

Usando análise de sentimentos para diferenciar bots e humanos na divulgação de publicações científicas sobre COVID-19 na mídia social X: um estudo com ChatGPT 3.5 e Gemini 1.5 Flash

Using sentiment analysis to differentiate bots and humans in the dissemination of scientific publications on COVID-19 on social media X: a study with ChatGPT 3.5 and Gemini 1.5 Flash

Usando análisis de sentimientos para diferenciar bots y humanos en la difusión de publicaciones científicas sobre COVID-19 en la red social X: un estudio con ChatGPT 3.5 y Gemini 1.5 Flash

Danielle Pompeu Noronha Pontes

Universidade do Estado do Amazonas, Manaus, Brasil

João de Melo Maricato

Universidade de Brasília, Brasília, Brasil

ORIGINAL

Resumo

Objetivo. O estudo tem como objetivo investigar a aplicação da análise de sentimentos para diferenciar contas automatizadas (bots) e humanas na divulgação de publicações científicas sobre COVID-19 na rede social X. Para isso, a pesquisa compara a eficácia das ferramentas de Large Language Models (LLMs) ChatGPT 3.5 e Gemini 1.5 Flash na análise de sentimento, através da classificação de polaridade, expressos em postagens sobre um artigo científico. O estudo busca compreender as diferenças de desempenho entre essas ferramentas, avaliar sua eficácia na tarefa de classificação de polaridades e identificar padrões de sentimentos que melhor distinguem bots de usuários humanos no contexto da disseminação científica. **Método.** Este estudo inicia com a coleta de uma amostra de postagens do X que fizeram menção à publicação estudada. Os posts foram coletados usando Python 3.12, a biblioteca BeautifulSoup 4.12 no ambiente Google Colab, resultando em um conjunto de 9.792 postagens e 5.601 perfis únicos. Em uma segunda etapa, esses perfis foram confrontados com um dataset previamente classificado como bot ou humano. Para aprimorar a confiabilidade da classificação, posteriormente foi realizada uma reclassificação manual de 41 contas que haviam feito mais de quatro postagens (terceira etapa), identificando 20 como bots e 21 como humanas. Essas contas geraram um total de 3.493 postagens, que foram submetidas à classificação de polaridade dos sentimentos (quarta etapa) utilizando as ferramentas ChatGPT 3.5 e Gemini 1.5 Flash. A classificação seguiu um prompt padronizado para categorizar os sentimentos como positivo, negativo ou neutro, sendo aplicada em lotes de 100 postagens devido às limitações de tokens das ferramentas. Na quinta etapa, 315 postagens analisadas foram classificadas manualmente. **Resultados.** A análise de 3.493 postagens sobre o artigo científico no X revelou uma predominância de sentimentos negativos (92,3%), com posts neutros (6,2%) e positivos (0,6%) sendo menos frequentes (0,9 não foram identificados). As ferramentas de inteligência artificial ChatGPT 3.5 e Gemini 1.5 Flash mostraram desempenhos semelhantes ao classificar sentimentos negativos, mas divergências surgiram em 315 postagens, com ChatGPT atingindo 85% de precisão em posts que o Gemini não conseguiu classificar. Bots apresentaram uma maior variabilidade emocional e foram mais críticos em relação à divulgação científica, enquanto as postagens humanas tenderam a ser mais neutras e consistentes, destacando diferenças relevantes para a detecção de bots. **Conclusões.** A análise de sentimentos realizada por ChatGPT e Gemini destaca a capacidade dessas ferramentas de classificar postagens em redes sociais relacionadas a artigos científicos, revelando padrões distintos entre bots e contas humanas. Os bots tendem a gerar conteúdos mais polarizados e predominantemente negativos, enquanto os humanos apresentam uma diversidade maior de sentimentos, com um equilíbrio entre postagens negativas, neutras e algumas positivas. Embora o

ChatGPT tenha se mostrado mais eficaz em cenários com poucos dados contextuais ou metadados para avaliar corretamente a polaridade emocional de um texto, o estudo sugere que uma análise mais abrangente é necessária para refinar as ferramentas e aprofundar a compreensão das interações entre contas humanas e contas robotizadas.

Palavras-chave: análise de sentimentos, chatGPT, gemini, X(twitter), bots

Abstract

Objective. This study aims to investigate the application of sentiment analysis to differentiate between automated accounts (bots) and human users in the dissemination of scientific publications about COVID-19 on the social network X. To achieve this, the research compares the effectiveness of Large Language Model (LLM) tools, specifically ChatGPT 3.5 and Gemini 1.5 Flash, in classifying the sentiment polarity expressed in posts about a scientific article. The study seeks to understand the performance differences between these tools, evaluate their effectiveness in polarity classification, and identify sentiment patterns that best distinguish bots from human users in the context of scientific dissemination. **Method.** This study begins with the collection of a sample of posts from X that mentioned the analyzed publication. The posts were collected using Python 3.12 and the Beautiful Soup 4.12 library within the Google Colab environment, resulting in a dataset of 9,792 posts and 5,601 unique profiles. In the second stage, these profiles were compared to a previously classified dataset of bots and human users. To enhance classification reliability, a subsequent manual reclassification was performed on 41 accounts that had posted more than four times (third stage), identifying 20 as bots and 21 as human users. These accounts generated a total of 3,493 posts, which were then subjected to sentiment polarity classification (fourth stage) using ChatGPT 3.5 and Gemini 1.5 Flash. The classification followed a standardized prompt to categorize sentiments as positive, negative, or neutral and was applied in batches of 100 posts due to token limitations of the tools. In the fifth stage, 315 of the analyzed posts was manually classified for validation. **Results.** The analysis of 3,493 posts about the scientific article on X revealed a predominance of negative sentiments (92.3%), with neutral posts (6.2%) and positive ones (0.6%) being less frequent (0.9% were unidentified). The AI tools ChatGPT 3.5 and Gemini 1.5 Flash showed similar performance in classifying negative sentiments, but discrepancies emerged in 315 posts, with ChatGPT achieving 85% accuracy in posts that Gemini failed to classify. Bots exhibited greater emotional variability and were more critical of scientific dissemination, whereas human posts tended to be more neutral and consistent, highlighting relevant differences for bot detection. **Conclusions.** Sentiment analysis performed by ChatGPT and Gemini highlights the capability of these tools to classify social media posts related to scientific articles, revealing distinct patterns between bots and human accounts. Bots tend to generate more polarized and predominantly negative content, while humans exhibit a greater diversity of sentiments, with a balance between negative, neutral, and some positive posts. Although ChatGPT proved to be more effective in scenarios with limited contextual data or metadata to accurately assess a text's emotional polarity, the study suggests that a more comprehensive analysis is needed to refine these tools and deepen the understanding of interactions between human and bot accounts.

Keywords: sentiment analysis, chatGPT, gemini, X (twitter), bots

Resumen

Objetivo. Este estudio tiene como objetivo investigar la aplicación del análisis de sentimientos para diferenciar entre cuentas automatizadas (bots) y usuarios humanos en la difusión de publicaciones científicas sobre COVID-19 en la red social X. Para ello, la investigación compara la eficacia de las herramientas de Modelos de Lenguaje de Gran Escala (LLMs), específicamente ChatGPT 3.5 y Gemini 1.5 Flash, en la clasificación de los sentimientos expresados en publicaciones sobre un artículo científico. El estudio busca comprender las diferencias de rendimiento entre estas herramientas, evaluar su eficacia en la clasificación de polaridades y identificar patrones de sentimientos que mejor distingan a los bots de los usuarios humanos en el contexto de la divulgación científica. **Método.** Este estudio comienza con la recopilación de una muestra de publicaciones en X que mencionaron la publicación analizada. Los posts fueron recolectados utilizando Python 3.12 y la biblioteca Beautiful Soup 4.12 en el entorno de Google Colab, lo que resultó en un conjunto de 9.792 publicaciones y 5.601 perfiles únicos. En una segunda etapa, estos perfiles fueron comparados con un conjunto de datos previamente clasificado como bot o humano. Para mejorar la confiabilidad de la clasificación, posteriormente se realizó una reclasificación manual de 41 cuentas que habían realizado más de cuatro publicaciones (tercera etapa), identificando 20 como bots y 21 como humanas. Estas cuentas generaron un total de 3.493 publicaciones, que fueron sometidas a análisis de sentimientos (cuarta etapa) utilizando las herramientas ChatGPT 3.5 y Gemini 1.5 Flash. La clasificación siguió un prompt estandarizado para categorizar los sentimientos como positivos, negativos o neutros, aplicándose en lotes de 100 publicaciones debido a las limitaciones de tokens de las herramientas. En la quinta etapa, 315 de las publicaciones analizadas fue clasificada manualmente para su validación. **Resultados.** El análisis de 3.493 publicaciones sobre el artículo científico en X reveló una predominancia de sentimientos negativos (92,3%), con posts neutros (6,2%) y positivos (0,6%), siendo estos últimos menos frecuentes (0,9% no fueron identificados). Las herramientas de inteligencia artificial ChatGPT 3.5 y Gemini 1.5 Flash mostraron desempeños similares al clasificar sentimientos negativos, pero surgieron divergencias en 315 publicaciones, con ChatGPT alcanzando un 85% de precisión en posts que Gemini no logró clasificar. Los bots presentaron una mayor variabilidad emocional y fueron más críticos en relación con la difusión científica, mientras que las

publicaciones humanas tendieron a ser más neutras y consistentes, destacando diferencias relevantes para la detección de bots.

Conclusiones. El análisis de sentimientos realizado por ChatGPT y Gemini destaca la capacidad de estas herramientas para clasificar publicaciones en redes sociales relacionadas con artículos científicos, revelando patrones distintos entre bots y cuentas humanas. Los bots tienden a generar contenidos más polarizados y predominantemente negativos, mientras que los humanos presentan una mayor diversidad de sentimientos, con un equilibrio entre publicaciones negativas, neutras y algunas positivas. Aunque ChatGPT demostró ser más eficaz en escenarios con pocos datos contextuales o metadatos para evaluar correctamente la polaridad emocional de un texto, el estudio sugiere que es necesaria un análisis más exhaustivo para refinar las herramientas y profundizar en la comprensión de las interacciones entre cuentas humanas y automatizadas.

Palabras clave: análisis de sentimientos, chatGPT, gemini, X (Twitter), bots

1 Introdução

As mídias sociais, notadamente o X, exercem uma influência substancial no ciclo de vida da produção científica (Darling et al., 2013). Ao impulsionar a interconexão de redes acadêmicas, acelerar o desenvolvimento de ideias e amplificar a difusão do conhecimento, essas plataformas proporcionam potenciais inegáveis. Contudo, a presença crescente de *bots* nesse ecossistema, alguns benignos e outros não, levanta a necessidade imperativa de diferenciar contas automatizadas e não automatizadas, além de avaliar seus impactos na disseminação da desinformação e na manipulação da opinião pública nas redes sociais. Esse cenário representa um desafio crescente para a sociedade, especialmente no contexto da divulgação científica, agravado pela proliferação de *bots* que simulam comportamentos humanos, dificultando a identificação de fontes confiáveis de informação.

Consequentemente, a participação de *bots* na disseminação de artigos científicos no X assume relevância crescente no contexto atual, marcado pela confluência da inteligência artificial (IA) e das plataformas de mídias sociais. Nesse cenário, a análise de sentimentos, através da classificação de polaridades, e a identificação de padrões de linguagem emergem como ferramentas potencialmente úteis para identificar a atuação de *bots* e aumentar a credibilidade da informação científica nas redes sociais. A capacidade de distinguir entre conteúdo gerado por humanos e por máquinas é fundamental para promover o debate público fundamentado em evidências e combater a disseminação de notícias falsas.

Nesse cenário, estudos recentes exploram a análise de sentimentos em contextos altmétricos, investigando correlações entre sentimentos expressos, plataformas de mídia social e impacto acadêmico. Arredondo (2018) destacou a falta de correlação entre sentimentos expressos e impacto social de publicações científicas; Halevi e Schimming (2018) propuseram uma análise de sentimentos em menções de mídias sociais e notícias científicas, facilitando a compreensão do conteúdo; Hassan et al. (2021) mostraram a eficácia de algoritmos de aprendizado de máquina na identificação de sentimentos em dados altmétricos; Hassan et al. (2023) exploraram interações sociais, especialmente em postagens no X, revelando variações nos sentimentos expressos em diferentes campos de pesquisa, fornecendo percepções sobre a circulação da pesquisa científica nas mídias sociais.

A aplicação de IA na análise de sentimentos em mídias sociais representa uma área em evolução no Processamento de Linguagem Natural (NLP), que visa interpretar dados não estruturados de várias fontes. Para Liu et al. (2020), essa evolução passa por três fases importantes: a era baseada no léxico, a era do aprendizado de máquina (ML) e a era atual, fundamentada na "arquitetura *Transformer*". Para esses autores, a análise de sentimentos, também conhecida como mineração de opiniões, combina recuperação da informação, NLP e inteligência artificial para determinar as opiniões de um grupo sobre um tema específico, classificando expressões como positivas, negativas ou neutras. Ela investiga a expressão de opiniões, polaridades, sentimentos, avaliações, atitudes e emoções em textos escritos em relação a entidades e seus atributos.

