

IA GENERATIVA: POTENCIAL E LIMITAÇÕES DOS LARGE LANGUAGE MODELS E PROMPTS NA PRODUÇÃO DE NOTÍCIAS

GENERATIVE AI: POTENTIAL AND LIMITATIONS OF LARGE LANGUAGE MODELS AND PROMPTS IN NEWS PRODUCTION

IA GENERATIVA: POTENCIAL Y LIMITACIONES DE LOS LARGE LANGUAGE MODELS Y PROMPTS EN LA PRODUCCIÓN DE NOTICIAS

Fabia Ioscote¹

Resumo

A expansão da Inteligência Artificial Generativa (GenAI), exemplificada pelo ChatGPT e MidJourney, destaca a crescente tendência de criação de conteúdos textuais e visuais por meio de *software*. Esse estudo examina como os *softwares* de GenAI podem auxiliar na produção de notícias, discutindo seus potenciais e limitações. Os objetos empíricos são ChatGPT (OpenAI), Gemini AI (Google) e MariTalk (Maritaca AI). A metodologia qualitativa e exploratória inclui uma revisão de literatura e o desenvolvimento de um protocolo de testes e critérios de avaliação, abordando aspectos computacionais e jornalísticos, como títulos, *leads*, uso de fontes e precisão. Os resultados indicam que, embora não gerem diretamente conteúdos noticiosos, os *softwares* de GenAI podem ser assistentes úteis nas redações, melhorando textos e sugerindo ideias, em um equilíbrio entre a automação e o julgamento editorial humano.

Palavras-chave: jornalismo; automação; geração de linguagem natural; *software*; comunicação digital.

Abstract

The expansion of GenAI, exemplified by ChatGPT, and MidJourney, highlights the growing trend of creating textual and visual content through software. This study examines how GenAI software can assist in news production, discussing their potential and limitations. The empirical objects are ChatGPT (OpenAI), Gemini AI (Google), and MariTalk (Maritaca AI). The qualitative and exploratory methodology includes a literature review and the development of a testing protocol and evaluation criteria, addressing computational and journalistic aspects such as titles, leads, font usage, and accuracy. The results indicate that, although they do not directly generate news content, GenAI software can serve as useful assistants in newsrooms, enhancing texts and suggesting ideas, striking a balance between automation and human editorial judgment.

Keywords: journalism; automation; natural language generation; software; digital communication.

Resumen

La expansión de la inteligencia artificial generativa (GenAI), ejemplificada por ChatGPT y MidJourney, destaca la creciente tendencia a crear contenidos textuales y visuales mediante *software*. Ese estudio examina cómo los programas de GenAI pueden ayudar en la producción de noticias, discutiendo sus potencias y limitaciones. Los objetos empíricos son ChatGPT (OpenAI), Gemini AI (Google) y MariTalk (Maritaca AI). La metodología cualitativa y exploratoria incluye una revisión de literatura y el desarrollo de un protocolo de pruebas y criterios de evaluación, desarrollando aspectos computacionales y periodísticos, como títulos, *leads*, uso de fuentes y precisión. Los resultados apuntan que, aunque no generan directamente contenidos de noticias, los *softwares* de GenAI pueden ser asistentes útiles en las redacciones, mejorando textos y sugiriendo ideas, en un equilibrio entre la automatización y el juicio editorial humano.

Palabras clave: periodismo; automatización; generación de lenguaje natural; *software*; comunicación digital.

¹ Doutoranda em Comunicação (PPGCOM-UFPR); mestre em comunicação, especialista em marketing pela Pontifícia Universidade Católica do Paraná (PUC-PR) em 2008. Licenciada em Docência para a Educação Básica - Letras (FAEL) em 2021. Bacharel em Comunicação Social - Jornalismo pela Universidade Tuiuti do Paraná (UTP) em 2007. E-mail: iosote.fabia@gmail.com.

1 Introdução

“Foi uma máquina que escreveu isso?” Esta pergunta ilustra a crescente capacidade da Inteligência Artificial (IA) em produzir textos cada vez mais persuasivos e humanizados. Nos últimos anos, houve avanços significativos no campo da IA e do Processamento de Linguagem Natural (PLN), impulsionados pelo desenvolvimento de modelos de linguagem como o GPT (*Generative Pre-trained Transformers*, em tradução livre para o português - Transformadores Generativos Pré-treinados), que conseguem assimilar e reproduzir a linguagem de uma maneira, muitas vezes, praticamente indistinguível da produção humana.

A criação de *software* e aplicativos baseados no modelo de linguagem GPT, como o ChatGPT da OpenAI (OpenAI, 2023a), é um marco na trajetória da IA nos últimos anos, a qual reverberou profundamente nos diálogos acadêmicos e sociais. A perspectiva de automatizar tarefas cotidianas e criar conteúdo textual utilizando *software* reconfigurou setores como *marketing*, agências de notícias e redação de artigos, promovendo a otimização de processos e o aumento da eficiência. No entanto, a possibilidade de *software* de IA desempenhar algumas tarefas ou funções, especialmente aquelas tradicionalmente humanas, também é motivo de preocupação, tanto em termos éticos quanto pela substituição de empregos. Cabe destacar que alguns estudos, como os de Lima Junior e de Paula (2021), ou Canavilhas (2024), discutem como os jornalistas podem se manter relevantes diante de um processo acelerado de revolução tecnológica.

A viabilidade da automação de notícias, ou de certas etapas do processo de produção, tem sido analisada por diversos autores, como Arce (2009), Dörr (2015), Santos (2016) e Carreira (2017). Esses pesquisadores delineiam a automação de *leads* e títulos jornalísticos, o uso de *Natural Language Generation* (NLG) e linguagens de programação como Python, além das possibilidades inerentes ao cenário de Big Data.

Entre os casos mais emblemáticos de *software* de redação de notícias, estão o Quakebot e The Homicide Report, ambos do Los Angeles Times, e o Heliograf, do The Washington Post. Esses são exemplos de programas desenvolvidos, internamente, dentro das redações (Carreira, 2017). Segundo os estudos de Canavilhas (2023), além dos mencionados, programas, como o WordSmith, desenvolvido pela empresa norte-americana Automated Insights e utilizado pela Associated Press e Yahoo News, estão em uso há mais de uma década, o que demonstra que, de alguma forma, sistemas de redação automatizada já fazem parte das redações há anos. Por outro lado, de acordo com Beckett (2019), o investimento em pesquisas e os requisitos

específicos do jornalismo para a aplicação de IA também contribuem para frear o avanço rápido da tecnologia nessa área.

Desde a segunda metade do ano de 2022, a disseminação da Inteligência Artificial Generativa (GenAI), exemplificada pelo ChatGPT, Dall-E 2, Midjourney e outros, trouxe à tona a crescente tendência de *software* para a produção de textos e imagens. Devido à sua acessibilidade e facilidade de uso, esses *softwares* se consolidaram como verdadeiros aliados do público em geral, simplificando tarefas cotidianas como a redação, tradução e formatação de textos. No cenário jornalístico, esses aplicativos são vistos como recursos que podem contribuir para a sugestão de pautas, redação e edição de artigos, simultaneamente, acelerando essas etapas e reduzindo os custos de produção (Pavlik, 2023).

