contexto grande oportunidade e, desde então, altos fluxos de material enganoso têm contribuído para atrapalhar as operações, resgates e acolhimentos locais, o que foi apontado pelo governo federal, através da Advocacia Geral da União (AGU), como um problema que contribui para o agravamento da tragédia no RS.

Sob intuito de testar as potencialidades e problematizações acerca do uso de ferramentas de checagem automática em períodos de grande comoção pública, foram analisadas as respostas de FátimaGPT para as principais desinformações que circularam sobre as enchentes. A velocidade e eficácia na checagem automática poderiam, hipoteticamente, contribuir para diminuição no compartilhamento de material enganoso e atestar a veracidade de informações em ambiente onde a situação está constantemente sendo alterada, com inclusão de sentimentos negativos, como medo, luto e traumas.

PROCESSO METODOLÓGICO

FátimaGPT possui em sua base de treinamento checagens realizadas pela agência Aos Fatos. Perguntas como, por exemplo, “a grama é verde?” foram desconsideradas e houve a preocupação de trazer desordem informacional que já havia circulado - portanto, já deveria ter sido verificada em algum momento. Ao mesmo tempo, foi considerada a hipótese de o usuário tomar conhecimento de desinformações que a Aos Fatos não havia checado, optando-se por coletar checagens da agência Lupa.

O recorte temporal abrange o período de 29 de abril (primeiro alerta meteorológico do Inmet) até 9 de junho, sendo possível a verificação de mais um mês de checagem sobre o assunto. No acervo do site da Lupa, foi selecionado “todos os conteúdos” para garantir que os marcadores não deixassem passar alguma possível checagem sobre as enchentes. Ao todo, a Lupa publicou 52 materiais sobre o assunto, sendo 4 reportagens e 48 checagens - destas, apenas 4 foram publicadas em junho e as outras 44 em maio, visto que as primeiras checagens sobre desinformação da situação no RS começam a ser publicadas no dia 3 de maio.

Foram coletadas checagens onde a desinformação tinha como alvo a própria catástrofe, as doações para o estado atingido, políticas sobre o clima, ações de empresários, celebridades e dos governos municipais, estaduais e federal sobre a tragédia, além de teorias da conspiração. Mesmo tratando-se do mesmo evento, a Lupa utilizou 12 marcadores diferentes, sendo 52,4% (22) marcado como “Rio Grande do Sul”. Sobre o selo, 82% (41) das informações checadas foram tipificadas como “Falso” pela agência, tendo também sido checados conteúdos que receberam o selo de “Falta Contexto” (4) e “Verdadeiro” (1).

A “CONVERSA”

Para facilitar a coleta do acervo da Agência Lupa e endereçamento das desinformações na conversa com a FátimaGPT, optou-se por utilizar o site onde encontra-se a robô checadora. O layout imita uma plataforma de troca de mensagens, simulando conversa com a ferramenta. Perto da logo, há a indicação de que se trata de versão de teste (*beta*). Há um painel do lado esquerdo da tela onde existe a possibilidade de se criar múltiplas conversas com FátimaGPT. Há um botão de reportar erros de execução, assim como indicação no rodapé: “As respostas produzidas por inteligência artificial podem conter erros. Se vir algum, reporte-o usando o botão à esquerda.”

FátimaGPT nos cumprimenta na primeira interação dizendo: “Olá, eu sou a Fátima, a robô checadora do Aos Fatos! Estou aqui para te ajudar a verificar se uma informação é verdadeira ou não”. Em seguida, ensina como podemos iniciar a conversa: “Para interagir comigo, faça uma pergunta ou envie uma mensagem de texto que você gostaria que fosse verificada”. É importante salientar que ainda há dois botões que podem direcionar para a conversa no WhatsApp e no Telegram. As 48 publicações da Lupa foram endereçadas transformando a desinformação em um questionamento. Por exemplo, geralmente iniciava-se a perguntar com “É verdade que...” seguindo as afirmativas dos produtos enganosos. Isso gerava uma resposta de FátimaGPT que era coletada, juntamente com as fontes apresentadas.

Houve a preocupação de indicar questionamentos simples para a robô checadora, já que havia o risco de o ChatGPT não compreender a pergunta. Na tentativa de verificação de 6 checagens, apesar de inúmeras reformulações das perguntas, FátimaGPT não conseguiu compreender o que estava sendo questionado. Observou-se, ainda, que em grande parte dos questionamentos não compreendidos, tratava-se de desinformação veiculada apenas como imagem ou vídeo.

Do total das 48 checagens, 8 questionamentos não possuíam dados suficientes no banco de treinamento para gerar uma resposta. Nesta situação, FátimaGPT exhibe a seguinte mensagem: “Desculpe, mas não tenho informações para responder a essa pergunta”. Necessário indicar conflito entre o entendimento da Lupa e de FátimaGPT sobre um objeto de desinformação: antes sendo classificado como falso, a robô checadora

indicou que se tratava de informação verdadeira, inclusive indicando fontes. Isso pode ser devido ao método de checagem e selos diferentes com os quais Lupa e Aos Fatos trabalham, além de poder ter sido alterado o entendimento sobre a desinformação em um mês. É interessante observar como a robô checadora teve outra compreensão e como isso tem capacidade de impactar no entendimento do usuário da plataforma - podendo gerar mais confusão, caso o indivíduo esteja em situação de vulnerabilidade em relação à desinformação.