Em redes sociais como o X, um dos principais desafios na classificação do sentimento das postagens é a expressão informal das opiniões, o que leva a erros ortográficos e uso descuidado da gramática (Barreto et al., 2023). Estudos recentes indicam que os sentimentos expressos em postagens no X podem prever o impacto social inicial de artigos de pesquisa, com artigos mencionados em postagens positivas e neutras apresentando um impacto maior do que aqueles mencionados ou não em postagens negativas (Hassan et al., 2023). No entanto, é válido questionar se essas impressões positivas, neutras ou negativas são originadas de contas automatizadas (*bots*) ou não.

O surgimento de ferramentas de modelos de linguagem como ChatGPT e Gemini representa uma abordagem inovadora para a detecção de *bots* aplicados a dados alométricos, por meio da análise de sentimentos. O ChatGPT é destacado como uma ferramenta potencialmente útil para análises de sentimentos em várias áreas de pesquisa. Um estudo conduzido por Silva e Serrano (2023) enfatiza sua eficácia, ressaltando que pode ser extremamente útil na interpretação automatizada de textos. Segundo os autores, o ChatGPT se destaca pela capacidade de compreender o contexto das mensagens, produzindo resultados comparáveis aos de análises conduzidas manualmente pelos pesquisadores.

Segundo Buscemi e Proverbio (2024) o Gemini é a mais recente família de modelos de linguagem multimodal desenvolvida pela Google DeepMind, lançada em dezembro de 2023, na época composta por Gemini Ultra, Pro e Nano. Os autores indicam que esses modelos, baseados em arquiteturas *Transformer* de decodificador único, ajustados para eficiência em Unidades de Processamento de Tensor (TPUs), são capazes de realizar tarefas complexas, incluindo reconhecimento de imagem, superando humanos em testes de Compreensão de Linguagem Multitarefa em Massa. O modelo Nano-2, adaptado para dispositivos de borda, possui 3,25 bilhões de parâmetros, enquanto os detalhes dos outros modelos permanecem não divulgados. Gemini foi treinado com dados multimodais e multilíngues, abrangendo texto, imagem, áudio e vídeo.

Segundo Rane et al. (2024) as principais diferenças entre Gemini e ChatGPT se concentram em suas áreas de especialização e estilo de interação. O Gemini se destaca na precisão factual graças à sua integração com a pesquisa do Google, permitindo respostas mais precisas e citações de fontes confiáveis. Ele é particularmente eficaz em tarefas que exigem alta precisão e integração de dados multimodais, como imagens e vídeos. Em contraste, o ChatGPT, da OpenAI, é reconhecido por sua fluidez conversacional e criatividade, sendo ideal para gerar textos envolventes e diversos, como poemas, código e composições, capturando nuances e infundindo personalidade nas interações.

Outra diferença significativa entre essas duas ferramentas de IA reside na acessibilidade e na disponibilidade. O ChatGPT tem maior acessibilidade ao oferecer uma versão gratuita e uma *premium* (ChatGPT Plus), enquanto o Gemini ainda está em desenvolvimento com acesso público limitado. Embora ambos os modelos enfrentem desafios como mitigação de viés e segurança, o Gemini se sobressai em precisão factual, enquanto o ChatGPT é preferido para interações mais criativas e fluidas. A escolha entre os dois depende do propósito: Gemini para precisão e verificação de fontes e ChatGPT para criatividade e fluência em diálogos (Rane et al., 2024).

Diante do exposto, este estudo avalia e compara a eficácia dos modelos de linguagem ChatGPT versão 3.5 e Gemini versão 1.5 Flash para analisar os sentimentos das postagens, através da classificação de polaridades, dos usuários do X sobre publicação científica que aborda estudos relacionados ao COVID19, com o intuito de explorar essa capacidade na diferenciação entre contas consideradas *bots* ou humanas, com base na polaridade das postagens (positivas, negativas ou neutras). Para isso, são utilizados dados de contas previamente classificadas como humanas ou *bots*, conforme o estudo de Pontes e Maricato (2023a). A pesquisa busca, também, identificar padrões que possibilitem a classificação das contas do X como *bots* ou humanas, a partir dos sentimentos expressos nas postagens.

A escolha dos modelos de linguagem ChatGPT e Gemini para essa investigação se justifica por suas avançadas capacidades de processamento de linguagem natural e de aprendizado de máquina. Esses modelos têm demonstrado um notável potencial para compreender nuances linguísticas e identificar padrões complexos em grandes volumes de dados que necessitam ser analisados. Isso tudo em frações de segundos para cada postagem. Além disso, ao escolher essas ferramentas, a pesquisa pode complementar estudos anteriores sobre a detecção de *bots* no contexto alométrico, a partir de novas ferramentas analíticas.

As questões de pesquisa que nortearam este estudo são:

Q1: Qual a diferença de resultados entre as ferramentas *Large Language Models* (LLMs) ChatGPT 3.5 e Gemini 1.5 Flash ao classificar os sentimentos de postagens realizadas no X sobre artigos científicos?

Q2: Qual a eficácia das ferramentas LLMs ChatGPT 3.5 e Gemini 1.5 Flash ao classificar os sentimentos de postagens realizadas no X sobre artigos científicos?

Q3: Quais são os padrões de sentimentos que mais diferenciam as contas *bots* das contas humanas na rede social X no contexto da divulgação científica?

2 Revisão de literatura

A análise de sentimentos tem emergido como uma ferramenta relevante na identificação e compreensão do comportamento de *bots* em mídias sociais, como o X. Ela tem evoluído significativamente nas últimas décadas, refletindo a crescente influência dessas plataformas na sociedade contemporânea e a necessidade de técnicas avançadas para processar dados visando identificar interações humanas autênticas e ações automatizadas. A literatura sobre esse tema revela uma evolução nas metodologias e técnicas empregadas para distinguir entre usuários humanos e *bots*, refletindo a crescente complexidade e sofisticação desses agentes e suas interações online.

Para entender o avanço das pesquisas nessa área, foi realizado uma busca na base de dados Scopus em junho de 2024 utilizando a chave de busca: (TITLE-ABS-KEY (*sentiment* AND *analysis*) AND TITLE-ABS-KEY (*bot* OR *bots*) AND TITLE-ABS-KEY (*twitter* OR *x*)). Como resultado foram identificados 106 artigos publicados entre os anos de 2011 e 2024, entre os quais serão destacados, em ordem cronológica, os mais relevantes na percepção dos autores desta pesquisa, tornando-se possível uma compreensão da evolução da área.

Mustafaraj e Metaxas (2011) partem da hipótese de que a interação genuína e o engajamento podem ser melhor medidos analisando-se *tweets* que exibem sinais de intervenção humana. Com isso, segundo os autores, a edição do texto de um *tweet* original antes de retuitá-lo pode revelar um engajamento consciente do usuário com o conteúdo. Esse artigo investiga os desafios de descobrir *retweets* que contêm comentários ou edições, e apresentam uma abordagem de utilização das características estruturais/comportamentais dos usuários do X, para adquirir rótulos para tais comentários, o que permitiria o treinamento de um classificador supervisionado. Ao invés de depender apenas do texto dos comentários, que são geralmente curtos e pouco informativos, os autores combinam essa informação com dados sobre a rede social dos usuários (quem segue quem, *retweets*, uso de *hashtags*) para inferir a orientação política e o nível de concordância entre os usuários. Essa abordagem permite criar um conjunto de treinamento mais robusto, mesmo com poucos dados textuais, e explorar a dinâmica de interação entre grupos políticos diferentes. Os resultados preliminares do trabalho de Mustafaraj e Metax indicam que essa técnica é promissora para identificar padrões de concordância e desacordo em discussões online, contribuindo para uma melhor compreensão da polarização política nas redes sociais.

Ibrahim et al. (2015) investigaram o uso do X para prever resultados eleitorais na Indonésia, combinando detecção de *bots* e análise de sentimentos. O estudo focou na remoção de *tweets* gerados por *bots* e usuários pagos para aprimorar a qualidade dos dados. A identificação de contas automatizadas foi baseada em observação empírica e aprendizado de máquina, detectando características como alta frequência de *tweets*, criação recente da conta e grande número de URLs mencionados nos últimos 100 *tweets*. Para automatizar esse processo, os pesquisadores treinaram um modelo de aprendizado de máquina que classificava contas entre "*buzzers*" e "*non-buzzers*", alcançando uma precisão superior a 86%.

Após a filtragem dos *bots*, a análise de sentimentos foi aplicada para classificar *tweets* e prever tendências eleitorais, demonstrando maior precisão que métodos tradicionais de pesquisa. No entanto, os autores ressaltam que a curta extensão dos *tweets* e a ausência de indicadores claros de sentimento podem dificultar a classificação. O estudo destacou a relevância da qualidade dos dados na análise de sentimentos e estabeleceu uma base para a utilização das redes sociais na previsão de eventos políticos. (Ibrahim et al., 2015).

O uso de *bots* nas mídias sociais tornou-se uma prática comum para automatizar interações e amplificar mensagens. No entanto, essa proliferação de *bots* trouxe preocupações, especialmente em relação à manipulação de debates públicos, à disseminação de desinformação e à influência na opinião pública. Para compreender o impacto dos *bots*, é crucial estudar como eles operam e como podem ser detectados. Estudos iniciais forneceram as bases para frameworks de detecção de *bots*. Dentre eles, Varol et al. (2017) apresentaram um *framework* para detectar *bots* no X, utilizando mais de mil características extraídas dos dados públicos e metadados dos usuários, incluindo conteúdo de *tweets* e padrões de sentimento. A pesquisa mostrou que entre 9% e 15% das contas ativas no X eram *bots* e destacou a capacidade dos *bots* de interagir de maneira sofisticada com usuários humanos. A análise de sentimentos foi uma das ferramentas utilizadas para identificar essas interações, revelando estratégias de *retweet* e menção usadas pelos *bots* para influenciar grupos-alvo (Varol et al., 2017).

Em 2018, Bello et al. (2018) aplicaram uma técnica de engenharia reversa para entender o comportamento de *bots* no X. O estudo propõe o conceito de análise de sentimentos diferencial como uma forma de compreender o comportamento dos bots em relação aos tópicos de sua rede, considerando tanto suas fontes de informação

(amigos) quanto seu público-alvo (seguidores) fornecendo insights sobre como os *bots* moldam a opinião pública e afetam o comportamento dos usuários.

Com o aumento da presença de *bots*, surgiram diferentes abordagens técnicas para sua detecção. Inicialmente, os métodos eram baseados em regras, mas, com o avanço do aprendizado de máquina, técnicas mais sofisticadas começaram a ser aplicadas. Modelos baseados em redes neurais e classificadores se tornaram ferramentas poderosas para a detecção de *bots* em grandes volumes de dados. Kudungunta e Ferrara (2018) apresentaram uma abordagem baseada em redes neurais profundas para a detecção de *bots*, que integrou características contextuais e de metadados dos usuários (data de criação da conta, número de seguidores e padrões de atividade da conta) com a análise de sentimentos. Essa abordagem dupla permitiu que o modelo aproveitasse uma gama mais ampla de informações, aprimorando sua capacidade de classificar os *tweets* com precisão. Segundo os autores, os metadados, embora sejam um preditor fraco por si só, contribuem significativamente para reduzir a taxa de erro quando combinados com o conteúdo do *tweet*. O uso de redes neurais LSTM (*Long Short-Term Memory*) permitiu a detecção de *bots* a partir de um único *tweet* com alta precisão (AUC > 99%), mostrando um avanço significativo na identificação de *bots* a partir de dados de texto e metadados.

Com o avanço das técnicas de aprendizado de máquina e inteligência artificial, os *bots* se tornaram mais sofisticados, aumentando os desafios para sua detecção. Os métodos baseados em aprendizado não supervisionado surgem como alternativas promissoras para enfrentar esses desafios e melhorar a precisão da distinção entre humanos e *bots*. Inuwa-Dutse et al. (2018) focaram na análise lexical dos *tweets* para distinguir entre *bots* e humanos. A pesquisa revelou que características lexicais, como diversidade e sofisticação lexical, ajudaram a melhorar a precisão na identificação de *bots*, oferecendo uma nova perspectiva sobre como o conteúdo dos *tweets* pode ser usado para detectar automatizações. O estudo identifica características lexicais, como diversidade lexical, proporção tipo-*token*, e uso de contrações e *emoticons*, como sinais distintivos entre humanos e *bots* no X. Humanos tendem a apresentar maior diversidade lexical, refletindo conteúdos mais variados e engajamento emocional e intelectual. No entanto, algumas contas automatizadas, especialmente aquelas de organizações ou figuras populares, podem ter alta diversidade lexical para engajar um grande público. Além disso, o uso de *emoticons* é mais frequente em contas de *bots*, indicando uma preferência por símbolos visuais em vez de uma linguagem variada com os humanos que podem expressar sentimentos por meio de um vocabulário mais rico (Inuwa-Dutse et al., 2018).

Bello e Heckel (2019) analisaram o comportamento de *bots* no contexto político do Brexit. Eles investigaram estratégias usadas por *bots* para influenciar o debate político, particularmente em relação à independência escocesa e a um segundo referendo do Brexit. Utilizando uma abordagem inovadora para engenharia reversa dos *bots*, descobriram que os *bots* tinham estratégias bem definidas para afetar a opinião pública e demonstraram a relevância da análise de sentimentos para entender a influência dos *bots* em contextos políticos. Segundo os autores os *bots* geralmente exibem um viés no sentimento contrário dos sentimentos dos seus seguidores. Esse contraste de sentimento indica uma abordagem estratégica dos *bots* para influenciar seu público.