Com base nesse breve contexto, formula-se a seguinte pergunta: até que ponto, e de que maneira, os *softwares* de IA Generativa podem contribuir para a produção de notícias? Este artigo tem como objetivo discutir o desempenho desses *softwares*, seus potenciais e limitações, ao mesmo tempo em que aborda discussões teóricas e conceituais sobre a produção de notícias. Os objetivos específicos são: a) descrever um protocolo que inclua critérios de avaliação e um roteiro de teste para a avaliação de *software* de GenAI voltados para o contexto jornalístico; e b) aplicar o roteiro de teste para avaliar o desempenho dos *softwares* de GenAI na produção de notícias.

2 Geração automatizada de texto e nlg

A geração automatizada de texto é uma das tarefas específicas dentro do campo do Processamento de Linguagem Natural (PLN), uma área da IA e da Ciência da Computação que visa possibilitar a interação entre humanos e computadores por meio de linguagem natural, como texto e fala. A geração automatizada de texto ganhou notoriedade nos últimos anos devido ao seu potencial transformador em diversos campos, incluindo produção de conteúdo, assistentes virtuais e outras aplicações, como tradução automática, legendagem, sumarização e resumos. Além disso, a geração automatizada de texto pode englobar diversas técnicas, desde regras simples de substituição de palavras até modelos avançados de linguagem, como aqueles baseados em redes neurais, como o GPT-3, “um *software* de computador projetado para gerar sequências de palavras, códigos ou outros dados, a partir de uma fonte de entrada, chamada de *prompt*” (Floridi; Chiriatti, 2020, p. 684)².

² No original: "a computer software designed to generate sequences of words, codes or other data, from an input source, called a prompt".

Os primórdios da geração automatizada de texto remontam os primeiros experimentos na década de 1950, quando cientistas da computação começaram a explorar o processamento de linguagem natural (Dörr, 2015). Nesse sentido, o matemático Claude Shannon se destacou em estudos sobre processamento estatístico de linguagem. No artigo *Prediction and Entropy of Printed English*, em tradução livre, Previsão e entropia do inglês impresso, publicado em 1951, o cientista apresentou um novo método para estimar a entropia e a redundância de uma linguagem, propondo um modelo para prever palavras que provavelmente seguirão outras (Shannon, 1951). Modelos de linguagem baseados na arquitetura *Transformer*, como o GPT, são capazes de fazer tais previsões. A arquitetura *Transformer* é um modelo de aprendizado de máquina, utilizado em tarefas de PLN, que se mostrou revolucionário em termos de desempenho e eficiência. Essa arquitetura foi apresentada no artigo intitulado *Attention Is All You Need*, em tradução livre, Atenção é tudo que você precisa, escrito por Vaswani *et al.* (2017), uma equipe da Google Research.

Desde 2018, com o surgimento de modelos de linguagem, como o GPT, a geração de texto alcançou um novo patamar, permitindo a criação de textos coerentes e contextuais em larga escala. Nesse cenário, por exemplo, empresas de *marketing* têm a possibilidade de utilizar algoritmos de geração de texto para criar conteúdo personalizados e direcionados a públicos específicos, enquanto assistentes virtuais podem responder às perguntas dos usuários de maneira mais natural e abrangente.

A Geração de Linguagem Natural, *Natural Language Generation* (NLG), é um subcampo específico dentro do Processamento de Linguagem Natural (PLN) que se concentra na criação automática de textos em linguagem natural a partir de dados estruturados ou informações não textuais (Reiter; Dale, 1997). A NLG é frequentemente utilizada em aplicações que necessitam converter dados em narrativas legíveis, facilitando a comunicação dessas informações de maneira mais acessível e compreensível para os seres humanos. Assim, pode transformar dados estatísticos em relatórios descritivos ou converter informações de um banco de dados em resumos informativos. Vale ressaltar que o Big Data tem contribuído para os avanços na NLG, incluindo sua aplicação em redações, particularmente em narrativas automatizadas (Santos, 2016). Por outro lado, embora possa ser útil para criar histórias a partir de dados, a NLG, ainda, não é capaz de fornecer explicações que exijam interpretação ou inferência dos dados.

Como observa Carreira (2017, p. 131), nessas circunstâncias, esses *softwares* operam como meros “processadores de texto, utilizando recursos de linguística computacional para criar um texto que se assemelha à linguagem natural, e podem até mesmo enganar aqueles que

interagem com eles”. No entanto, a autora enfatiza que isso não implica que esses *softwares* realmente entendam o conteúdo que estão produzindo.

3 Geração automatizada de texto no contexto jornalístico

As aplicações da geração automatizada de texto são vastas e abrangem diversas áreas. Na produção de *storytelling*, por exemplo, os modelos GPT podem ser usados para criar histórias, roteiros e até diálogos realistas para jogos e filmes. No jornalismo, é possível gerar *leads* para notícias ou até mesmo artigos completos sobre eventos.

Estudos como os de Arce (2009), que teorizaram a possibilidade de automatizar o *lead*, entendendo-o como uma fórmula matemática inspirada nos estudos de Nilson Lage, e as narrativas automatizadas usando NLG vistas em Carreira (2017), lançam luz sobre essa perspectiva da geração automatizada de texto no jornalismo.

Historicamente, Canavilhas (2023) aponta que os primeiros experimentos com produção automatizada de texto ocorreram por volta de 2010, cobrindo notícias econômicas e esportivas. Isso se deve ao fato de que essas áreas possuem uma abundância de dados disponíveis na web, com informações padronizadas, o que favorece narrativas estruturadas.

Lage (1987) destaca que, tradicionalmente, a produção de textos envolve restrições ao código linguístico, o que reduz o número de palavras e regras operacionais utilizadas, facilitando o trabalho e permitindo o controle de qualidade. De acordo com Carreira (2017), considerando as características do jornalismo - que requer o processamento de informações em larga escala e para consumo imediato - as variáveis formais na criação de um texto são ainda mais drasticamente reduzidas. “Essa lógica, além de simplificar o trabalho do jornalista, também facilita o trabalho do algoritmo, pois este opera a partir de um conjunto menor de palavras e regras linguísticas” (Carreira, 2017, p. 129).

Segundo Lage (1987), há três elementos na estrutura do *lead* jornalístico: sujeito, predicado e circunstâncias. O autor também destaca três campos semânticos nas notícias: deslocamentos (“indo”), transformações (“fazendo”) e enunciações (“dizendo”). Carreira (2017) explica que a estrutura lógica do *lead* e a classificação dos verbos possibilitam a criação de “padrões análogos”, representados por equações ou algoritmos, tornando-os computáveis. Isso foi teorizado por Arce (2009) usando a seguinte fórmula: $F(x, y, z, \dots)$ “onde F corresponde ao verbo, e ‘x’, ‘y’, ‘z’..., aos argumentos, ou seja, o sujeito x (argumento externo) e complementos y, z... (argumentos internos) do verbo” (Arce, 2009, p. 9). Com base nesse conceito, o método fundamental de escrita de notícias,

tradicionalmente realizado por jornalistas humanos, foi, de certa forma, transferido para algoritmos que produzem textos automatizados, seguindo um conjunto de regras e modelos pré-definidos (Carreira, 2017).

A estrutura da notícia envolve a reportagem de uma série de fatos, começando pelo mais importante, seguindo o mesmo padrão para cada fato (Lage, 1987). Essa estrutura, semelhante ao passo a passo de um *lead*, possibilita a automação e a programação de algoritmos, uma ideia promovida por Nilson Lage em 1997, no XX Congresso Intercom (Santos, 2016). Essa possibilidade ocorre porque as notícias consistem em fatos ou conjuntos de fatos, tornando viável a perspectiva de automação (Arce, 2009; Carreira, 2017).