Para exemplificar, a Lupa checkou se as águas do Guaíba haviam alcançado o topo do Muro da Mauá, estrutura de proteção contra enchentes na Avenida Mauá, em Porto Alegre. Enquanto para Maiquel Rosauero, que assina a apuração e cita como fonte o Departamento Municipal de Água e Esgoto de Porto Alegre (Dmae), o selo atribuído foi de “Falso”, para FátimaGPT tratava-se de informação verdadeira. Há indicação de vídeo gravado no México em 2020 com animais carregados que circulou como se fosse no RS neste ano. Por não haver nenhuma menção ao Muro da Mauá ou às águas do Guaíba na checagem da Aos Fatos sobre o vídeo mexicano, acredita-se que se trata de uma alucinação da IA.

CONSIDERAÇÕES FINAIS

Considerando que a ferramenta está em fase de testes e que tratamos aqui de desinformações sobre evento de alta comoção pública, responder 70,83% das perguntas foi considerado um resultado positivo acerca da utilização da robô checadora. Importante, no entanto, ressaltar que 14 questionamentos não possuíam dados para resposta ou não foram compreendidos por FátimaGPT, além das 4 ocorrências onde possivelmente há produção de alucinação da IA. Para este trabalho, a plataforma mostrou-se eficaz na análise qualitativa: o modelo apresentado possui layout interessante das respostas e traz fontes confiáveis (no caso, da própria agência). Além disso, tratando-se de situação de emergência e tragédia, FátimaGPT, mais de uma vez, finalizou respostas com frases como “É importante verificar informações atualizadas diretamente com fontes confiáveis para obter os dados mais recentes sobre a situação [...]” (FátimaGPT, 2024). Caso o projeto continue e receba investimentos que ampliem sua capacidade e velocidade de coleta de dados, poderá se tornar ferramenta importante de literacia midiática e uma aliada no combate à desinformação.

REFERÊNCIAS

- AOS FATOS. Brasil. 2015. Disponível em: <https://www.aosfatos.org/> Acesos em 16 mar. 2021.
- BECKETT, Charlie; YASEEN, Mira. *Generating change: a global survey of what news organisations are doing with AI*. Disponível em: <https://www.journalismai.info/research/2023-generating-change> Acesso em: 04 out. 2023.
- CAZES, Leonardo. **Apura Verdade – Leonardo Cazes (Episódio 003)**. 2021. (1 h; 10 min). Disponível em: <https://open.spotify.com/episode/64MEFiCBAz8DC9NtsBd0rc?si=2VsUYaJLS0moCjFIFDTbZQ> Acesso em: 06 dez. 2023.
- EMPOLI, Giuliano Da. **Os engenheiros do caos**. [tradução Arnaldo Bloch] – 1. ed.; 3. reimp. – São Paulo: Vestígio, 2020. (Espírito do Tempo)
- FATOS, Aos. **Conheça a Fátima, a robô checadora do Aos Fatos**. 2020. (1 min; 30 s). Disponível em: <https://www.youtube.com/watch?v=yziW9TWgMXU> Acesso em: 06 dez. 2023.
- FUTURE OF LIFE INSTITUTE. *Policymaking in the pause: what can policymakers do now to combat risks from advanced AI systems?* - Narberth, EUA: Future of Life Institute, 2023. Disponível em: futureoflife.org. Acesso em: 15 ago. 2023.
- LUPA. Brasil. 2023. Disponível em: <https://lupa.uol.com.br/> Acesso em: 13 ago. 2023.
- PESSÔA, Paulo; BOMFIM, Ivan. I.A., (des)informação e (des)contextualização no jornalismo. **Vozes & Diálogos**, v. 22, n. 2, 2023. Disponível em: <https://doi.org/10.14210/vd.v22n2.p26-35> Acesso em: 28 jun. 2024.
- SANTAELLA, Lucia. **Há como deter a invasão do ChatGPT?** - 1. ed. - São Paulo: Estação das Letras e Cores, 2023a. (Coleção Interrogações).
- SANTAELLA, Lúcia. Balanço crítico preliminar do ChatGPT. **Revista FAMECOS**, v. 30, n. 1, p. 1 - 12, jan. / dez., 2023b. Disponível em: revistaseletronicas.pucrs.br. Acesso em: 27 mai. 2024.
- SIMON, Felix. *Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy*. **Digital Journalism**, v. 10, n. 10, p. 1832 - 1854, 2022. Disponível em: www.tandfonline.com. Acesso em: 14 abr. 2024.

Inserir aqui as referências bibliográficas em fonte Times New Roman, em corpo 11 (onze), com espaçamento simples entre as linhas. As referências bibliográficas, no fim do trabalho, devem ter os dados completos e seguir as normas da ABNT 6023 para trabalhos científicos. Cada referência deve ocupar um parágrafo e devem estar separados por dois espaços simples.