Ainda sobre a influência na opinião pública, Liu (2019) examinou a influência dos *bots* nas marcas e nas informações relacionadas a elas, usando uma abordagem de *big data* e análise de sentimentos para identificar como os *bots* manipulam a percepção das marcas. O estudo destacou a eficácia das técnicas de detecção de *bots* e a necessidade de distinguir opiniões autênticas das criadas artificialmente. A análise indicou que o sentimento positivo, postados por *bots* sociais, não teve um impacto significativo na viralidade dos *tweets*. Em contraste, descobriu que o sentimento negativo tem um efeito significativo na proliferação da postagem (viralidade), ou seja, *tweets* com conteúdo negativo tinham maior probabilidade de serem compartilhados e disseminados amplamente. Essa descoberta destaca que os *bots* sociais podem aproveitar as emoções negativas para aumentar seu alcance e influência no X reforçando a ideia de que mensagens negativas são uma ferramenta poderosa para *bots* sociais. O estudo descobriu que mensagens neutras não foram eficazes para motivar o compartilhamento entre os usuários. Isso sugere que os *tweets* neutros não envolvem os usuários da mesma forma que os *tweets* positivos ou negativos, o que os torna menos propensos a se tornarem virais. As descobertas indicam que os *bots* sociais têm mais sucesso em divulgar informações quando utilizam sentimentos negativos em vez de sentimentos positivos ou neutros. Em resumo, o estudo revelou que, embora o sentimento positivo não aumente significativamente a possibilidade de viralizar, o sentimento negativo desempenha um papel crucial na disseminação de informações por *bots* sociais, com o sentimento neutro sendo ineficaz no engajamento dos usuários.

Com a consolidação cada vez mais próximas das criptomoedas em 2020, Kraaijeveld e De Smedt (2020) exploraram o poder preditivo do sentimento público no X para prever preços de criptomoedas. Embora o foco principal fosse a previsão financeira, o estudo identificou a presença de *bots* e a influência deles nas discussões sobre criptomoedas, ressaltando a relevância da análise de sentimentos para detectar a manipulação de mercado por *bots*.

Em 2020 o contexto político impulsionou pesquisas como as desenvolvidas por Uyheng e Carley (2020), Pastor-Galindo et al. (2020) e Tavazoe et al. (2020). Uyheng e Carley (2020) ampliaram a análise de sentimentos para contextos geopolíticos diversos, investigando o impacto dos *bots* nas eleições na Ásia-Pacífico. A pesquisa mostrou variações no impacto dos *bots* nas mudanças de sentimento público e estrutura das comunidades online, contribuindo para a literatura sobre desinformação online e técnicas de análise de sentimentos. Os estudos de Pastor-Galindo et al. (2020) e Tavazoe et al. (2020) reforçam a importância da análise de sentimentos na detecção de *bots* e na identificação de padrões de comportamento malicioso. Pastor-Galindo et al. (2020) apresentaram um conjunto de dados valioso para a pesquisa sobre *bots* políticos, enquanto Tavazoe et al. (2020) utilizaram a análise de sentimentos para distinguir entre *bots* e usuários legítimos, aplicando modelos de classificação supervisionada para melhorar a detecção de *bots*.

Anwar e Yaqub (2020) usaram aprendizado não supervisionado para detectar *bots* durante as eleições canadenses de 2019, aplicando análise de sentimentos como uma ferramenta adicional para isolar contas automatizadas. Sua abordagem inovadora revelou a eficácia da análise de sentimentos em combinação com técnicas de *clustering* para a identificação de *bots*.

No ano de 2021, destacam-se trabalhos voltados para investigação sobre o interesse dos usuários das mídias sociais em torno do COVID-19, tais como o desenvolvido por Graham et al. (2021). Que exploraram a dinâmica polarizada das discussões no X sobre as restrições ao COVID-19 em Victoria, Austrália. Eles utilizaram uma abordagem mista que combinou análise estatística, análise de sentimentos e detecção de *bots*. Este estudo revelou como *bots* e atores hiper-partidários manipularam conversas e ampliaram suas mensagens por meio de técnicas específicas, destacando a utilidade da análise de sentimentos para identificar e entender a influência de *bots* em campanhas políticas.

A análise de sentimentos tem se destacado como uma ferramenta valiosa para diferenciar interações humanas de *bots* nas mídias sociais. Ao estudar os sentimentos expressos nas postagens, os pesquisadores podem identificar padrões emocionais típicos de humanos que muitas vezes estão ausentes ou são replicados de maneira artificial por *bots*. Isso pode ser fatal durante crises, como emergências de saúde pública, onde a disseminação de informações confiáveis é crítica. *Bots* podem distorcer o debate e disseminar desinformação, tornando a análise de sentimentos uma ferramenta crucial para monitorar a qualidade da informação que circula, especialmente em tópicos científicos e técnicos. Dentro dessa narrativa, Yousefinaghani et al. (2021) analisaram sentimentos e opiniões sobre vacinas COVID-19, descobrindo diferenças nas discussões entre contas de *bots* e usuários humanos. A análise revelou que a oposição à vacinação frequentemente vinha de contas de *bots* ou ativistas políticos, enquanto figuras conhecidas e organizações apoiavam a vacinação. Esse estudo evidenciou como a análise de sentimentos pode ajudar a distinguir entre fontes de informação confiáveis e não confiáveis.

Ainda em 2021, Chen et al. (2021) focaram em entender a imagem de países durante a pandemia de COVID-19, usando análise de sentimentos baseada em aspectos (os autores definiram sete categorias diferentes de aspectos, a saber: política, economia, relações exteriores, cultura, situação epidêmica, medidas antiepidêmicas e racismo em relação à China) para examinar como a imagem da China foi afetada. A pesquisa revelou variações significativas nas atitudes de diferentes grupos de usuários, incluindo *bots*. O estudo demonstrou como a análise de sentimentos pode ser aplicada para diferenciar entre respostas humanas e automatizadas, fornecendo descobertas sobre o impacto dos *bots* na percepção pública. Eles descobriram que os *bots* sociais tendiam a expressar mais sentimentos negativos em relação à situação epidêmica, às medidas antiepidêmicas e ao racismo em comparação com usuários humanos. Essa distinção na expressão de sentimentos destacou as diferentes motivações e comportamentos entre contas automatizadas e usuários humanos.

Ainda acerca do COVID, em 2022, Suarez-Lledo e Alvarez-Galvez (2022) investigaram o papel dos *bots* durante a pandemia de COVID-19. Utilizando a API do X e o Botometer para classificar contas, eles identificaram diferenças notáveis no conteúdo e no tom dos *tweets* entre contas de *bots* e humanas. *Bots* tendiam a disseminar informações críticas e questionadoras, frequentemente com um tom negativo, enquanto as contas humanas apresentavam uma abordagem mais positiva e informativa, sublinhando como os *bots* podem influenciar a percepção pública durante crises de saúde.

Galgoczy et al. (2022) aplicaram a análise de sentimentos ao examinar o papel dos *bots* no primeiro impeachment do presidente Donald Trump. Através da análise de sentimento e métricas de NLP, descobriram que *bots* contribuíam significativamente para um tom negativo na discussão e estavam direcionados principalmente contra democratas. A pesquisa também revelou que *bots* utilizavam fontes fortemente partidárias, reforçando a ideia de que *bots* são usados como ferramentas de propaganda computacional para moldar debates políticos.

Entre os artigos mais atuais nessa temática, destaca-se o trabalho escrito por Cai et al. (2023) que abordam a influência dos *bots* sociais na difusão de informações durante emergências de saúde pública, no caso, a pandemia de COVID-19. Utilizando métodos de aprendizado de máquina e regressão causal, o estudo revelou que os *bots* sociais são mais propensos a disseminar informações com sentimentos negativos do que positivos. A análise dos dados coletados no Weibo e a comparação com informações relacionadas a surtos em diferentes cidades mostraram que os *bots* desempenham um papel relevante na transmissão de tópicos específicos, afetando a mecânica da difusão de informações na rede social. Além disso, a análise de sentimentos e a aplicação de modelos de tópicos como o *Structural Topic Model* (STM) ajudaram a identificar as diferenças entre os conteúdos disseminados por humanos e *bots*.

Babur et al. (2023) focam na identificação de contas de *bots* no X usando técnicas de engenharia de características e aprendizado de máquina. O estudo coletou *tweets* e metadados para criar um conjunto de dados robusto e utilizou técnicas de análise de sentimentos para identificar características relevantes dos *tweets* dos *bots*. A análise comparativa de três modelos diferentes demonstrou a eficácia de características combinadas, como a análise de sentimentos e a proporção de amigos/seguidores, na identificação de contas de *bots*. Esse trabalho destaca o potencial da engenharia de características associada à análise de sentimentos para melhorar a detecção de *bots*, estabelecendo uma base para pesquisas futuras.

A pesquisa de Bär et al. (2023), focou nas atividades de mídia social da Bellingcat e encontrou um aumento significativo na interação com *bots* durante eventos como a invasão russa da Ucrânia. A análise de sentimento revelou uma polarização crescente nas respostas dos seguidores, com *bots* contribuindo para um aumento no tom negativo.

Outro trabalho que aborda contexto de guerra é o apresentado em Breve et al. (2024) que conduziram uma análise da percepção global do conflito Rússia-Ucrânia no X, incluindo uma análise de sentimentos e detecção de *bots*. A pesquisa revelou uma correlação entre o volume de *tweets* e o sentimento predominante, com a detecção de muitos *bots* envolvidos na disseminação de *hashtags* pró-conflito. A análise de sentimentos destacou a prevalência de sentimentos negativos e a importância de métodos confiáveis para detectar *bots* que influenciam a opinião pública durante conflitos.

Mouronte-López et al (2024) exploraram os padrões de comportamento de humanos e *bots* em conversas sobre sustentabilidade no X. A análise combinou modelos matemáticos para identificar tipologias de contas e uma análise de sentimentos baseada em léxico. Os autores descobriram na pesquisa que, apesar de mensagens predominantemente positivas tanto de *bots* quanto de humanos, *bots* tendiam a ter uma maior proporção de mensagens neutras. Segundo as descobertas do artigo, os *bots* exibiram uma porcentagem maior de mensagens neutras em comparação com usuários humanos. Isso sugere que, embora os *bots* possam desenvolver conversas, eles geralmente o fazem sem fortes expressões emocionais, levando a um tom mais neutro em suas postagens. Além disso, os *bots* não apenas tiveram uma polaridade geral mais alta em suas postagens, mas também exibiram um número significativo de postagens com valores extremos de sentimento. Isso significa que, embora possam postar positivamente, eles também tendem a ter reações mais extremas, positivas ou negativas, em comparação com usuários humanos.

Esses estudos demonstram a evolução e a importância crescente da análise de sentimentos na identificação e compreensão dos *bots* em plataformas de mídia social. Desde a análise de debates políticos e questões de saúde até a previsão de preços financeiros, a capacidade de distinguir entre interações humanas e automatizadas é crucial para mitigar os impactos negativos dos *bots* e melhorar a integridade das interações online.

Desde os primeiros estudos que utilizaram análise lexical e métodos de classificação Bayesiana (Velayutham & Tiwari, 2017), houve um avanço significativo para técnicas mais complexas, como redes neurais profundas (Kudugunta & Ferrara, 2018) e modelos de aprendizado não supervisionado (Wu et al., 2022). A integração de métodos avançados, como redes neurais LSTM e BERT, tem aprimorado a precisão na identificação de *bots* e na análise de sentimentos.

Os estudos revisados destacam a aplicação da análise de sentimentos em diversos contextos, desde previsões eleitorais Ibrahim et al. (2015) até debates sobre saúde pública e criptomoedas Gatkal et al. (2021). A análise de sentimentos tem demonstrado sua eficácia em distinguir entre interações humanas e automatizadas, contribuindo para a compreensão das influências dos *bots* em diferentes domínios.

Os impactos dos *bots* são evidentes em várias áreas, incluindo a polarização política e a manipulação de percepções públicas (Çetinkaya, 2020; Zhang et al., 2023). Contudo, desafios persistem, como a adaptação das técnicas para diferentes idiomas e a identificação de *bots* cada vez mais sofisticados. Estudos recentes têm abordado essas questões com abordagens inovadoras, como a utilização de léxicos aprendidos automaticamente e modelos híbridos (Chawla & Kapoor, 2023).

O futuro da pesquisa na análise de sentimentos e identificação de *bots* promete avanços com a integração de modelos de linguagem de grande escala e técnicas de aprendizado profundo. A contínua evolução das técnicas de detecção e análise permitirá uma maior compreensão e mitigação da influência dos *bots* nas redes sociais.

A revisão demonstra que a combinação de análise de sentimentos com técnicas avançadas de detecção de *bots* é importante para entender e controlar o impacto desses agentes nas mídias sociais. Apesar das descobertas dos estudos oferecerem uma base sólida para futuras pesquisas e práticas na área observa-se que o problema de pesquisa persiste sem solução. Neste contexto, observa-se que poucas pesquisas exploram e avaliam a eficácia de novas ferramentas de Inteligência Artificial Conversacional, como Chat GPT e Gemini, na classificação de sentimentos, apresentando assim uma lacuna a ser explorada. Outros trabalhos foram analisados e em outras buscas exploratórias, porém, não foram encontrados nenhum que analisasse dados de circulação da produção científica, ou seja, que contribuíssem com a identificação de *bots* por meio de análise de sentimentos em dados altmétricos.

