

Entrevistando a IA sobre as eleições municipais brasileiras: aproximações metodológicas para aferição de confiabilidade de quatro chatbots de Inteligência Artificial Generativa¹

Zanei Ramos Barcellos²

Fernanda Vasques Ferreira³

Marlise Viegas Brenol⁴

Rafiza Varão⁵

Resumo expandido

A partir do problema de pesquisa resumido na pergunta “os chatbots inteligentes colaboram mais com a informação ou com a desinformação nos períodos pré-eleitorais?”, e com o objetivo de verificar se os chats de inteligência artificial generativa (IAG) abertos fornecem informações corretas aos cidadãos sobre as eleições municipais de 2024 no Brasil, escolheu-se aplicar a alguns deles um questionário estruturado com respostas abertas para um estudo quali-quantitativo.

Selecionou-se como objetos de pesquisa chatbots movidos por IAGs e as informações contidas nas resposta às 60 perguntas formuladas pelos pesquisadores autores

¹ Trabalho apresentado no GT Regimes de verificação em tempos de IA do IV Encontro Virtual da ABCiber – Associação Brasileira de Pesquisadores em Ciberultura. Perspectivas Interdisciplinares e Reconfigurações na Ciberultura: Dados, Algoritmos e Inteligência Artificial. Realização da UNIFAE, nos dias 20 e 21 de junho de 2024.

² Professor do Departamento de Jornalismo, Faculdade de Comunicação da Universidade de Brasília, Doutor em Gestão Urbana. E-mail: zaneibarcellos@unb.br.

³ Professora do Departamento de Jornalismo, Doutora em Comunicação, Faculdade de Comunicação da Universidade de Brasília. E-mail: fernanda.jornalista82@gmail.com.

⁴ Professora do Departamento de Comunicação Organizacional, Faculdade de Comunicação da Universidade de Brasília, Doutora em Comunicação e Informação. E-mail: marlise.brenol@fac.unb.br.

⁵ Professora do Departamento de Jornalismo, Doutora em Teorias e Tecnologias da Comunicação, Faculdade de Comunicação da Universidade de Brasília. E-mail: rafiza@unb.br.

deste trabalho, sobre as eleições em questão, fornecidas pelas quatro IAGs escolhidas: Copilot (Microsoft), Sapiens (Sapiens Technology), ChatGPT (OpenAI) e LuzIA (Meta).

A proposta inicial era escolher os sistemas mais populares, de uso gratuito, criados por grandes desenvolvedores de tecnologia (Microsoft, Google, OpenAI e Meta). O Google, porém, decidiu que sua IAG, a Gemini, não responderia a questões sobre eleições em qualquer país onde se realizassem pleitos neste ano. Na falta de outra IA popular e aberta de grandes desenvolvedores acessíveis no Brasil, optou-se por incluir uma IAG nacional, a Sapiens, mesmo não sendo muito conhecida, principalmente para comparar seu comportamento com as internacionais.

Os *prompts* formulados atenderam a duas perspectivas principais: a) perguntas que o cidadão comum poderia fazer para tirar dúvidas sobre os processos eleitorais ou simplesmente para conhecê-los melhor; e, b) perguntas que poderiam ser feitas por pessoas desinformadas, não no sentido de não terem acesso à informação, mas de orientarem suas dúvidas a partir de conteúdos desinformativos. Nesse segundo caso, portanto, a palavra desinformado significa alguém que, no ecossistema desinformativo, consome conteúdo falso, inverídico, equivocado – mas que crê nesses conteúdos como potencialmente verdadeiros. O movimento duplo se deu sobretudo por duas prerrogativas: 1) a ideia de que os cidadãos buscam informações para orientar suas escolhas eleitorais nas tecnologias digitais -- sobretudo mídias sociais (Ruediger, Grassi, 2020; Mitchelstein, Boczkowski, 2018) e agora as ferramentas de IA generativas (Sullivan-Paul, 2023); 2) a concepção de que a IA, a despeito de seus ganhos, é também uma ferramenta para disseminação de desinformação, incluindo aquela sobre política (Kaplan, 2020; Bontridder, Poullet, 2021; Frias, 2021).