Estudos contemporâneos, como os de Santos (2016), demonstraram como os jornalistas podem se beneficiar de linguagens de programação, como Python, para automatizar o *lead*. Em um sentido mais amplo de notícias automatizadas, pesquisas como as de Dörr (2015) relacionam o potencial dos dados estruturados na produção de notícias usando NLG. As notícias esportivas e financeiras são mais fáceis de automatizar porque envolvem números e estatísticas, são processáveis, estruturadas e amplamente disponíveis no ambiente digital (Dörr, 2015; Canavilhas, 2023).

4 LLMs: GPTs e LAMDA

Um Modelo de Linguagem de Grande Escala, também conhecido como LLM (*Large Language Model*), é um tipo de algoritmo de aprendizado profundo (*deep learning*) capaz de entender, resumir, traduzir, prever e gerar texto, e outros tipos de conteúdo, com base no conhecimento adquirido a partir de vastos conjuntos de dados (Lee, 2023). O GPT, desenvolvido pela OpenAI e lançado em 2018, é um exemplo proeminente de LLM. O GPT é um modelo de linguagem baseado na arquitetura *Transformer* e foi treinado com grandes quantidades de dados textuais para aprender padrões e representações linguísticas em diferentes contextos. Seu treinamento prévio permite que o modelo preveja as próximas palavras em uma sequência de texto, tornando-o ideal para a tarefa de geração automatizada de texto. De acordo com Floridi e Chiriatti (2020), um grande volume de dados é necessário para que esses modelos produzam resultados relevantes.

A versão inicial do GPT, lançada em 2018, utilizava 110 milhões de parâmetros de aprendizado, que são os valores otimizados pela rede neural durante o treinamento. Em apenas um ano, o GPT-2 aumentou significativamente esse número para 1,5 bilhões. O GPT-3 utiliza 175 bilhões de parâmetros (Floridi; Chiriatti, 2020). O modelo GPT-4, da OpenAI, lançado em

2023, foi treinado com um número significativo de parâmetros, embora a OpenAI não tenha divulgado publicamente o total exato. Essa versão do ChatGPT, baseada na arquitetura GPT-4, oferece avanços significativos no processamento de linguagem natural. O GPT-4 demonstra uma melhor compreensão e gestão do contexto, mantendo a coerência em conversas mais longas. Com uma base de conhecimento mais ampla e capacidades multimodais, pode processar tanto texto quanto imagens, aumentando sua versatilidade.

Quando usado para a tarefa de NLG, o GPT pode receber uma entrada específica, conhecida como *prompt* (ou contexto inicial), e, com base nesse contexto, o GPT gera continuamente um texto que faz sentido e é coerente com a entrada fornecida. Floridi e Chiriatti (2020) explicam que, por não entender o que está sendo produzido, o GPT segue a sequência de palavras digitadas no *prompt*, independentemente do nível de dificuldade da tarefa.

Do ponto de vista técnico de como os LLMs funcionam, Cortiz (2023) explica que o aprendizado ocorre em um espaço multidimensional, no qual um grande conjunto de textos é processado. Por meio de uma representação matemática, os LLMs distribuem a probabilidade de sequências de palavras. Com a arquitetura *Transformer*, os LLMs são capazes de aprender padrões de relações entre palavras. Assim, como se trata de um cálculo de probabilidade entre palavras, podem ocorrer erros nas informações geradas, conhecidos como alucinações da IA.

Outro ponto a ser destacado é o treinamento dos dados utilizados nesses *softwares*. O ChatGPT 3.5, por exemplo, utilizou dados de treinamento anteriores a setembro de 2021. Somente em setembro de 2023, o *software* passou a ser capaz de acessar a internet para fornecer informações atualizadas. O GPT-4 do ChatGPT é gratuito com limites de uso. Desde maio de 2024, inclui uma funcionalidade para fornecer links da internet, aprimorando as respostas com informações atualizadas (OpenAI, 2024).

Enquanto o GPT é um exemplo de LLM, com foco na geração automatizada de texto, o LaMDA (*Language Model for Dialogue Applications*, em tradução livre, Modelo de Linguagem para Aplicações de Diálogo), desenvolvido pelo Google, é um avanço recente que visa melhorar a capacidade de conduzir diálogos naturais e interativos. Ao contrário dos modelos anteriores, que tendiam a produzir respostas contextuais isoladas, o LaMDA busca uma compreensão mais profunda do contexto da conversa para fornecer respostas mais coerentes e abrangentes. “No seu núcleo, o LaMDA não se baseia em conjuntos de dados altamente selecionados ou diálogos roteirizados” (Lug, 2023). Além disso, o LaMDA é projetado para ser mais flexível em suas respostas, permitindo uma gama mais ampla de

informações e, até mesmo, fazendo perguntas para obter esclarecimentos durante a interação (Pichai, 2023). O LaMDA usa bilhões de palavras e dados em sua base e seu crescimento “é medido pela coleta de respostas de seu modelo pré-treinado, modelo ajustado e avaliadores humanos para o diálogo de dois autores de múltiplos turnos” (Lug, 2023).

Apesar dos avanços significativos proporcionados por modelos como o GPT e o LaMDA, é importante mencionar que eles ainda enfrentam desafios éticos e de controle de qualidade. Questões como viés e confiabilidade nas respostas devem ser cuidadosamente consideradas para garantir que as interações com esses modelos sejam seguras e benéficas para os usuários (Pichai, 2023).

Assim, tanto o ChatGPT (baseado no modelo de linguagem GPT) quanto o Gemini AI (baseado no modelo de linguagem LaMDA) pertencem à categoria dos LLMs e são utilizados para automatizar a geração de texto em diversas aplicações, o que os torna *softwares* importantes de Geração de Linguagem Natural (NLG).

4.1 Uma nova fase com a GenAI

A disponibilidade de *software* de GenAI levanta a possibilidade de facilitar a produção de notícias por meio da inserção do *prompt*, sem exigir conhecimento de linguagens de programação. Com base nessa perspectiva, em maio de 2023, Berti (2023) testou pautas, pesquisa, edição e distribuição de textos jornalísticos com o ChatGPT. Para o autor, o *software* pode revolucionar o jornalismo, mas ele pondera que “é perigoso quando o utilizamos para fazer o básico por nós ou o usamos aleatoriamente como uma ferramenta jornalística” (Berti, 2023, p. 143). Outro estudo recente utilizando o ChatGPT foi conduzido por Santos (2023). O pesquisador usou o método de entrevista (EIAF) com o *software*. Santos (2023) concluiu que o controle humano é essencial em todos os textos gerados por GenAI, o que é uma das suas principais limitações. Pavlik (2023) testou o ChatGPT e considera que esse *software* pode ser relevante para melhorar a qualidade e a eficiência do trabalho jornalístico, especialmente no que diz respeito à otimização do tempo.

É importante acrescentar que outros detalhes que fazem parte da estrutura da notícia, como o uso de fontes, a precisão da informação, a gramática e a ortografia, também devem ser observados ao usar o software de GenAI (Martins Filho, 1997). É necessário ter atenção aos problemas conhecidos como alucinações da IA, que se caracterizam como uma espécie de invenção dos fatos.