Neste cenário, o presente estudo avança nas pesquisas sobre a disseminação da produção científica ao propor uma abordagem para a identificação de *bots* no contexto de dados altmétricos, utilizando análise de sentimentos e *Large Language Models* (LLMs). O estudo ganha importância ao avaliar a eficácia de LLMs, especificamente Gemini e ChatGPT, na classificação de sentimentos aplicados à dados altmétricos. O recente avanço dessas ferramentas representa uma oportunidade para aprimorar métodos de detecção de *bots*, mas também exige uma análise de suas limitações e vieses. Assim, este trabalho não apenas amplia as possibilidades metodológicas no campo da altmetria, mas também contribui para o refinamento do uso de LLMs na ciência da informação, destacando desafios e potenciais aplicações futuras.

3 Metodologia

O conjunto de dados em análise inclui uma amostra de postagens no X que fizeram menção ao editorial, publicado no periódico “*Dying in a Leadership Vacuum*” em 7 de outubro de 2020, com o título “*Dying in a Leadership Vacuum*” (DOI: 10.1056/NEJMe2029812). Cabe destacar que essa publicação é da área de medicina e saúde, com foco no COVID e está listada entre as que possuem maiores *scores* altmétricos na plataforma Altmetric.

Essas características podem, hipoteticamente, representar comportamentos sociais, incluindo a intensidade do uso de bots. Portanto, como se trata de assunto que tem forte influência política e interesse social, onde o negacionismo científico se fez presente, os comportamentos observados nesse conjunto de artigos e respectivas contas no Twitter, provavelmente contam com dinâmicas peculiares de adesão social que podem ser diferentes em outros contextos temáticos e tipos de publicação. (Pontes & Maricato, 2023a)

Ressalta-se que Pontes e Maricato (2023a) classificaram, como *bots* e humanos, os perfis que comentaram essa publicação no X, através do Botometer e disponibilizaram os dados classificados em um *dataset*. A classificação realizada pelo Botometer é baseada em aprendizagem de máquina supervisionada, que é supostamente capaz de distinguir contas similares a *bots* e contas similares a humanos, com base em seus recursos ou características (Pontes & Maricato, 2023a).

Uma vez que as postagens não foram disponibilizadas no *dataset* publicado por Pontes e Maricato (2023a), foi realizada a coleta de uma amostra de postagens do X que fizeram menção à publicação. Os posts foram extraídos da plataforma Altmetric através de um algoritmo em Python 3.12 usando a biblioteca Beautiful Soup 4.12 e o ambiente de desenvolvimento Google Colab. Com isso, foi gerado um conjunto de dados composto por 9.792 postagens do X que comentaram especificamente o referido editorial. Dessas postagens, foram identificados 5.601

perfis únicos, que foram confrontados com os perfis classificados e disponibilizadas no *dataset* gerado por Pontes e Maricato (2023a). Das contas que continham a classificação (*bot* ou humano) foram selecionadas 41 contas que fizeram mais de 4 postagens.

Segundo o *dataset* disponibilizado por Pontes e Maricato (2023b), 10 contas foram classificadas pelo botometer como *bot* e 31 como humanas. Considerando que Pontes e Maricato (2023a) citam a limitação na classificação das contas da rede de atenção atômica pelo Botometer foi realizada uma classificação manual das 41 contas selecionadas. A classificação manual das contas selecionadas tomou como critério a análise da quantidade de postagens, o horário da postagem, o intervalo de tempo entre as postagens, a data de criação das contas, a imagem da foto. Nesta classificação manual, foram identificados que 20 contas são *bots* e 21 contas são humanas. No Quadro 1 essas diferenças podem ser observadas.

As contas classificadas realizaram um total de 3.493 postagens¹. De acordo com a classificação manual das contas, 20 *bots* foram responsáveis por 86% das postagens totalizando 3004 posts. As 21 contas classificadas como humanas totalizaram somente 489 postagens representando 14% do total. Daqui para frente são apresentadas somente a classificação manual nas análises.

Quadro 1

Classificação de contas manual x Botometer

Contas	Bot manual	Humana manual	Bot botometer	Humana botometer	Quantidade posts
CONTA01	4		4		4
CONTA02	5		5		5
CONTA03		7	7		7
CONTA04	7		7		7
CONTA05	7		7		7
CONTA06	9		9		9
CONTA07	12		12		12
CONTA08	18		18		18
CONTA09	27		27		27
CONTA10	66		66		66
CONTA11	30			30	30
CONTA12		9		9	9
CONTA13		58		58	58
CONTA14		9		9	9
CONTA15	24			24	24
CONTA16	12			12	12
CONTA17		8		8	8
CONTA18		64		64	64
CONTA19	7			7	7
CONTA20		24		24	24
CONTA21	9			9	9
CONTA22		12		12	12
CONTA23	2071			2071	2071
CONTA24		14		14	14
CONTA25	592			592	592
CONTA26		11		11	11
CONTA27	8			8	8
CONTA28	68			68	68

¹ Os dados das 41 contas com as 3.493 postagens e todos os resultados das classificações realizadas estão disponíveis no link: <https://zenodo.org/records/14919674>.

Contas	Bot manual	Humana manual	Bot botometer	Humana botometer	Quantidade posts
CONTA29	8			8	8
CONTA30		16		16	16
CONTA31		44		44	44
CONTA32		8		8	8
CONTA33		8		8	8
CONTA34		47		47	47
CONTA35		49		49	49
CONTA36		10		10	10
CONTA37		18		18	18
CONTA38	20			20	20
CONTA39		34		34	34
CONTA40		27		27	27
CONTA41		12		12	12
Quantidade de posts	3.004	489	162	3.331	3.493
Percentagem	86%	14%	5%	95%	
Total bot/humana	20	21	10	31	
Percentagem	49%	51%	24%	76%	

Nota. Fonte: Elaborado pelos Autores (2024). [Descrição do Quadro] Quadro com 41 linhas representando contas na plataforma X (antigo Twitter) e 6 colunas. As colunas estão divididas em: “CONTAS”, “MANUAL – BOT”, “MANUAL – HUMANA”, “BOTOMETER – BOT”, “BOTOMETER – HUMANA” e “QTD POSTS”. Os dados indicam a quantidade de postagens de cada conta classificadas como automatizadas (bots) ou humanas, segundo avaliação manual e por Botometer. As últimas três linhas da tabela trazem totais e percentuais: 86% das postagens foram atribuídas a bots pela avaliação manual, mas apenas 5% segundo o Botometer. Já a classificação total de contas mostra 49% bots e 51% humanas na avaliação manual, contra 24% bots e 76% humanas no Botometer. [Fim da descrição].

Por fim, os sentimentos expressos nessas 3.493 postagens foram classificados utilizando o ChatGPT 3.5 da OpenAi e o Gemini 1.5 Flash da Google. A Análise de sentimentos dos posts foram realizados utilizando o *Prompt* apresentado na Figura 1.

O *Prompt* utilizado consiste em cinco itens e duas variáveis. As variáveis *abstract* e *listadepost* são usadas para que o *Prompt* possa ser replicado, ou seja, utilizado repetidamente para analisar outros posts. No item 1 é dado a instrução para classificação de polaridade dos sentimentos entre Positivo, Negativo e Neutro. O item 2 informa o formato de input da lista de post que será usado. O item 3 indica como o resultado deve ser apresentado. O item 4 é usado para informar o conteúdo da variável *abstract* e o item 5 é usado para informar o conteúdo da lista de posts. Devido ao limite de caracteres do *Prompt*, essas etapas foram repetidas para a classificação de conjuntos de 100 posts por vez.

Figura 1

Prompt utilizado para classificação de sentimentos em publicações científicas de postagens da rede X nas ferramentas Gemini e ChatGPT

1. Forneça uma avaliação classificando o sentimento como Positivo, Negativo ou Neutro de cada post da **[listadeposts]**, com base no **[abstract]**
2. A lista de posts é formada por: "codigodopost : Post"
3. Apresente os resultados indicando o codigodopost e a classificação gerada com a seguinte formatação "codigopost:classificacao"
4. Abstract: Covid-19 has created a crisis throughout the world. This crisis has produced a test of leadership. With no good options to combat a novel pathogen, countries were forced to make hard choices about how to respond. Here in the United States, our leaders have failed that test. They have taken a crisis and turned it into a tragedy. The magnitude of this failure is astonishing. According to the Johns Hopkins Center for Systems Science and Engineering,¹ the United States leads the world in Covid-19 cases and in deaths due to the disease, far exceeding the numbers in much larger countries, such as China. The death rate in this country is more than double that of Canada, exceeds that of Japan, a country with a vulnerable and elderly population, by a factor of almost 50, and even dwarfs the rates in lowermiddle-income countries, such as Vietnam, by a factor of almost 2000. Covid-19 is an overwhelming Challenger, and many factors contribute to its severity. But the one we can control is how we behave. And in the United States we have consistently behaved poorly.
5. Listadeposts:

Nota. Fonte: Elaborado pelos Autores (2024). *[Descrição da imagem]* Imagem contendo uma tela com cinco instruções numeradas, escritas em português, relativas à análise de sentimentos de postagens em redes sociais. A instrução 1 solicita uma avaliação classificando o sentimento das postagens como Positivo, Negativo ou Neutro, com base em um resumo indicado como **[abstract]**. A instrução 2 define o formato da lista de posts como pares no modelo "codigodopost: Post". A instrução 3 solicita a apresentação dos resultados no formato "codigopost:classificacao". A instrução 4 apresenta o conteúdo do resumo, em inglês, descrevendo o impacto da pandemia de Covid-19 nos Estados Unidos, com forte crítica à liderança do país. Por fim, a instrução 5 menciona a variável **[listadeposts]**. O texto está organizado em estrutura clara, com numeração, e faz referência a variáveis em colchetes. *[Fim da descrição]*.

Nas duas ferramentas analisadas, a classificação do sentimento dos posts tomou como base o contexto geral do artigo apresentado no *abstract*. Segundo o entendimento das IAs o *abstract* descreve a crise global causada pela Covid-19 e critica a resposta dos líderes nos Estados Unidos, rotulando-a como um fracasso que resultou em uma tragédia. O tom geral é altamente negativo em relação à gestão da pandemia nos EUA. Com base nisso, as ferramentas atribuíram uma classificação de sentimento a cada post da seguinte maneira: Positivo: Se o post elogiar a resposta dos líderes dos EUA à pandemia ou defender ações positivas tomadas. Negativo: Se o post criticar a resposta dos líderes dos EUA à pandemia ou expressar descontentamento com as ações tomadas. Neutro: Se o post não expressar claramente uma opinião sobre a resposta dos líderes dos EUA à pandemia.

Após a classificação realizada pelas ferramentas observou-se que 63 postagens não tiveram polaridade identificada pelo Gemini e em 315 postagens as IAs discordavam. Considerando esses casos foi realizada uma classificação manual, por um dos autores deste artigo, considerando os seguintes critérios de classificação:

1) Positivo:

- a) Presença de palavras e frases positivas – Termos como "prestigious", "agrees", "powerful words", "save lives", e "I will end this" sugerem apoio, confiança e otimismo.
- b) Apoio explícito a uma causa ou figura política – Muitas postagens demonstram suporte a Joe Biden e ao artigo.
- c) Hashtags e expressões motivacionais – A hashtag #VoteBlueToEndTheNightmare sugere um tom de mobilização e esperança, o que pode ser associado a um sentimento positivo dentro do contexto político.
- d) Tom encorajador e inspirador – Frases como "More than ever, our vote can save lives." e "I will make sure we have a plan." transmitem um senso de propósito e otimismo.

2) Negativo:

- a) Uso de Palavras com Carga Negativa - Expressões como "fail", "tragedy", " Hands are COVERED in BLOOD " entre outras, transmitem forte conotação negativa.
- b) Crítica Direta a Pessoas ou Instituições - As postagens contêm ataques diretos a figuras políticas e partidos, especialmente ao ex-presidente Donald Trump e ao Partido Republicano, enfatizando falhas na gestão da pandemia.
- c) Contexto de Perda e Sofrimento - Muitas mensagens citam números elevados de mortes e apontam responsabilidade política, o que reforça a percepção de negatividade.
- d) Tom de Urgência e Indignação - Uso de letras maiúsculas, pontos de exclamação e frases imperativas como "Vote como se sua vida dependesse disso!" aumentam o tom emocional negativo.
- e) Polarização e Conflito - As postagens frequentemente estabelecem uma oposição entre grupos políticos e atribuem culpa a um lado específico, o que intensifica a negatividade.

3) Neutro:

- a) Ausência de palavras fortemente carregadas emocionalmente;
- b) Apenas compartilhamento - Posts compostos apenas por identificação do artigo (título, doi ou link) incluindo, ou não, marcação de contas (@).

4 Resultados e discussão

No total, foram analisados os sentimentos de 3.493 posts que mencionaram a publicação científica, utilizando as ferramentas ChatGPT da OpenAI e Gemini da empresa Google, como já foi detalhado na seção anterior. Segundo as duas tecnologias utilizadas, a grande maioria das postagens eram de conotação negativa (92,3%). Os posts neutros representaram em média 6,2%, enquanto os positivos foram em média 0,6%. As postagens não identificadas pelo Gemini representam 0,9%.