Há aqui, portanto, a percepção de que a entrada da IA no cotidiano de usuários comuns interfere na própria construção do espaço público, que passa a ser atravessado, no amplo contexto das mídias digitais, por instrumentos não-humanos capazes de interferir nas

decisões humanas, por sua vez. Esse fenômeno já era perceptível a partir da ação dos chamados *bots* (Ferreira, Varão, Boselli, 2022), mas cuja operação ainda se distanciava de comandos simples, efetuados por usuários comuns no ambiente digital. Com os novos aparatos de IA, contudo, esses comandos se tornam cada vez mais acessíveis, além de produzirem conteúdos cada vez mais verossimilhantes. Desse modo, seria possível confiar nas respostas dos chatbots aqui listados como fonte de informação em eleições? Há pesquisas que indicam erros, falhas e alucinações nas respostas elaboradas pelos chatbots de IAG (Lemos, 2024), por isso este estudo se torna ainda mais relevante ao se propor à compreensão das potencialidades e limites das novas formas de comunicação para a cidadania.

Estimou-se que 60 perguntas diferentes gerariam volume de respostas suficientes para embasar análises capazes de realizar o objetivo da pesquisa. Cada um dos quatro pesquisadores envolvidos formulou 15 delas, para, desta forma, favorecer a criação de um formulário diverso, abrangente e com diferentes pontos de vista.

As perguntas foram elaboradas a partir da lista de informações mais buscadas no site do Tribunal Superior Eleitoral (TSE) e com base nas principais *fake news* sobre eleições. Acrescentamos ao universo de perguntas aquelas que poderiam figurar no repertório de cidadãos que pretendem se informar ou daqueles, eventualmente, mal informados que buscam reforçar suas convicções, confirmá-las ou ainda daqueles que desejam tirar dúvidas pontuais sobre as eleições municipais em 2024. Nessa última situação, baseando-nos em diversas *fake news* publicadas em sites de checagem, mas que não figuram em listas.

Para avaliar a precisão das respostas dos chatbots, elaboraram-se seis parâmetros padrões, cada um associado a um valor numérico inteiro de referência, de zero a cinco, conforme a lista abaixo:

5 – Resposta correta com informações complementares pertinentes ou desacompanhada de outras informações;

- 4 – Resposta correta acompanhada de outra(s) informação(ões), sendo alguma(s) incorreta(s) ou não pertinente(s);
- 3 – Informações corretas, mas sem responder diretamente, total ou parcialmente, à pergunta;
- 2 – Disse que não sabia/podia responder, indicando ou não outras fontes;
- 1 – Resposta incorreta, mas com outra(s) informação(ões) pertinente(s) correta(s);
- 0 – Resposta incorreta, acompanhada ou não de outra(s) informação(ões) pertinente(s) ou não incorreta(s), ou mesclando informação(ões) correta(s) com incorreta(s).

A avaliação da correção da resposta, portanto, deveria ser qualitativa, tendo como fontes de aferição de veracidade as informações disponíveis nos sites da Justiça Eleitoral e/ou nas matérias publicadas por veículos jornalísticos de referência/credibilidade⁶. Coube a cada pesquisador avaliar as respostas fornecidas pelo chat que entrevistou e classificá-las em alguma das seis categorias.

A média de cada chatbot obtém-se somando o valor atribuído a cada resposta e dividindo o resultado por 60. Assim, as médias obtidas não significam a associação a um dos seis parâmetros propostos, mas servem para estabelecer um grau de proximidade maior ou menor a um dos extremos: desinformação (mais próximo ao valor zero) e informação (mais próximo ao valor 5).

Entre 28 de abril e 1º de maio de 2024 foi realizado um pré-teste do questionário. Quinze perguntas foram aplicadas a três dos chats em investigação. Não foram aplicadas ao Sapiens porque a IAG ainda não estava definida como uma das que seriam investigadas. O pré-teste comprovou que a proposta metodológica funcionava, mas evidenciou que os textos de alguns dos seis parâmetros de avaliação precisavam ser ajustados para tornar o

⁶ Cabe destacar aqui os aspectos de referência e credibilidade utilizados, uma vez que no contexto da desinformação, existem veículos que se dizem jornalísticos com formato e padrões de apresentação próximos do jornalismo de referência que se apropriam dessas características para falsear a realidade, distorcer, enganar, confundir e até mesmo manipular.

conjunto o mais abrangente possível. Mesmo após os ajustes, admitiu-se a impossibilidade de enquadrar todas respostas exatamente em uma das seis possibilidades. Nestes raros casos, o pesquisador deveria classificá-la na opção mais próxima da verdade.