4.2 Engenharia de *prompt*

Com a adoção generalizada da IA Generativa, o conceito de engenharia de *prompts* ganhou destaque no discurso e na prática. A engenharia de *prompts* pode ser entendida como o processo de criação de instruções específicas para guiar modelos de linguagem, como os de geração de texto por IA, a fim de obter resultados desejados em tarefas, como a criação de notícias. Ao ajustar essas instruções por meio da engenharia de *prompts*, os jornalistas podem moldar os resultados do modelo para atender às necessidades da produção de notícias. Vale mencionar que, além do uso de *prompts* eficientes, algumas redações estão desenvolvendo seus próprios LLMs para a produção de conteúdo - como a Bloomberg com o BloombergGPT. O BloombergGPT está sendo desenvolvido e treinado com aproximadamente 50 bilhões de parâmetros. Por uma breve comparação, o GPT-3 foi treinado com 175 bilhões de parâmetros em 2020 (Floridi; Chiriatti, 2020).

5 Metodologia

A estratégia metodológica adotada nesse estudo é qualitativa e exploratória, consistindo em duas etapas distintas. Primeiramente, é realizada uma revisão de literatura, que envolve a definição de critérios para a avaliação de *software* de GenAI. Essa etapa também inclui o desenvolvimento de um protocolo e um roteiro de teste. A revisão de literatura, como enfatizado por Lakato e Marconi (2003), vai além da simples síntese do conhecimento existente, oferecendo uma oportunidade para explorar o tema sob uma nova perspectiva e originar conclusões inovadoras.

A segunda etapa envolve a seleção de *software* para teste, seguida pela aplicação do protocolo definido, usando o roteiro de teste e a análise subsequente dos resultados. A revisão de literatura desempenha um papel fundamental na construção da base teórica para a compreensão da geração automática de texto, especialmente no contexto de modelos de linguagem em grande escala, como GPT (*Generative Pre-trained Transformers*) e LaMDA (*Language Model for Dialogue Applications*). Esses modelos formam a base dos objetos empíricos escolhidos para exame nesse estudo. Além disso, a revisão de literatura auxilia na definição dos critérios de avaliação e na formulação do protocolo de teste e do roteiro de aplicação.

Os critérios de avaliação, que são baseados nos conceitos de estrutura da notícia e *lead* (Lage, 1987), foram considerados os mais apropriados, tendo em vista estudos anteriores (Arce, 2009; Santos, 2016) que teorizaram e demonstraram a possibilidade de construção de textos jornalísticos utilizando *software*. Os critérios de avaliação incluem: título, gravata (subtítulo),

lead, corpo do texto, uso de fontes, precisão, gramática e ortografia, componentes vistos em manuais de redação jornalística (Martins Filho, 1997).

Além disso, com base nos estudos de Reiter e Belz (2009), que discutem a validade de algumas métricas para avaliar automaticamente *software* de geração de linguagem natural, considera-se as avaliações humanas desses *softwares* como essenciais. Ao longo do tempo, os textos gerados por NLG têm sido avaliados de diferentes formas, “a maioria das quais pode ser classificada como avaliações baseadas em desempenho de tarefa, julgamentos e avaliações humanas, ou comparação com textos de corpora usando métricas automáticas” (Reiter; Belz, 2009, p. 531, tradução livre)³. Nesse sentido, de acordo com Reiter e Belz (2009), em comparação com outras formas de avaliação, as avaliações humanas são mais rápidas e baratas, além de não exigirem o suporte de especialistas na área. Adiciona-se que essas avaliações são essenciais para compreender a adequação e aceitação dos textos gerados, uma vez que a linguagem é repleta de nuances e preferências. No entanto, reconhece-se que, para garantir resultados consistentes ao comparar diferentes *softwares*, é primordial criar um protocolo e um roteiro de teste. Enfatiza-se que, por se tratar de um estudo exploratório, o protocolo e o roteiro descritos na próxima subseção, inspirados nos conceitos de testes de aceitação, não constituem um esquema fixo, mas uma abordagem que outros pesquisadores podem explorar e analisar criticamente, com o objetivo de promover a construção e o avanço contínuo da pesquisa. (Sommerville, 2007).

A escolha de uma metodologia qualitativa e exploratória é justificada pela necessidade de uma compreensão detalhada da avaliação dos programas de GenAI. A segunda etapa, envolvendo a aplicação prática do protocolo de teste em modelos avançados como GPT e LaMDA, assegura a relevância e a validação empírica dos conceitos teóricos. Essa metodologia combina uma base teórica com uma análise prática detalhada, contribuindo para o avanço do conhecimento na área de GenAI e suas possíveis aplicações no jornalismo.

Os objetos empíricos selecionados para teste nesse estudo são três *softwares* de geração de texto em linguagem natural (NLG): ChatGPT da OpenAI, Gemini AI do Google e MariTalk da Maritaca AI. O ChatGPT, lançado em 30 de novembro de 2022, foi escolhido por sua ampla popularidade entre o público geral. Por sua vez, o Gemini AI, que se tornou disponível no Brasil posteriormente, em 12 de julho de 2023, foi incluído no estudo por ser considerado o principal concorrente do ChatGPT. O MariTalk foi escolhido por se tratar de um projeto de uma *startup*

³ No original: “most of which can be classified as evaluations based on task performance, human judgments and evaluations, or comparison with corpus texts using automatic metrics”.

brasileira e possuir capacidades similares às do ChatGPT, utilizando o modelo Sabiá-65B (Pires *et al.*, 2023).

É relevante enfatizar que, apesar de esses softwares de NLG utilizarem diferentes modelos de linguagem, como GPT-3.5, LaMDA e Sabiá-65B, todos são relevantes para o propósito desse estudo. Vale ressaltar que os testes foram realizados em dois momentos, ao longo de um período de quase um ano. Assim, foi possível verificar se as atualizações desses programas forneceram mudanças nos resultados. Além disso, cada teste foi realizado com um único *prompt*, sem repetições ou tentativas de refinamento. Essa abordagem foi escolhida para que o objetivo não fosse treinar a Inteligência Artificial para obter melhores resultados, mas compreender como o *software* responde à primeira solicitação.

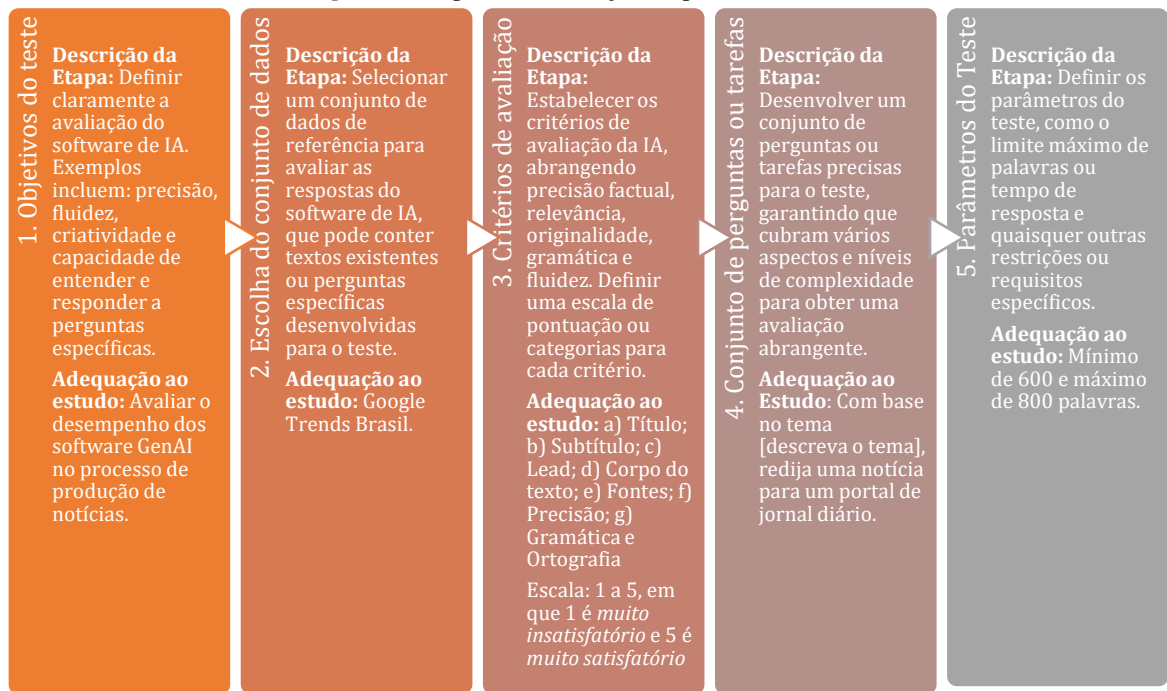
Os testes foram conduzidos em 14 e 16 de agosto de 2023, segunda e quarta-feira, respectivamente, pela manhã. Outra rodada de testes foi realizada em 14 e 16 de julho de 2024, também segunda e quarta-feira. O objetivo foi coletar dados para uma análise comparativa não apenas entre os *softwares*, mas também em diferentes dias e anos, e verificar possíveis nuances nos resultados.