Em relação a diferença de desempenho entre as ferramentas LLMs ChatGPT 3.5 e Gemini 1.5 Flash ao classificar os sentimentos de postagens realizadas no X sobre artigos científicos, observou-se um desempenho muito semelhante na classificação dos posts negativos.

No contexto geral, os classificadores concordaram em 3.178 posts mostrando uma conformidade de 90,98% em relação a análise de sentimentos realizada pelas IAs conversacionais avaliadas. Entretanto, na classificação de sentimentos pelo Gemini observa-se que essa IA não conseguiu identificar o sentimento em 63 postagens, o que representa uma falha significativa em sua capacidade de processamento. As justificativas fornecidas pelo Gemini incluem mensagens como "*unclassifiable*" e "*sentiment unclear (limited by retweet)*", além de uma série de respostas que indicam a incapacidade da ferramenta de realizar a tarefa, atribuída ao fato de ser um modelo de linguagem limitado. Avaliando manualmente cada uma dessas 63 postagens, observou-se que, no geral, os posts continham apenas o link do artigo. Isso sugere uma limitação técnica, onde o modelo não é capaz de lidar com certos tipos de textos ou contextos, como *retweets* ou postagens complexas. Esse resultado se assemelha ao citado por Ibrahim et al., (2015) que enfatiza a dificuldade de classificar com precisão os sentimentos de *tweets* muito curtos ou que não tem indicadores claros de sentimento. Essa falha pode ter impacto importante na qualidade da análise, especialmente em cenários onde a precisão da classificação emocional é crítica para a tomada de decisões, como análise de opinião pública ou comportamento eleitoral.

O ChatGPT, por sua vez, classificou as 63 postagens que o Gemini não conseguiu classificar. De acordo com os resultados da classificação do ChatGPT, 43 foram classificadas como negativas e 20 como neutras. Considerando essa distorção foi realizada uma análise manual, por um dos autores do artigo, com o intuito de entender melhor as limitações do Gemini.

Analisando manualmente as 63 postagens verificou-se que o GPT acertou 85% das classificações, ou seja, ele foi capaz de identificar corretamente o sentimento de 54 postagens. A análise manual encontrou 41 postagens negativas, 19 neutras, 3 positivas e 1 não identificável. Dentro deste conjunto de 63 postagens não classificadas pelo Gemini, a precisão de 85% do GPT é bastante expressiva, indicando que ele tem uma boa capacidade de identificação de sentimentos em uma ampla gama de contextos, mostrando uma eficácia que se mostra mais relevante do que aquela demonstrada pelo Gemini. No entanto, há margem para melhoria, especialmente no que diz respeito à distinção entre sentimentos neutros e positivos, que parece ser mais desafiadora para o modelo.

Ao comparar o resultado do ChatGPT na classificação dessas 63 postagens com a análise manual observou-se uma discrepância no resultado, principalmente na identificação de posts positivos (3 positivos na análise manual contra 0 pelo ChatGPT) e em uma postagem cujo sentimento não estava claro. Esses dados sugerem que o ChatGPT tende a classificar como neutras postagens que poderiam ser percebidas como positivas, demonstrando uma certa "cautela" no reconhecimento de positividade. Essa diferença entre a análise automática e a manual também pode estar relacionada a nuances contextuais que os modelos automáticos ainda não conseguem captar completamente, como sarcasmo, ironia, ou o tom implícito de certos textos, principalmente em plataformas como X, onde a comunicação é muitas vezes ambígua e curta.

Considerando esses aspectos, o Gemini, quando comparado ao GPT, demonstrou uma dificuldade maior na análise de posts do X sobre o artigo, principalmente na identificação de nuances e ironias presentes em comentários complexos. A combinação de emojis, sarcasmo e referências indiretas, como aparece nos posts não classificados, gerou uma dificuldade para o Gemini, que pode não processar completamente a mensagem em sua totalidade. Essa limitação pode ser atribuída a um treinamento com base em um conjunto de dados menos robusto e diversificado, resultando em menor capacidade de inferência e uma tendência a analisar palavras isoladamente, em vez do contexto geral da mensagem. Consequentemente, o Gemini pode ter dificuldades em captar o sentimento geral do autor do *tweet*, especialmente quando este se expressa de forma sutil ou ambígua.

Refletindo sobre a eficácia dos LLMs para classificação de sentimentos, é possível observar que o fato de o ChatGPT atingir 85% de precisão, dentro do conjunto de posts não identificados pelo Gemini, mostra o potencial da IA para tarefas de análise de sentimento, mas também destaca a importância de uma supervisão humana, especialmente em contextos sensíveis. Modelos de IA podem fornecer uma base rápida e eficaz para análise em larga escala, mas, devido a suas limitações, é crucial que haja verificação manual para garantir a precisão. A falha do Gemini em classificar um número significativo de postagens também aponta para a necessidade de modelos mais robustos e flexíveis, capazes de lidar com uma maior variedade de formatos de conteúdo, como repostagens, mensagens curtas ou postagens complexas.

Para finalizar a análise a respeito da eficácia das ferramentas LLMs ChatGPT e Gemini, ao classificar os sentimentos de postagens realizadas no X, observou-se que, em 315 posts os classificadores divergiram em relação ao sentimento. Para entender a procedência da divergência os 315 posts foram analisados e classificados manualmente por um dos autores do artigo. O Gráfico 1 apresenta o resultado dessa classificação manual onde foi contabilizado que o Gemini acertou 177 sentimentos e o GPT acertou 119. Ao comparar a eficácia dos dois LLMs, observa-se que ambos possuem boa capacidade em classificar sentimentos negativos, entretanto, dentro das classificações divergentes, observou-se que o Gemini foi mais eficaz na classificação.

Gráfico 1

Conferência manual de sentimentos dos posts que divergiram na classificação Gemini x GPT



Nota. Fonte: Elaborado pelos Autores (2024). [Descrição do gráfico] Gráfico de duas colunas empilhadas, na cor azul escuro para acertou e na cor laranja para errou. Exibe o desempenho de dois modelos de inteligência artificial — GEMINI e GPT — na tarefa de classificação de sentimentos. O eixo vertical representa a quantidade total de

classificações. Cada modelo é representado por uma coluna composta por duas cores: azul escuro para os acertos e laranja para os erros. A coluna do modelo GEMINI mostra 177 acertos e 138 erros. Já a coluna do modelo GPT exibe 119 acertos e 196 erros. A comparação indica que o modelo GEMINI teve melhor desempenho geral, com maior número de acertos e menor número de erros do que o modelo GPT. *[Fim da descrição].*

Nas classificações realizadas pelo ChatGPT, observa-se que a maioria das postagens feitas por bots apresentaram sentimento negativo e uma quantidade relevante de postagens foi classificada com sentimento neutro. Um número limitado de postagens positivas foi identificado.

O modelo Gemini também classificou a maioria das postagens de bots como negativas, com números próximos aos do ChatGPT, como pode ser observado na Figura 2. Houve uma menor quantidade de postagens com sentimento neutro em comparação ao ChatGPT. Assim como o GPT, o Gemini classificou poucas postagens como positivas.

Figura 2

Comparação da classificação de sentimento das postagens no X através do Chat GPT e Gemini por tipo de conta (bot e humana)

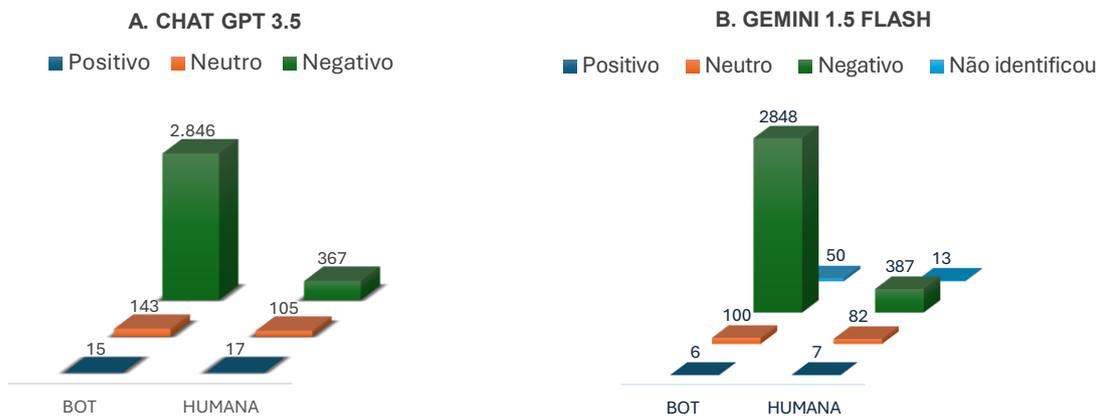


Tabela A

	GPT BOT		GPT HUMANA		GEMINI BOT		GEMINI HUMANA	
Negativo	2.846	81%	367	11%	2.848	82%	387	11%
Neutro	143	4%	105	3%	100	3%	82	2%
Positivo	15	0,4%	17	0,5%	6	0,2%	7	0,2%
Não identificado					50	1%	13	0,4%

Tabela B

	GPT BOT		GPT HUMANA		GEMINI BOT		GEMINI HUMANA	
Negativo	2.846	95%	367	75%	2.848	95%	387	79%
Neutro	143	5%	105	21%	100	3%	82	17%
Positivo	15	0,5%	17	3,5%	6	0,2%	7	1,4%
Não identificado					50	2%	13	2,7%
Total	3.004		489		3.004		489	

Nota. Percentual baseado no número de posts de cada classe. Fonte: Elaborado pelos Autores (2024). *[Descrição da imagem]* A imagem exibe dois gráficos de colunas tridimensionais lado a lado e duas tabelas comparativas abaixo, representando a distribuição de sentimentos (positivo, neutro, negativo e não identificado – azul escuro, laranja, verde e azul claro, respectivamente) atribuídos por dois modelos de inteligência artificial — ChatGPT 3.5 e Gemini 1.5 Flash — a postagens classificadas como provenientes de contas bots ou humanas. O gráfico à esquerda mostra os resultados do ChatGPT: entre os bots, 2.846 postagens foram negativas, 143 neutras e 15 positivas; entre humanos, 367 foram negativas, 105 neutras e 17 positivas. O gráfico à direita apresenta os dados do Gemini: para bots, 2.848 negativas, 100 neutras, 6 positivas e 50 não identificadas; para humanos, 387

negativas, 82 neutras, 7 positivas e 13 não identificadas. A Tabela A mostra os percentuais baseados no total de postagens, e a Tabela B apresenta os percentuais relativos ao total de postagens dentro de cada classe (bot ou humana). [Fim da descrição].

Nas classificações realizadas pelo ChatGPT a maioria das postagens feitas por humanos também foi classificada como negativa (73%). Comparativamente, as contas humanas apresentam uma quantidade menor de postagens neutras em relação aos *bots*, entretanto entre as postagens humanas as neutras representam 21% enquanto as neutras realizadas pelos *bots* representam 5%. Um número muito pequeno de postagens positivas foi registrado para contas humanas. O Gemini apresentou uma classificação similar ao ChatGPT, classificando a maioria das postagens humanas como negativas (79%).

Tanto *bots* quanto humanos mostram uma alta predominância de sentimentos negativos. No entanto, os *bots* parecem ser mais intensamente críticos. Os humanos tendem a postar mais conteúdo neutro em comparação com bots, o que pode indicar uma maior variabilidade ou moderação nas opiniões humanas em relação a artigos científicos. Tanto *bots* quanto humanos têm uma baixa quantidade de postagens positivas. Porém, o ChatGPT identificou um número ligeiramente maior de postagens positivas em contas humanas do que o Gemini.

No tocante aos sentimentos negativos, tanto ChatGPT quanto Gemini, mostram uma alta prevalência desse tipo de sentimento nas contas de *bots*, corroborando com os trabalhos Suarez-Lledo e Alvarez-Galvez (2022), Galgoczy et al. (2022), Cai et al. (2023) e Bär et al. (2023). Essa consistência sugere que, para o artigo avaliado, os *bots* estiveram envolvidos na disseminação de conteúdos de natureza crítica ou pessimista em relação à divulgação científica. No caso das contas humanas, apesar de também haver uma presença significativa de sentimentos negativos, principalmente quando analisadas pelo Gemini (63% negativos), há uma distribuição mais ampla, com uma quantidade considerável de postagens neutras e positivas em comparação com os *bots*.

Considerando os sentimentos neutros, é possível afirmar que a proporção de postagens neutras é significativamente maior nas contas humanas, tanto na análise do ChatGPT quanto na de Gemini. Isso sugere que as contas humanas tendem a compartilhar informações de maneira mais imparcial ou objetiva, sem inclinação emocional forte. *Bots*, por outro lado, mostram uma baixa porcentagem de postagens neutras. Isso pode indicar que os *bots* são programados para provocar reações emocionais mais intensas e polarizadas no público.

Com relação aos sentimentos positivos, o ChatGPT identificou um percentual muito baixo de positividade entre os *bots*, com apenas 1%. Já as contas humanas tiveram 5% de sentimentos positivos, mostrando que os humanos são mais propensos a transmitir mensagens otimistas ou de incentivo corroborando com Suarez-Lledo e Alvarez-Galvez (2022). Gemini não identificou nenhum sentimento positivo entre os *bots* e uma porcentagem muito baixa nas contas humanas (5%) como pode ser observado no Quadro 2.