Se todos os quatro pesquisadores participaram conjuntamente da estruturação da lista das 60 perguntas, quando coube a cada um formular 15 delas, a aplicação do formulário e avaliação das respostas foi individual. Cada pesquisador se dedicou a um dos chatbots em questão. Maio de 2024 foi o recorte temporal para a aplicação dos *prompts* às quatro IAGs por ser um mês sem muita movimentação político-eleitoral, sem notícias relevantes sobre o assunto e, principalmente, para que todos os chats fossem entrevistados mais ou menos no mesmo período. Rafiza Varão

Os valores numéricos atribuídos pelos pesquisadores a todas as respostas de cada IAG foram anotados em uma planilha eletrônica, para facilitar os cálculos de médias, obtenção de outras estatísticas, evidenciar fatos e facilitar as análises qualitativas. A planilha consta de três tabelas acopladas. Fórmulas foram inseridas para cálculos automatizados, cujos resultados mudam e aparecem a cada valor inserido. A primeira tabela tem cinco colunas, a primeira para numerar as 60 perguntas do formulário aplicadas, e as demais colunas dedicadas às avaliações numéricas atribuídas às respostas únicas fornecidas pelos quatro chatbots investigados. A última linha fornece a média aritmética das classificações obtidas por cada IAG e a média geral atribuída aos chats. A segunda tabela conta quantas classificações de zero a cinco cada chatbot obteve e fornece a soma do número destes conceitos atribuídos pelos avaliadores a cada pergunta. A terceira tabela faz o mesmo que a segunda, porém fornece os resultados em percentuais. Sempre que o resultado não for número inteiro, a precisão é de centésimos.

Os resultados preliminares, em caráter exploratório, mostram as médias obtidas pelos quatro chats próximos a 4: Copilot, 3,92; Sapiens, 4,33; ChatGPT, 4,47, e LuzIA, 4,50. A média geral foi 4,30. Assim, pode-se dizer que, pelos parâmetros estabelecidos para

a pesquisa, os chats avaliados tendem a colaborar mais para a informação que para a desinformação do cidadão sobre as eleições municipais brasileiras de 2024, em que pesem respostas que tergiversaram as perguntas feitas ou, ainda, apresentaram respostas desatualizadas e, em alguns casos, alucinações. Isso significa que a IA Generativa, quando motivada apenas por questionamentos humanos tende a repetir fatos verídicos, recolhidos da sua base de dados – o que não significa que não possa errar ou inventar.

Palavras-chave

Inteligência Artificial Generativa; Chatbots; Desinformação; Confiabilidade; Eleições.

Referências

- BONTRIDDER, N., POULLET, Y. The role of artificial intelligence in disinformation. **Data & Policy**, v. 32, p. 1-18, n. 3, 2021.
- DE FRIAS, E. S. Inteligencia artificial, desinformación y populismo digital: Cómo las plataformas digitales impulsan los movimientos de extrema derecha. **Razón y palabra**, v. 25, n. 112, p. 23, 2021.
- FERREIRA, Fernanda Vasques; VARÃO, Rafiza; BOSELLI, Marco Aurélio. Desinformação sobre a vacina da Covid-19 no Brasil: medição de alcance e impactos das fake news na saúde. **Razón y Palabra**, v. 25, p. 122-139, 2022.
- KAPLAN, A. Artificial intelligence, social media, and fake news: Is this the end of democracy. In: GÜL, A. A., ERTÜRK, Y. D., ELMER, P. (Org). **Digital transformation in media and society**. Istanbul: Istanbul University Press, 2020, p. 149-161.
- LEMOS, André Luiz Martins. Erros, falhas e perturbações digitais em alucinações das IA generativas: Tipologia, premissas e epistemologia da comunicação. **MATRIZES**, v. 18, n. 1, p. 75-91, 2024.
- MITCHELSTEIN, Eugenia; BOCZKOWSKI, Pablo J. Juventud, estatus y conexiones. Explicación del consumo incidental de noticias en redes sociales. **Revista mexicana de opinión pública**, n. 24, p. 131-145, 2018.
- RUEDIGER, M. A.; GRASSI, A. (Coord.). The digital ecosystem in the 2020 municipal elections in Brazil: the buzz of distrust in the electoral system on Facebook, YouTube and Twitter. Policy paper. Rio de Janeiro: FGV DAPP, 2020.
- SULLIVAN-PAUL, M.. How would ChatGPT vote in a federal election? A study exploring algorithmic political bias in artificial intelligence. Tóquio: Escola de Políticas Públicas, Universidade de Tóquio, 2023. (Tese de doutorado).