6 Protocolo e roteiro de teste

Esta subseção descreve as etapas para a elaboração do protocolo e do roteiro de teste. Aqui, inspira-se nos conceitos de Engenharia de *Software*, especificamente no que diz respeito aos testes de aceitação de *software*. “Os testes de aceitação também podem revelar problemas de requisitos onde as funcionalidades do software não atendem às necessidades do usuário ou o desempenho do software é inaceitável” (Sommerville, 2007, p. 28).

As etapas para a elaboração do protocolo de teste (Figura 1) consistem em descrever os objetivos do teste; escolher um conjunto de dados; determinar os critérios de avaliação; desenvolver um conjunto de perguntas ou tarefas às quais o *software* será submetido e estabelecer parâmetros para o teste.

Figura 1: Etapas na elaboração do protocolo de teste



Fonte: autora (2023).

Com base no protocolo delineado, um roteiro de teste (Figura 2) foi desenvolvido para orientar o processo de pesquisa. Esse roteiro compreende componentes essenciais, incluindo uma descrição do *software* a ser testado (ChatGPT, Gemini AI e MariTalk), o cenário (redação de uma notícia sobre um tema específico), dados de entrada (adaptados para esse estudo com base nos resultados do Google Trends Brasil) e resultados esperados (uma notícia adequada para um portal de notícias diário, avaliada de acordo com os critérios estabelecidos). Esclareceu-se que o critério de 600 a 800 palavras foi estabelecido, considerando as peculiaridades dos LLMs. Esse intervalo melhora a eficiência, evita respostas prolixas e facilita a avaliação em escala, além de simular cenários reais em que a concisão é fundamental para a usabilidade.

O roteiro de teste foi ajustado para enfatizar os julgamentos humanos como um aspecto fundamental da avaliação intrínseca, alinhando-se às práticas na experimentação de *software* de Geração de Linguagem Natural (NLG) desde meados da década de 1990 (Reiter; Belz, 2009). Essa abordagem visa garantir uma avaliação abrangente e detalhada do desempenho do *software* na geração de notícias, levando em consideração os critérios estabelecidos nas fases anteriores da pesquisa.

Figura 2: Etapas na elaboração do protocolo de teste

Software	Cenário	Dados de entrada	Resultados esperados	Critérios de avaliação	Título	Subtítulo	Lead	Corpo do texto	Fontes	Precisão	Gramática e Ortografia
Nome 1				Escala de pontuação (1 to 5)							
Nome 2											
Nome 3											

Fonte: autora (2023).

O primeiro teste desse estudo foi realizado na manhã de segunda-feira, 14 de agosto de 2023. Nessa data, o tópico “Flamengo vs. São Paulo” era a principal tendência no Google Trends Brasil, com 2 milhões de buscas. A partida de futebol entre Flamengo e São Paulo, válida pela 19ª rodada do Campeonato Brasileiro (Brasileirão), ocorreu no domingo (13) no Estádio do Maracanã, no Rio de Janeiro, resultando em um empate de 1 a 1.

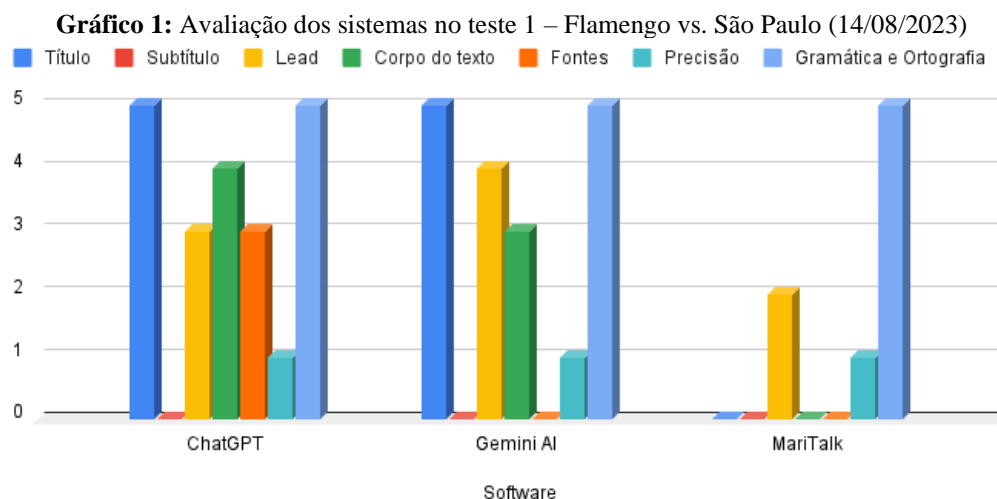
O segundo teste foi realizado na manhã de quarta-feira, 16 de agosto de 2023. O tópico “preços da gasolina” apareceu como a principal tendência no Google Trends Brasil, com 400.000 buscas. No dia anterior (15/08/2023), a Petrobrás havia anunciado um aumento de R\$0,41 por litro no preço da gasolina e R\$ 0,78 por litro no preço do diesel.

O terceiro teste foi conduzido na segunda-feira, 15 de julho de 2024. O tópico no Google Trends Brasil foi “Dia dos Homens”, com aproximadamente 100.000 buscas. Esse dia foi criado no Brasil para conscientizar a população masculina sobre a importância de cuidar da saúde.

O quarto teste foi realizado na quarta-feira, 17 de julho de 2024. O tópico no Google Trends Brasil foi “Nubank”, com mais de 200.000 buscas. A instituição financeira Nubank enfrentou falhas em seu sistema, deixando os clientes sem acesso às suas contas bancárias.

7 Resultados e discussão

Testou-se cada programa, iniciando uma nova conversa com o *prompt*: “Com base no tema Flamengo vs. São Paulo, escreva uma notícia para um portal de notícias diário”. Os resultados (Gráfico 1) mostraram um desempenho semelhante entre o ChatGPT e o Gemini AI. Além disso, houve semelhança em certos critérios de avaliação entre os três *softwares* avaliados.



Fonte: autora (2024).