Sobre os padrões de sentimentos que mais diferenciam as contas *bots* das contas humanas na rede social X no contexto da divulgação científica, os dados apresentados no Quadro 2 relacionam as 41 contas do X analisadas e o resultado da análise de sentimento de cada uma delas. Comparando os resultados das classificações pelas duas tecnologias, observa-se que o resultado de classificação dos *tweets* foram iguais para 10 contas humanas e para 5 contas classificadas como *bot*, sugerindo que as contas humanas são mais claras e objetivas em suas postagens. Quinze contas classificadas como *bot* apresentaram divergências entre as classificações realizadas pelo ChatGPT e Gemini demonstrando que as postagens realizadas por *bots* não possuem clareza em seus significados. Isso é possível observar na conta *Bot01* e *Bot05* que apresentam resultados de classificação discrepantes. Entre as contas classificadas como humanas observa-se que 11 contas tiveram classificações diferentes e 10 iguais.

Quadro 2

Contas selecionadas para análise

Conta	Chat GPT Negativo	Chat GPT Neutro	Chat GPT Positivo	Gemini Não identificado	Gemini Negativo	Gemini Neutro	Gemini Positivo
BOT01	6	45	15		62	4	
BOT02	10	20			10	20	
BOT03	5				4	1	
BOT04	22	2			24		
BOT05	10	2			2	10	
BOT06	10	2			8	4	
BOT07		7		6			1
BOT08	7				7		
BOT09		9			1	8	
BOT10	2062	9		22	2028	20	1
BOT11	565	27		15	559	17	1
BOT12	3	4		2	5		
BOT13	8			1	6	1	
BOT14	68				68		
BOT15	8			4	4		
BOT16		9			5	3	1
BOT17	27				27		
BOT18		4				4	
BOT19	17	1			13	4	1
BOT20	18	2			15	4	1
HUMANA01	7				7		
HUMANA02	9				9		
HUMANA03	49	9			53	5	
HUMANA04	3	6			5	4	
HUMANA05	8				8		
HUMANA06	36	28		13	41	10	
HUMANA07	24				24		
HUMANA08	7	5			7	5	
HUMANA09	4	10			9	5	
HUMANA10	11				10		1
HUMANA11	16				16		
HUMANA12	23	5	16		18	20	6
HUMANA13	8				6	2	
HUMANA14		8				8	
HUMANA15	47				47		
HUMANA16	49				49		
HUMANA17	10				7	3	
HUMANA18	8	10			7	11	
HUMANA19	31	3			25	9	
HUMANA20	5	21	1		27		
HUMANA21	12				12		
Quantidade posts	3213	248	32	63	3235	182	13
Porcentagem	92%	7%	1%	2%	93%	5%	0%

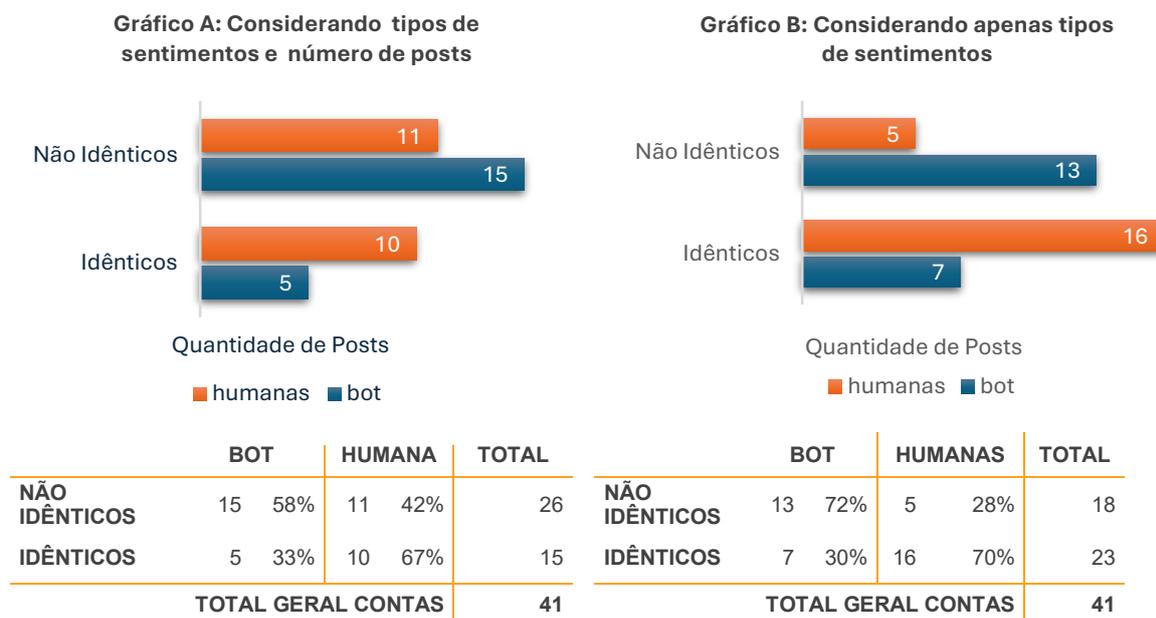
Nota. Fonte: Elaborado pelos Autores (2024). *[Descrição do quadro]* Quadro com 41 linhas representando contas na plataforma X (antigo Twitter) e 7 colunas. As colunas estão organizadas da seguinte forma: “CONTA”, três colunas com classificações de sentimentos atribuídas pelo modelo ChatGPT — “NEGATIVO”, “NEUTRO” e “POSITIVO” — e quatro colunas atribuídas pelo modelo Gemini — “NÃO IDENTIFICADO”, “NEGATIVO”, “NEUTRO” e “POSITIVO”. Os dados indicam a distribuição de sentimentos gerados por cada modelo para as publicações de 20 contas automatizadas (BOT) e 21 humanas (HUMANA). No total, o ChatGPT atribuiu sentimentos negativos a 3.213 postagens (92%), neutros a 248 (7%) e positivos a 52 (1%). O Gemini classificou 3.235 como negativas (93%), 182 como neutras (5%), 13 como positivas (0%) e 63 como “não identificadas” (2%). *[Fim da descrição]*

Na classificação dos sentimentos de postagens das 41 contas do X, apresentadas no Quadro 21, pelo ChatGPT e pelo Gemini, observou-se que 15 contas tiveram classificação de sentimentos idênticas (considerando os sentimentos e o número de posts) quando classificadas pelo GPT e Gemini. Dessas 10 eram contas humanas e 5 contas bots. As 26 contas restantes, tiveram classificações de sentimentos diferentes (ou sentimentos diferentes ou número de postagens diferentes para cada sentimento) quando classificadas pelo GPT e pelo Gemini. Dessas 15 eram bots e 11 humanas) como pode ser analisado no Gráfico A da Figura 3.

A partir dos resultados da Figura 3, diversas análises podem ser elaboradas para compreender o comportamento de contas bots e humanas em relação à classificação de sentimentos.

Figura 3

Comportamento de contas bots e humanas em relação à classificação de sentimentos



Nota. Fonte: Elaborado pelos Autores (2024). *[Descrição da imagem]* A imagem apresenta dois gráficos de barras horizontais (Gráficos A e B) e duas tabelas-resumo, analisando o comportamento de 41 contas — bots e humanas — em relação à classificação de sentimentos gerada por modelos de IA. O Gráfico A compara a quantidade de contas com classificações de sentimentos idênticas ou não idênticas, considerando tanto o tipo de sentimento (positivo, neutro, negativo) quanto o número de postagens. Entre as contas com classificações não idênticas, destacam-se 15 bots (azul) e 11 humanas (laranja); nas classificações idênticas, 5 bots (azul) e 10 humanas (laranja). O Gráfico B considera apenas o tipo de sentimento, desconsiderando o volume de posts. Nele, 13 bots e 5 humanas apresentaram classificações não idênticas, enquanto 7 bots e 16 humanas tiveram classificações idênticas. As tabelas abaixo do gráfico detalham os percentuais. No Gráfico A, 58% dos bots e 42% das humanas tiveram classificações não idênticas. Já no Gráfico B, 72% dos bots e 28% das humanas foram não idênticos. Os dados evidenciam maior consistência na classificação de sentimentos atribuída às contas humanas em comparação às bots. *[Fim da descrição]*

Em relação a consistência entre classificadores, 67% das contas humanas (10 de 15) tiveram classificações de sentimentos idênticas pelos dois classificadores, enquanto apenas 33% das contas *bots* (5 de 15) tiveram classificações idênticas (Gráfico A da Figura 3). Isso sugere que as postagens de contas humanas podem ser mais consistentes em termos de sentimentos, o que facilita a concordância entre os dois modelos de classificação (ChatGPT e Gemini). Já as contas *bots* mostram maior variabilidade ou complexidade, o que pode causar discordância entre classificadores.

Considerando as diferenças nas classificações dos tipos de sentimentos, foram encontradas maiores divergências nas contas *bots*, onde 58% das contas que tiveram classificações diferentes entre GPT e Gemini (15 de 26) conforme apresentado no Gráfico A da Figura 3. Isso sugere que *bots* podem gerar conteúdo que é interpretado de formas diferentes pelos classificadores. Pode ser uma indicação de que as mensagens criadas por *bots* são mais ambíguas ou que os *bots* têm uma programação que gera sentimentos mais variados.

Levando em consideração somente o número de sentimentos apresentados pelas contas (desconsiderando o número de postagens em cada sentimento), foram identificadas 18 contas com número de sentimentos diferentes (13 *bots* e 5 humanas) e 23 contas com número de sentimentos iguais (7 *bots* e 16 humanas) conforme pode ser observado no Gráfico B da Figura 3. Observa-se que a maior parte dessas contas eram *bots* (72%). Isso indica que *bots* podem variar mais em termos de número de sentimentos que expressam, talvez por estratégias de automação que tentam replicar comportamento humano, mas falham ao não manter consistência.

Considerando as contas com número de sentimentos iguais, a maioria são humanas, totalizando 70%. Isso reforça a hipótese de que contas humanas são mais consistentes na expressão de sentimentos, enquanto os *bots* tendem a oscilar mais.

Os resultados sugerem que os *bots* tendem a produzir conteúdo com maior variação tanto no número quanto nos tipos de sentimentos. Isso pode ser uma estratégia dos *bots* para parecer mais humanos, mas também pode revelar limitações no design dos *bots*, que geram inconsistências perceptíveis pelos classificadores.

A diferença nos resultados de classificação entre GPT e Gemini para as contas *bots* pode estar relacionada às dificuldades desses modelos em lidar com linguagem gerada artificialmente. Como os *bots* podem utilizar padrões de linguagem diferentes dos humanos, os modelos podem ter mais dificuldade em categorizar os sentimentos corretamente.

As inconsistências nos sentimentos entre *bots* e humanos podem ser usadas para melhorar algoritmos de detecção de *bots*. Se *bots* mostram padrões de sentimentos mais variados ou inconsistentes, isso pode ser um sinal distintivo para classificadores futuros. Por outro lado, a maior concordância nos sentimentos de contas humanas pode ser um indicativo de autenticidade.

A análise sugere que contas humanas tendem a ser mais consistentes na expressão de sentimentos, tanto em termos de conteúdo quanto de número de sentimentos. *Bots*, por outro lado, apresentam maior variabilidade, o que pode gerar divergências entre classificadores de sentimentos como o GPT e o Gemini. Essa variabilidade pode ser uma chave para identificar e diferenciar *bots* de contas humanas, contribuindo para estudos de comportamento online e detecção de automação.

Esse tipo de análise pode ser expandido com mais dados ou modelos, mas já sugere uma tendência de como *bots* e humanos interagem com o conteúdo digital e como suas postagens podem ser analisadas para entender seu impacto no ecossistema de informação científica online.

5 Conclusões

A análise realizada evidencia a viabilidade do uso das ferramentas ChatGPT e Gemini na classificação de sentimentos de postagens em redes sociais, particularmente de um artigo científico que trata da Covid 19. Ambas as ferramentas demonstraram uma capacidade consistente de interpretar o conteúdo dos resumos dos artigos e utilizá-los como base para identificar e categorizar os sentimentos expressos nas postagens. Essa abordagem automatizada se mostra eficaz para o processamento de grandes volumes de dados, oferecendo uma solução escalável para a análise de interações e opiniões dos usuários, contribuindo para uma compreensão mais profunda das dinâmicas sociais e informacionais no ambiente digital.

Q1: Qual a diferença de resultados entre as ferramentas *Large Language Models* (LLMs) ChatGPT 3.5 e Gemini 1.5 Flash ao classificar os sentimentos de postagens realizadas no X sobre artigos científicos?

A conclusão que se pode traçar a partir dessa análise é que, embora o desempenho do ChatGPT 3.5 e do Gemini 1.5 Flash seja semelhante na classificação de posts negativos, o primeiro demonstrou uma maior eficácia ao lidar com postagens que continham apenas links de artigos científicos. O Gemini mostrou uma postura mais cautelosa, não classificando 63 posts devido à falta de informações textuais associadas aos links, enquanto o ChatGPT atribuiu uma classificação em todos esses casos.

Q2: Qual a eficácia das ferramentas LLMs ChatGPT 3.5 e Gemini 1.5 Flash ao classificar os sentimentos de postagens realizadas no X sobre artigos científicos?

Após uma verificação manual, observou-se que o ChatGPT cometeu poucos erros (9 em 63), demonstrando uma precisão significativa, o que sugere que, em cenários onde há pouca informação textual, o ChatGPT ainda consegue fornecer análises razoavelmente corretas. Isso indica que, no contexto da análise de sentimentos sobre artigos científicos na plataforma X, a eficácia do ChatGPT 3.5, mesmo em situações desafiadoras, pode ser mais útil do que a postura de cautela adotada pelo Gemini, dependendo do objetivo da análise.

A comparação entre a classificação automática e a manual ressalta a eficácia do GPT em identificar sentimentos, mas também evidencia áreas onde melhorias são necessárias, especialmente na identificação de sentimentos positivos e no tratamento de postagens ambíguas ou complexas. O progresso contínuo em modelos de linguagem e aprendizado de máquina é necessário para que essas ferramentas possam oferecer resultados ainda mais confiáveis e úteis para os pesquisadores e analistas.

Q3: Quais são os padrões de sentimentos que mais diferenciam as contas *bots* das contas humanas na rede social X no contexto da divulgação científica?

Tanto ChatGPT quanto Gemini apontam que os *bots* tendem a veicular conteúdo mais polarizado, com uma forte predominância de negatividade. Isso pode estar relacionado ao uso desses *bots* para influenciar a percepção pública de forma mais agressiva ou com a intenção de gerar controvérsia. As contas humanas, por outro lado, apresentam uma distribuição mais variada de sentimentos, o que sugere uma abordagem mais equilibrada e diversa na interação com o conteúdo científico. Humanos demonstram um maior espectro de reações emocionais, com uma boa parcela de postagens neutras e até algumas positivas.

Os *bots* parecem predominantemente produzir postagens negativas. Isso pode indicar uma tendência dos *bots* em se envolver em discussões mais críticas ou polarizadas. Embora uma parte das postagens seja neutra, o número de postagens positivas é bastante baixo, sugerindo que os *bots* estão mais focados em compartilhar críticas ou informações que geram menos engajamento positivo o que pode estar relacionada a estratégias de manipulação da opinião pública, frequentemente utilizadas para gerar polarização e desconfiança. O tom negativo pode aumentar o engajamento e a viralização das postagens, uma vez que conteúdos críticos ou polêmicos tendem a provocar mais interações nas redes sociais. Essa abordagem também pode refletir interesses específicos de grupos políticos ou ideológicos que utilizaram os *bots* para enfraquecer narrativas institucionais e promover discursos alternativos.

As postagens feitas por humanos também tendem a ser majoritariamente negativas, mas há uma presença maior de sentimentos neutros em comparação aos *bots*. Isso pode indicar que, embora os humanos também compartilhem críticas, eles têm uma abordagem mais equilibrada, sem o mesmo foco crítico extremo observado nos *bots*. No entanto, assim como no caso dos *bots*, há uma quantidade muito baixa de postagens positivas, sugerindo que o engajamento positivo com artigos científicos pode ser limitado nas discussões online.

A análise dos sentimentos de *bots* e humanos em postagens sobre artigos científicos revela um cenário predominantemente negativo, com *bots* sendo mais críticos e humanos apresentando uma distribuição um pouco mais equilibrada entre sentimentos negativos e neutros. A baixa presença de postagens positivas, tanto de *bots* quanto de humanos, sugere que a divulgação de artigos científicos está envolvida em debates focados em críticas ou questões controversas, em vez de um engajamento positivo ou de elogios.

Em conclusão, a análise de sentimentos realizada por ChatGPT e Gemini revela padrões distintos entre *bots* e contas humanas na divulgação científica. Os *bots* tendem a veicular conteúdos mais polarizados e predominantemente negativos, o que sugere um uso intencional dessas contas para intensificar debates e influenciar a percepção pública de maneira mais agressiva. Em contraste, as contas humanas exibem uma maior diversidade de sentimentos, com interações mais equilibradas, incluindo uma significativa quantidade de

postagens neutras e algumas positivas. Esses resultados indicam que *bots* podem estar sendo usados para polarizar o debate em torno de questões científicas, enquanto os humanos contribuem com uma gama mais ampla de emoções e perspectivas, favorecendo uma interação mais rica e diversificada no ambiente digital.

Esses achados fornecem indícios importantes para pesquisadores e profissionais interessados na análise de sentimentos em mídias sociais, destacando a importância de considerar a autenticidade das interações e a eficácia de ferramentas de análise automatizada. Além disso, esses resultados podem ser aplicados para aprimorar a detecção de *bots*, combinando a análise de sentimentos com outras métricas, como padrões de engajamento e redes de conexão. A tendência dos *bots* a expressarem sentimentos mais polarizados e negativos pode ser usada para modelagem preditiva, contribuindo para classificadores híbridos que integrem aprendizado de máquina e LLMs. Considerando a necessidade de desenvolver métodos mais eficazes para detecção de *bots*, esses achados podem contribuir para o desenvolvimento de estratégias como:

- a) Modelagem de padrões emocionais: Incorporar a análise de sentimentos como uma métrica complementar na identificação de bots, associando padrões de polarização emocional à atividade automatizada.
- b) Integração com outras métricas: Combinar a análise de sentimentos com métricas de engajamento, frequência de postagens e rede de conexões para criar um modelo mais refinado de detecção.
- c) Desenvolvimento de classificadores híbridos: Criar modelos que integrem LLMs com técnicas supervisionadas de aprendizado de máquina, treinadas especificamente para diferenciar sentimentos expressos por bots e humanos.
- d) Automação da reclassificação manual: Incorporar heurísticas baseadas nos achados do estudo para reduzir a necessidade de revisão manual, otimizando o processo de diferenciação entre contas automatizadas e humanas.

Embora o estudo tenha apresentado uma análise consistente e original sobre a identificação de *bots* por meio da análise de sentimentos em dados altmétricos, algumas limitações podem ser consideradas para aprimorar futuras pesquisas. A amostra, centrada em um único artigo científico, oferece uma visão válida, mas a inclusão de uma variedade maior de publicações científicas poderia expandir a aplicabilidade dos resultados. A utilização das ferramentas ChatGPT 3.5 e Gemini 1.5 Flash trouxe contribuições valiosas, entretanto, essas ferramentas, embora avançadas, são treinadas com grandes volumes de dados que podem incluir informações enviesadas ou desatualizadas. Além disso, sua incapacidade de compreender nuances contextuais complexas ou a limitação na identificação de sentimentos sutis pode afetar a classificação, especialmente em postagens mais curtas ou em contextos específicos como a divulgação científica. Uma análise mais aprofundada desses vieses poderia melhorar a interpretação dos resultados e ajustar as ferramentas para um desempenho mais preciso.

Uma sugestão para trabalhos futuros visando melhorar a eficácia do ChatGPT e do Gemini na análise de sentimentos em contextos ambíguos ou com poucos dados é aprimorar a engenharia de prompt e incorporar metadados contextuais aos modelos. Além disso, estratégias como ajustar os modelos para a análise científica, utilizar múltiplos modelos em comparação e aplicar feedback iterativo podem aumentar a precisão na classificação. Essas melhorias aperfeiçoam a capacidade das ferramentas de identificar diferentes sentimentos e fortaleceriam sua aplicação na detecção automatizada de *bots*, além de contribuir para o entendimento da disseminação de publicações científicas em redes sociais.

Outro ponto relevante é o uso de 41 contas manualmente classificadas, que, embora representativas, não fornecem um panorama completo de comportamentos e postagens. A ampliação dessa amostra, assim como a exploração de outras ferramentas de LLMs como Deep Seek e Llama, poderia contribuir para validar e refinar a metodologia proposta. Além disso, é recomendável verificar se o ChatGPT e o Gemini terão a mesma resposta se aplicado em momentos diferentes para análise de conteúdo, além de avaliar minuciosamente o impacto das interações entre contas humanas e automatizadas nas mídias sociais, especialmente no contexto acadêmico. Isso permitirá uma melhor compreensão das disparidades comportamentais e seu consequente efeito nas interações sociais online, ao mesmo tempo em que fortalece as estratégias de detecção de *bots*.

Referências

- Anwar, A., & Yaqub, U. (2020). Bot detection in Twitter landscape using unsupervised learning. In *The 21st Annual International Conference on Digital Government Research 2020* (pp. 329–330). ACM. <https://dl.acm.org/doi/10.1145/3396956.3401801>
- Arredondo, L. (2018). A study of altmetrics using sentiment analysis. *Honors Capstones*, 70, 1–6. <https://huskiecommons.lib.niu.edu/studentengagement-honorscapstones/70/>
- Babur, Z., Bekdemir, U., Sen, A., Carkit, S. O., Genc, O., Gulcu, A., Gumustas, C., & Soyak, E. G. (2023). Unmasking Twitter bots: Feature engineering and machine learning for bot account identification. In *Innovations in Intelligent Systems and Applications Conference, ASYU 2023* (pp. 16–21). IEEE. <https://doi.org/10.1109/ASYU58738.2023.10296706>
- Bär, D., Calderon, F., Lawlor, M., Lickleder, S., Totzauer, M., & Feuerriegel, S. (2023). Analyzing social media activities at Bellingcat. In *Proceedings of the 15th ACM Web Science Conference 2023* (pp. 163–173). ACM. <https://dl.acm.org/doi/10.1145/3578503.358360>
- Barreto, S., Moura, R., Carvalho, J., Paes, A., & Plastino, A. (2023). Sentiment analysis in tweets: An assessment study from classical to modern word representation models. *Data Mining and Knowledge Discovery*, 37(1), 318–380. <https://doi.org/10.1007/s10618-022-00853-0>
- Bello, B. S., & Heckel, R. (2019). Analyzing the behaviour of Twitter bots in post-Brexit politics. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 61–66). IEEE. <https://doi.org/10.1109/SNAMS.2019.8931874>
- Bello, B. S., Heckel, R., & Minku, L. (2018). Reverse engineering the behaviour of Twitter bots. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 27–34). IEEE. <https://doi.org/10.1109/SNAMS.2018.8554675>
- Breve, B., Caruccio, L., Cirillo, S., Deufemia, V., & Polese, G. (2024). Analyzing the worldwide perception of the Russia-Ukraine conflict through Twitter. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00921-w>
- Buscemi, A., & Proverbio, D. (2024). ChatGPT vs Gemini vs LLaMA on multilingual sentiment analysis. *arXiv*, 1–11. <https://doi.org/10.48550/arXiv.2402.01715>
- Cai, M., Luo, H., Meng, X., Cui, Y., & Wang, W. (2023). Network distribution and sentiment interaction: Information diffusion mechanisms between social bots and human users on social media. *Information Processing & Management*, 60(2), 103197. <https://doi.org/10.1016/j.ipm.2022.103197>
- Çetinkaya, Y. M., Toroslu, İ. H., & Davulcu, H. (2020). Developing a Twitter bot that can join a discussion using state-of-the-art architectures. *Social Network Analysis and Mining*, 10(1), 1–21. <https://doi.org/10.1007/s13278-020-00665-4>
- Chawla, V., & Kapoor, Y. (2023). A hybrid framework for bot detection on Twitter: Fusing digital DNA with BERT. *Multimedia Tools and Applications*, 82(20), 30831–30854. <https://doi.org/10.1007/s11042-023-14730-5>
- Chen, H., Zhu, Z., Qi, F., Ye, Y., Liu, Z., Sun, M., & Jin, J. (2021). Country image in COVID-19 pandemic: A case study of China. *IEEE Transactions on Big Data*, 7(1), 81–92. <https://doi.org/10.1109/TBDATA.2020.3023459>
- Darling, E., Shiffman, D., Côté, I., & Drew, J. (2013). The role of Twitter in the life cycle of a scientific publication. *Ideas in Ecology and Evolution*, 6, 32–43. <https://doi.org/10.4033/iee.2013.6.6.f>
- Galgoczy, M. C., Phatak, A., Vinson, D., Mago, V. K., & Giabbanelli, P. J. (2022). (Re)shaping online narratives: When bots promote the message of President Trump during his first impeachment. *PeerJ Computer Science*, 8, 1–26. <https://doi.org/10.7717/peerj-cs.947>
- Gatkal, S., Panjwani, D., Barhate, S., Mangla, R., & Kazi, F. (2021). Community detection and impact of bots on sentiment polarity of Twitter networks. *2021 Asian Conference on Innovation in Technology (ASIANCON 2021)*, 1–6. <https://doi.org/10.1109/ASIANCON51346.2021.9544691>