Os títulos *Flamengo e São Paulo jogam partida eletrizante no Maracanã* (ChatGPT) e *Flamengo vence São Paulo e permanece no topo do Brasileirão* (Gemini AI) receberam nota 5, considerada muito satisfatória na escala de pontos. O MariTalk não foi avaliado nesse critério, pois não entregou um título. Em relação ao subtítulo, nenhum programa apresentou resultados. É importante notar que, nesse ponto, a precisão do título não foi considerada, uma vez que o resultado da partida foi um empate, e não uma vitória do Flamengo, como indicado pelo Gemini AI.

No critério *lead*, observou-se variações nas notas atribuídas aos diferentes *softwares* testados. O *lead* “é o relato do fato principal de uma série, o que é mais importante ou interessante” (Lage, 1987, p. 27). Ao analisar os textos gerados com base no esquema clássico de Laswell (o que - quem - quando - onde - como e por que), o Gemini AI obteve resultados superiores. O *software* apresentou os elementos do *lead* e expandiu o “como” e o “porquê” ao longo do corpo do texto. Nesse critério, o ChatGPT falhou ao não mencionar que se tratava de uma partida do Campeonato Brasileiro (o que). O MariTalk não forneceu informações sobre onde a partida aconteceu.

Na avaliação do Corpo do texto, o ChatGPT e Gemini AI apresentaram desempenhos semelhantes, com uma leve vantagem para o primeiro, já que o desenvolvimento dos parágrafos foi mais eficiente. No entanto, é importante destacar que, por se tratar de uma notícia de futebol, o Gemini AI forneceu uma ficha técnica do jogo, item relevante em textos de cobertura esportiva. Nesse aspecto, o MariTalk não foi avaliado, pois entregou apenas um parágrafo de texto.

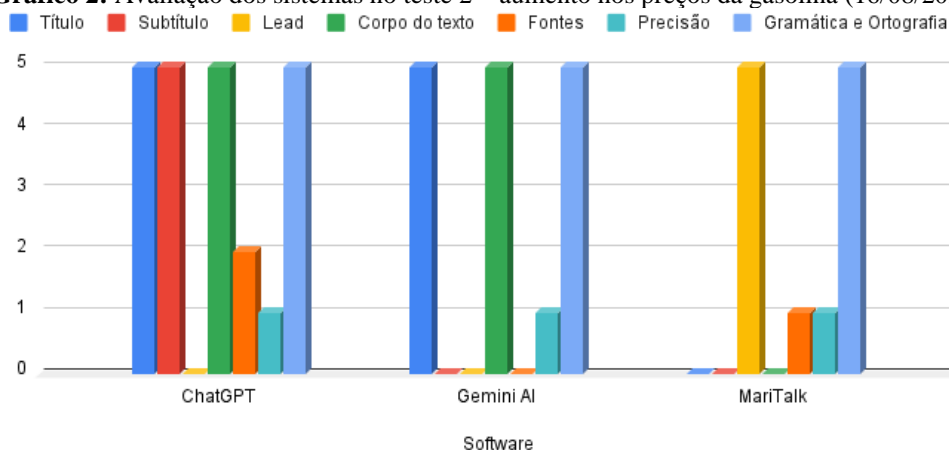
No critério fontes, o ChatGPT foi o único a fornecer uma declaração. “Foi um grande espetáculo para os fãs de futebol. As equipes mostraram muita determinação e qualidade em campo, disse o técnico do Flamengo”. Esses resultados estão alinhados às observações de Berti

(2023), que identificou a capacidade do ChatGPT de gerar textos, fazendo referência a figuras reais quando solicitado. No teste realizado, embora o programa não tenha mencionado o nome do técnico da equipe, formulou uma frase contextualizada no cenário do futebol. Contudo, é importante considerar as possíveis alucinações da IA Generativa (Cortiz, 2023).

Em termos de precisão, todos os sistemas receberam nota 1, considerada muito insatisfatória, pois não atenderam aos requisitos necessários, como nomes, datas, relações entre pessoas e informações (Martins Filho, 1997). Quanto à gramática e ortografia, todos os *softwares* receberam nota 5, sem erros identificados. Ao testar o ChatGPT, Santos (2023) também constatou que os textos eram bem estruturados. Isso destaca a perspectiva técnica sobre o funcionamento dos modelos de linguagem baseados em *Large Language Models* (LLMs). Essencialmente, os modelos de linguagem GPT são capazes de criar textos de forma estatística, calculando a probabilidade de uma palavra se encaixar na sequência de outra, com base em treinamento prévio extenso e grandes conjuntos de dados (Floridi; Chiriatti, 2020; Lee, 2023). Apesar das explicações de Lug (2023) e Pichai (2023) sobre o funcionamento do modelo de linguagem LaMDA - que também depende de avaliadores humanos para garantir a eficácia das respostas do *software* -, o modelo é composto por bilhões de palavras e dados baseados em estatísticas para prever as próximas palavras.

No segundo teste, foram avaliados os programas criando uma nova conversa com o *prompt*: “Com base no tema Aumento dos Preços da Gasolina, escreva uma notícia para um portal de notícias diário”. Aqui, enfatiza-se a necessidade de iniciar sempre uma nova conversa com o *software* para evitar que ele aprenda com *prompts* anteriores usados na mesma seção de chat. De modo geral, observou-se que os resultados (Gráfico 2) mostraram uma leve melhoria em comparação ao teste anterior.

Gráfico 2: Avaliação dos sistemas no teste 2 – aumento nos preços da gasolina (16/08/2023)



Fonte: autora (2024).

Os títulos Preços da Gasolina Alcançam Novo Pico, Gerando Preocupações Econômicas (ChatGPT) e Aumento da Gasolina Prejudica a Economia e a População (Gemini AI) receberam nota 5, considerada muito satisfatória. O MariTalk, mais uma vez, não entregou um título. Nessa nova rodada de testes, o ChatGPT apresentou o subtítulo A alta contínua nos preços dos combustíveis gera debate sobre o impacto no bolso dos consumidores e na inflação, que foi considerado muito satisfatório.

Quanto ao *lead*, o MariTalk foi o único a fornecer um resultado que incluía todos os elementos, segundo o modelo de Laswell. No entanto, apesar dessa eficácia, o *software* produziu apenas um parágrafo, o que impossibilitou a avaliação no critério corpo do texto. Nesse aspecto, tanto o ChatGPT quanto o Gemini AI tiveram um desempenho muito bom, inclusive em comparação ao teste anterior. Vale destacar que o ChatGPT desenvolveu um texto seguindo uma estrutura com título, subtítulo, data, nome do jornalista e local (cidade e estado). O Gemini AI trouxe uma característica distinta ao incluir uma seção com dicas para economizar gasolina, semelhante à ficha técnica fornecida no primeiro teste, conhecida no jornalismo como informação de serviço.