- Graham, T., Bruns, A., Angus, D., Hurcombe, E., & Hames, S. (2021). #IStandWithDan versus #DictatorDan: The polarised dynamics of Twitter discussions about Victoria's COVID-19 restrictions. *Media International Australia*, 179(1), 127–148. <https://doi.org/10.1177/1329878X20981780>
- Halevi, G., & Schimming, L. (2018). An initiative to track sentiments in altmetrics. *Journal of Altmetrics*, 1(1), 1–7. <https://doi.org/10.29024/joa.1>
- Hassan, S. U., Aljohani, N. R., Tarar, U. I., Safder, I., Sarwar, R., Alelyani, S., & Nawaz, R. (2023). Exploiting tweet sentiments in altmetrics large-scale data. *Journal of Information Science*, 49(5), 1229–1245. <https://doi.org/10.1177/016555152111043713>
- Hassan, S. U., Saleem, A., Soroya, S. H., Safder, I., Iqbal, S., Jamil, S., Bukhari, F., Aljohani, N. R., & Nawaz, R. (2021). Sentiment analysis of tweets through Altmetrics: A machine learning approach. *Journal of Information Science*, 47(6), 712–726. <https://doi.org/10.1177/0165551520930917>
- Ibrahim, M., Abdillah, O., Wicaksono, A. F., & Adriani, M. (2015). Buzzer detection and sentiment analysis for predicting presidential election results in a Twitter nation. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)* (pp. 1348–1353). IEEE. <https://doi.org/10.1109/ICDMW.2015.113>
- Inuwa-Dutse, I., Bello, B. S., & Korkontzelos, I. (2018). Lexical analysis of automated accounts on Twitter. In *Proceedings of the International Conferences on WWW/Internet 2018 and Applied Computing 2018* (pp. 75–82).
- Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, 65, 101188. <https://doi.org/10.1016/j.intfin.2020.101188>
- Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. *Information Sciences*, 467, 312–322. <https://doi.org/10.1016/j.ins.2018.08.019>
- Liu, H., Chatterjee, I., Zhou, M., Lu, X. S., & Abusorrah, A. (2020). Aspect-based sentiment analysis: A survey of deep learning methods. *IEEE Transactions on Computational Social Systems*, 7(6), 1358–1375. <https://doi.org/10.1109/TCSS.2020.3033302>
- Liu, X. (2019). A big data approach to examining social bots on Twitter. *Journal of Services Marketing*, 33(4), 369–379. <https://doi.org/10.1108/JSM-02-2018-0049>
- Mouronte-López, M. L., Gómez Sánchez-Seco, J., & Benito, R. M. (2024). Patterns of human and bots behaviour on Twitter conversations about sustainability. *Scientific Reports*, 14(1), 1–14. <https://doi.org/10.1038/s41598-024-52471-z>
- Mustafaraj, E., & Metaxas, P. T. (2011). What edited retweets reveal about online political discourse. In *Proceedings of the 5th AAAI Conference on Analyzing Microtext 2011* (pp. 38–43). AAAI Press. <https://dl.acm.org/doi/10.5555/2908630.2908637>
- Pastor-Galindo, J., Zago, M., Nespola, P., Bernal, S. L., Celdran, A. H., Perez, M. G., Ruiperez-Valiente, J. A., Perez, G. M., & Marmol, F. G. (2020). Spotting political social bots in Twitter: A use case of the 2019 Spanish general election. *IEEE Transactions on Network and Service Management*, 17(4), 2156–2170. <https://doi.org/10.1109/TNSM.2020.3031573>
- Pontes, D., & Maricato, J. (2023b). Classificação de contas do Twitter através do Botometer. In *Encontro Nacional de Pesquisa em Informação 2023*, Aracaju, SE. *Anais* [...]. Aracaju, SE. <https://enancib.ancib.org/index.php/enancib/xxxii/enancib/paper/view/1957>
- Pontes, D., & Maricato, J. (2023a). *Dataset with classification of accounts that disclose scientific articles on Twitter between bot, human and cyborg through Botometer* [Dataset]. Zenodo. <https://zenodo.org/records/8137161>
- Rane, N., Choudhary, S., & Rane, J. (2024). Gemini versus ChatGPT: Applications, performance, architecture, capabilities, and implementation. *Journal of Applied Artificial Intelligence*, 5(1), 69–93. <https://doi.org/10.48185/jaai.v5i1.1052>
- Silva, M. E. M., & Serrano, P. H. S. M. (2023). Análise de sentimentos em textos de redes sociais: Uma comparação entre o ChatGPT e métodos tradicionais. <https://doi.org/10.5902/2316882X84828>
- Suarez-Lledo, V., & Alvarez-Galvez, J. (2022). Assessing the role of social bots during the COVID-19 pandemic: Infodemic, disagreement, and criticism. *Journal of Medical Internet Research*, 24(8), 1–12. <https://doi.org/10.2196/36085>

- Tavazoe, F., Buscaldi, D., Mola, F., & Conversano, C. (2020). Empowering detection of malicious social bots and content spammers on Twitter by sentiment analysis. *Electronic Journal of Applied Statistical Analysis*, 13(2), 375–389. <https://doi.org/10.1285/i20705948v13n2p375>
- Uyheng, J., & Carley, K. M. (2020). Bot impacts on public sentiment and community structures: Comparative analysis of three elections in the Asia-Pacific. In *Lecture Notes in Computer Science (Vol. 12268, pp. 12–22)*. Springer. https://doi.org/10.1007/978-3-030-61255-9_2
- Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. *Proceedings of the 11th International Conference on Web and Social Media (ICWSM 2017)*, 280–289. <https://doi.org/10.1609/icwsm.v11i1.14871>
- Velayutham, T., & Tiwari, P. K. (2017). Bot identification: Helping analysts for right data in Twitter. In *2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall)* (pp. 1–5). IEEE. <https://doi.org/10.1109/ICACCAF.2017.8344722>
- Wu, J., Teng, E., & Cao, Z. (2022). Twitter bot detection through unsupervised machine learning. In *Proceedings - 2022 IEEE International Conference on Big Data (Big Data 2022)* (pp. 5833–5839). IEEE. <https://doi.org/10.1109/BigData55660.2022.10020983>
- Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., & Sharif, S. (2021). An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, 108, 256–262. <https://pubmed.ncbi.nlm.nih.gov/34052407/>
- Zhang, L., Gou, Z., Fang, Z., Sivertsen, G., & Huang, Y. (2023). Who tweets scientific publications? A large-scale study of tweeting audiences in all areas of research. *Journal of the Association for Information Science and Technology*, 74(13), 1485–1497. <https://doi.org/10.1002/asi.24830>

Dados de publicação

Danielle Pompeu Noronha Pontes

Mestre

Universidade do Estado do Amazonas, Escola Superior de Tecnologia, Manaus, AM, Brasil

Universidade de Brasília, Faculdade de Ciência da Informação, Brasília, DF, Brasil

dnoronha@uea.edu.br

<https://orcid.org/0000-0003-0036-6638>

Graduada em Processamento de Dados pela Universidade Federal do Amazonas (1996) e mestrado em Sistemas Digitais pela Universidade de São Paulo (2012). Concluiu pós-graduação lato sensu em Informática na Universidade Federal do Ceará (2002). Atualmente é professora da Universidade do Estado do Amazonas. Aluna do programa de Doutorado em Ciência da Informação da Universidade de Brasília nas áreas de Almetria e Inteligência Artificial.

João de Melo Maricato

Doutor

Universidade de Brasília, Faculdade de Ciência da Informação, Brasília, DF, Brasil

jmmaricato@gmail.com

<https://orcid.org/0000-0001-9162-6866>

Ph.D. Doutor em Ciência da Informação pela Escola de Comunicações e Artes da Universidade de São Paulo – ECA/USP (2010). Bacharelado em Biblioteconomia e Ciência da Informação pela Universidade Federal de São Carlos – UFSCar (2002). Professor do curso de graduação em Biblioteconomia da Faculdade de Ciência da

Informação (FCI) da Universidade de Brasília (UnB). Professor do Programa de Pós-Graduação em Ciência da Informação da UnB (PPGCinf).

Originalidade

Declaro que o texto é original e não está sendo revisado por nenhuma outra publicação. Caso eu decida cancelar o processo de publicação, concordo em informar imediatamente a equipe editorial da Revista Biblios para que o envio possa ser arquivado.

Preprint

O manuscrito não foi submetido a nenhuma plataforma de Preprints.

Informações sobre o trabalho

Este artigo foi apresentado no 9º Encontro Brasileiro de Bibliometria e Cientometria (EBBC 2024). Na revista Biblios, passou por novas rodadas de avaliação por pares e foi aprimorado para fins de publicação.

Agradecimentos

Não se aplica.

Contribuição dos autores

Concepção e elaboração do manuscrito: D.P.N. Pontes, J.M. Maricato

Coleta de dados: D.P.N. Pontes

Análise de dados: D.P.N. Pontes

Discussão dos resultados: D.P.N. Pontes, J.M. Maricato

Revisão e aprovação: D.P.N. Pontes, J.M. Maricato

Uso de inteligência artificial

Foram utilizadas as ferramentas de IA ChatGPT e Gemini para auxiliar na redação, revisão e tradução do trabalho.

Financiamento

Não se aplica.

Permissão para usar imagens

Não se aplica.

Aprovação do Comitê de Ética em Pesquisa

Não se aplica.

Conflito de interesses

Não se aplica.

Declaração de disponibilidade de dados

Os dados estão disponibilizados em arquivo Excel em:

Pontes, D., & Maricato, J. (2025). *Sentiment Analysis of COVID-19 Scientific Publication Dissemination on Social Media X: A Dataset Analyzed with ChatGPT 3.5 and Gemini 1.5 Flash* [Dataset]. Zenodo.

<https://zenodo.org/records/14919674>.

O conjunto de dados disponibilizado inclui uma amostra de postagens no X que fizeram menção ao editorial, publicado no periódico “*Dying in a Leadership Vacuum*” em 7 de outubro de 2020, com o título “*Dying in a Leadership Vacuum*” (DOI: <https://10.1056/NEJMe2029812>).

Foi realizada a coleta de uma amostra de postagens do X que fizeram menção à publicação. Os posts foram extraídos da plataforma Altmeter através de um algoritmo em Python 3.12 usando a biblioteca BeautifulSoup 4.12 e o ambiente de desenvolvimento Google Colab. Com isso, foi gerado um conjunto de dados composto por 9.792 postagens do X que comentaram especificamente o referido editorial. Dessas postagens, foram identificados 5.601 perfis únicos, que foram confrontados com os perfis classificados e disponibilizadas no *dataset* gerado por Pontes

e Maricato (2023b). Das contas que continham a classificação (*bot* ou humano) foram selecionadas 41 contas que fizeram mais de 4 postagens.

Segundo o *dataset* disponibilizado por Pontes e Maricato (2023b), 10 contas foram classificadas pelo Botometer como *bot* e 31 como humanas. Considerando que Pontes e Maricato (2023a) citam a limitação na classificação das contas da rede de atenção atômica pelo Botometer foi realizada uma classificação manual das 41 contas selecionadas. A classificação manual das contas selecionadas tomou como critério a análise da quantidade de postagens, o horário da postagem, o intervalo de tempo entre as postagens, a data de criação das contas, a imagem da foto. Nesta classificação manual, foram identificados que 20 contas são *bots* e 21 contas são humanas.

As contas classificadas realizaram um total de 3.493 postagens que estão disponibilizados neste *dataset* e utilizados nas análises apresentadas no artigo.

A estrutura dos metadados que compõem o *dataset* está apresentado no Quadro 1 a seguir.

Quadro 1

Esquema de metadados do conjunto de dados

Nome da variável	Tipo de dado	Descrição	Valores possíveis
CONTA	String (Texto)	Código anonimização da conta para preservar a identidade do usuário.	ACCOUNT + número sequencial
CLASS CONTA (BTM)	Categórico	Classificação automática da conta usando uma ferramenta como Botometer.	humano, bot
CLASS CONTA (MANUAL)	Categórico	Classificação manual da conta baseada na análise do pesquisador.	humano, bot
CONTEUDO POST	Texto (String)	Texto completo do post coletado.	
CLASS SENT (GPT)	Categórico	Classificação de sentimento atribuída pelo ChatGPT.	positivo, neutro, negativo
CLASS SENT (GEMINI)	Categórico	Classificação de sentimento atribuída pelo Gemini.	positivo, neutro, negativo
GPT X GEM (IGUAL/DIF)	Binário	Indica se a classificação dos sentimentos foi igual ou diferente entre ChatGPT e Gemini.	igual, diferente
CLASS POST (MANUAL)	Categórico	Classificação manual do sentimento do post.	positivo, neutro, negativo
RESULT GPT	Categórico	Avaliação da classificação do ChatGPT em relação ao padrão manual.	correto, incorreto
RESULT GEMINI	Categórico	Avaliação da classificação do Gemini em relação ao padrão manual.	correto, incorreto

Nota. Fonte: Elaborado pelos Autores (2024).

Pontes, D., & Maricato, J. (2023a). *Classificação de contas do Twitter através do botometer*. [Artigo publicado]. XXXIII Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação, Aracaju, SE, Brasil.

<https://enancib.ancib.org/index.php/enancib/xxxii/enancib/paper/view/1957>

Pontes, D., & Maricato, J. (2023b). *Dataset with classification of accounts that disclose scientific articles on Twitter between bot, human and cyborg through botometer* [Dataset]. Zenodo <https://zenodo.org/records/8137161>.

Licença de uso - uso exclusivo do periódico

Os autores concedem à Biblios direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob uma Licença *Creative Commons Atribuição (CC BY) 4.0 Internacional*. Esta licença permite que terceiros remixem, adaptem e desenvolvam o trabalho publicado, dando os devidos créditos pela autoria e publicação inicial neste periódico. Os autores estão autorizados a firmar acordos adicionais separados para distribuição não exclusiva da versão publicada do trabalho no periódico (por exemplo, publicação em um repositório institucional, em um site pessoal, publicação de uma tradução ou como um capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

Editor

Publicado pelo Sistema de Bibliotecas Universitárias da Universidade de Pittsburgh. Responsabilidade compartilhada com universidades parceiras. As ideias expressas neste artigo são dos autores e não representam necessariamente as opiniões dos editores ou da universidade.

Editora

Lucia da Silveira.

Histórico

Recebido: 15-09-2024 – Aprovado: 01-04-2025 – Publicado: 03-07-2025



The articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by [Pitt Open Library Publishing](http://pittopenlibrarypublishing.com).