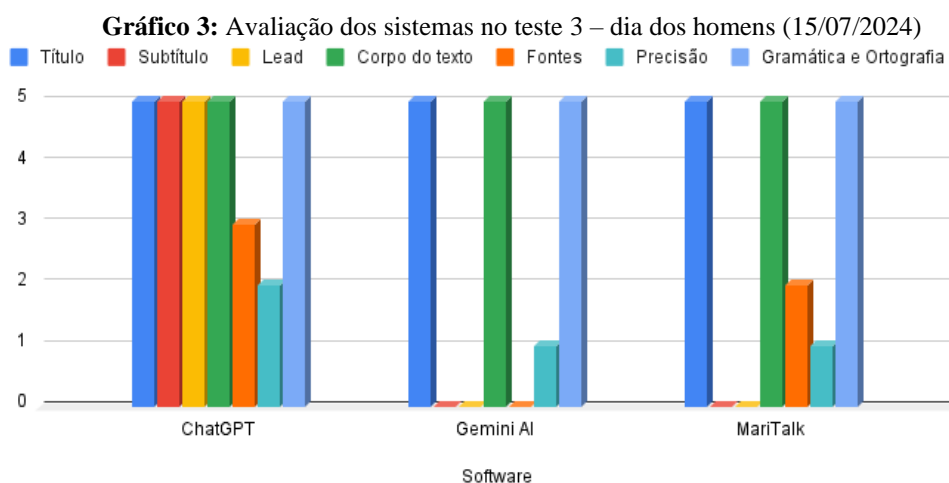
No critério fontes, o ChatGPT apresentou um texto sugerindo o espaço para edição das informações: segundo dados fornecidos pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), o preço médio da gasolina atingiu [valor] por litro em [data], marcando um aumento significativo em relação ao mês anterior. Essa abordagem foi mais eficaz do que o uso de aspas, como no teste 1. No mesmo critério, o MariTalk parafraseou declarações da Petrobras e do Governo Federal.

No critério precisão, a avaliação foi novamente insatisfatória. Nenhum dos *softwares* apresentou atributos que pudessem ser considerados factuais. O Gemini AI inseriu a informação de que o litro da gasolina custava R\$7,00 no texto, enquanto o MariTalk mencionou um aumento de 10% no preço da gasolina. Isso destaca, mais uma vez, a importância das possíveis alucinações nesses sistemas, conforme enfatizado por Cortiz (2023). Vale ressaltar que a versão GPT-3.5 do ChatGPT, testada nesse estudo nos primeiros e segundos testes, foi treinada com dados anteriores a setembro de 2021, e foi apenas em setembro de 2023 que o *software* começou a ter acesso a conteúdo disponíveis na internet. Um ponto relevante a ser mencionado é o número de parâmetros com os quais cada *software* foi treinado. Enquanto o GPT-3.5 é baseado em 175 bilhões de parâmetros, o modelo de linguagem em português, MariTalk utiliza 65 bilhões de parâmetros (Pires *et al.*, 2023). No entanto, todos os programas obtiveram novamente uma nota 5, muito satisfatória, no critério gramática e ortografia. Isso garante os propósitos

desses *softwares* como modelos de linguagem em larga escala, capazes de calcular estatisticamente a sequência de palavras (Floridi; Chiriatti, 2023).

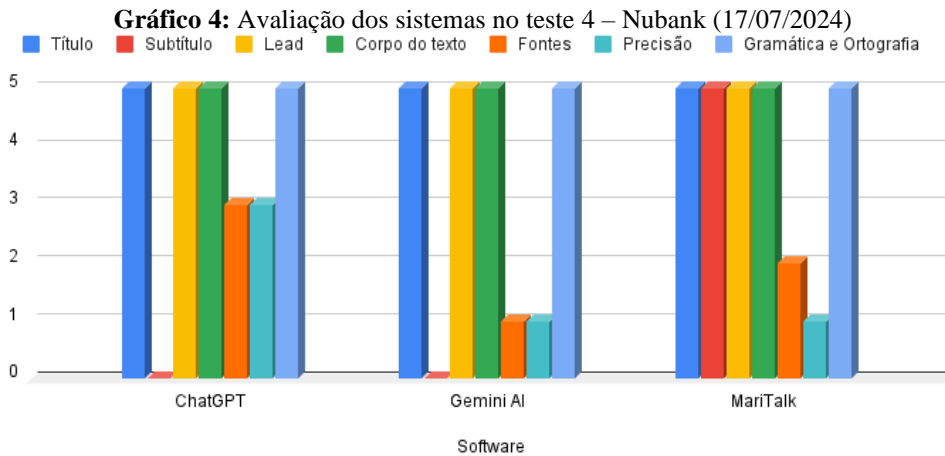
A nova rodada de testes foi realizada em julho de 2024. Todos os sistemas testados nesse estudo passaram por atualizações. Conseqüentemente, os testes 3 e 4 (Gráfico 3; gráfico 4) mostraram uma leve melhoria em relação aos resultados anteriores.

O terceiro teste, conduzido com o *prompt* “Com base no tema Dia dos Homens, escreva uma notícia para um portal de notícias diário”, destacou-se, especialmente, pelos resultados do ChatGPT (Gráfico 3). Isso porque o *software* forneceu fontes e dados para contextualizar a notícia. Embora as fontes tenham sido nomeadas como sendo do Ministério da Saúde, não foram especificados nomes individuais. No entanto, os dados apresentados, provenientes do Instituto Nacional do Câncer (INCA), são reais. Nessa segunda rodada de testes, o MariTalk também mostrou melhorias, como um título e um corpo de texto mais completos.



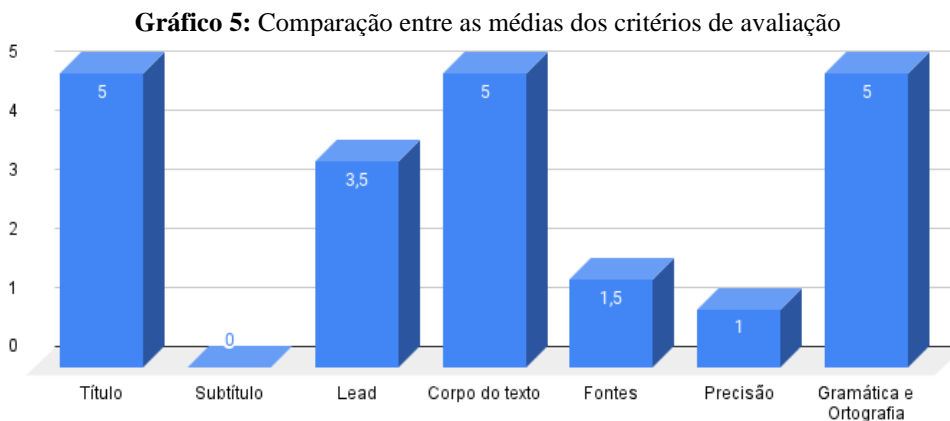
Fonte: autora (2024).

No quarto e último teste, utilizando o *prompt* “Com base no tema Falha no aplicativo do Nubank, escreva uma notícia para um portal de notícias diário”, tanto o ChatGPT quanto o MariTalk merecem destaque especial (Gráfico 4). Além de todos os parâmetros apresentados nos testes anteriores, no quarto teste, o ChatGPT incluiu corretamente o endereço de e-mail da assessoria de imprensa do Nubank e uma nota de atualização para a notícia. Por outro lado, o MariTalk se destacou ao fornecer dois segmentos, um no início e outro no final do artigo, explicando que o texto gerado era um exemplo genérico de notícia.



Fonte: autora (2024).

Outro resultado observado é que nenhum dos programas alcançou o parâmetro mínimo de 600 palavras nos quatro testes realizados. Em resumo, ao comparar os testes e considerar as implicações à luz das conclusões de Berti (2023) e Santos (2023), pode-se concluir que os programas analisados nesse estudo têm potencial para auxiliar na geração de textos. Na média entre os quatro testes realizados (Gráfico 5), observa-se que alguns critérios possuem melhor desempenho como título, corpo do texto e gramática e ortografia.



Fonte: autora (2024).

No entanto, não é viável classificá-los como *software* de produção de notícias, uma vez que elementos fundamentais para a estrutura jornalística, como o *lead*, a incorporação de fontes e a precisão (Lage, 1987; Martins Filho, 1997), não demonstraram um desempenho substancial. Na verdade, essa limitação não invalida o uso desses *softwares* em redações, especialmente considerando a eficácia demonstrada pela IA Generativa em sugerir pautas e editar textos, como evidenciado por Berti (2023).

8 Considerações finais

Esse artigo buscou explorar os limites e o potencial dos *softwares* de IA Generativa (GenAI) para a produção de notícias, respondendo à seguinte questão: até que ponto e de que maneira os *softwares* de GenAI podem contribuir para a produção de notícias? A análise revelou que, embora os *softwares* de GenAI não gerem conteúdo jornalístico *per se*, demonstraram sua viabilidade como assistentes em redações.

A discussão sobre o desempenho dos *softwares*, bem como seus potenciais e limitações, mostrou que, apesar de representarem um apoio no processo de geração de textos, os objetos empíricos desse estudo ainda não atingiram um nível de maturidade que permita realizar o trabalho perfeitamente sem intervenção humana. A constatação de que o ChatGPT, Gemini AI e MariTalk podem atuar como assistentes de escrita, sugerindo ideias e estruturas, é acompanhada pela compreensão de que seu uso exige um equilíbrio cuidadoso entre automação e discernimento editorial humano.

Destaca-se também que esses sistemas continuam sendo aprimorados, e que testes futuros podem ou não alcançar resultados melhores, uma vez que eles não foram necessariamente projetados para tarefas jornalísticas. Por outro lado, salienta-se que a solução de problemas é uma característica central das pesquisas em computação. Esse fator sugere que o rápido avanço da área possa, em breve, oferecer soluções técnicas para falhas que ainda persistem.

É importante observar que o uso desses sistemas exige competências na elaboração de *prompts* adequados ao contexto jornalístico. Portanto, esses testes também podem sugerir caminhos potenciais para o futuro ou presente do jornalismo, particularmente por meio de uma função conhecida como especialista em *prompts*. Propõe-se uma reflexão sobre como as competências necessárias para atuar com a GenAI podem contribuir para novas funções, ou até mesmo perfis nas redações, assim como as mídias sociais desencadearam o *social media*.

Mesmo com as limitações apontadas, também se enxerga a possibilidade de utilizar esses *softwares* para produzir material complementar, desdobramentos temáticos e até para transformar textos já existentes, que poderiam ser empregados em outros produtos jornalísticos além dos textos em portais. Outra perspectiva, que poderia ser um caminho alternativo para o jornalismo diante dos LLMs, seria o investimento em *softwares* proprietários, como iniciativas do tipo BloombergGPT. No entanto, as especificidades do jornalismo e os custos de desenvolvimento de *software*, como lembra Beckett (2019), podem representar um grande desafio para a implementação desses sistemas pelas empresas de mídia.

Quanto aos procedimentos utilizados nessa pesquisa, destacou-se possíveis fragilidades no método empregado, especialmente no desenvolvimento do protocolo e do roteiro de testes. Reconhece-se que, embora os testes tenham sido sistemáticos, essa abordagem é suscetível a adaptações e melhorias para pesquisas futuras.

No tocante a estudos futuros, sugere-se a ampliação dessa pesquisa para um contexto de redação jornalística cotidiana, englobando a colaboração direta entre esses *softwares* e jornalistas profissionais. A aplicação prática em uma redação permitiria uma compreensão mais abrangente de como os *softwares* de GenAI interagem com fluxos de trabalho já existentes, que podem melhor clarificar as dinâmicas entre Inteligência Artificial e a criatividade humana na produção de notícias.

Por fim, embora os *softwares* de IA Generativa, atualmente, não substituam integralmente o papel humano na produção de notícias, demonstraram seu valor como assistentes em redações e facilitadores potenciais na criação e ampliação de conteúdo jornalístico. À medida em que se avança para uma era cada vez mais tecnológica, é imperativo continuar explorando e aprimorando essa interação entre Inteligência Artificial e Jornalismo, mantendo um olhar crítico sobre suas aplicações e implicações éticas. Como as questões éticas em torno da IA permanecem amplamente debatidas na sociedade, é essencial que a formação jornalística inclua discussões entre professores e futuros profissionais, integrando esses debates à prática em sala de aula.

Referências

ARCE, T. O lead automatizado: uma possibilidade de tratamento da informação para o jornalismo impresso diário. **Revista Exacta**, São Paulo, v. 2, n. 3, p. 253-262, 2009. DOI: <http://dx.doi.org/10.18674/exacta.v2i3.253>. Disponível em: <https://revistas.unibh.br/dcet/article/view/253>. Acesso em: 18 dez. 2024.

BECKETT, C. New powers, new responsibilities: a global survey of journalism and artificial intelligence. The London School of Economics and Political Science. Polis Journalism and Society. **Google News Initiative**, 2019. Disponível em: <https://bit.ly/3P4tf4G/>. Acesso em: 26 July 2023.

BERTI, O. M. C. **ChatGPT**: evolução ou fim do jornalismo? São Luís: EdUESPI, 2023.

CARREIRA, K. A. C. **Notícias automatizadas**: a evolução que levou o jornalismo a ser feito por não humanos. 2017. 205 f. Dissertação (Mestrado em Comunicação Social) — Programa de Pós-Graduação em Comunicação Social, Universidade Metodista de São Paulo, São Bernardo do Campo, 2017. Disponível em: https://bdtd.ibict.br/vufind/Record/METO_edfc0ee1d54839c2b5477b8e783ec6ce. Acesso em: 18 dez. 2024.

CANAVILHAS, J. **Manual de jornalismo na web**. Covilhã: LabCom UBI, 2024.

CANAVILHAS, J. Produção automática de texto jornalístico com IA: contributo para uma história. **Textual & Visual Media**, Madri, v. 17, p. 22-40, 2023.

DOI: <https://doi.org/10.56418/txt.17.1.2023.2>. Disponível em:

<https://textualvisualmedia.com/index.php/txtvmedia/article/view/319>. Acesso em: 18 dez. 2024.

CORTIZ, D. **Entendendo as alucinações do ChatGPT**. Disponível em:

<https://bit.ly/45sKu4O/>. Acesso em: 06 jul. 2023.

DÖRR, K. Mapping the field of algorithmic journalism. **Digital Journalism**, London, v. 4, n. 6, p. 700-722, 2015. DOI:10.1080/21670811.2015.1096748. Disponível em:

https://www.researchgate.net/publication/282642995_Mapping_the_field_of_Algorithmic_Journalism. Acesso em: 18 Dec. 2024.

FLORIDI, L.; CHIRIATTI, M. GPT-3: its nature, scope, limits, and consequences. **Minds and Machines**, Dordrecht, v. 30, p. 681-694, 2020. Disponível em:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3827044. Acesso em: 18 Dec. 2024.

LAGE, N. **Estrutura da notícia**. São Paulo: Editora Ática, 1987.

LAKATO, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003.

LEE, A. Para que são usados os grandes modelos de linguagem? **Nvidia**, 29 mar. 2019.

Disponível em: <https://bit.ly/43ISspe>. Acesso em: 10 jul. 2023.

LIMA JUNIOR, W. T.; PAULA, F. G. Funções do jornalista em ecossistema informativo digital conectado composto por máquinas sociais pré-cognitivas. **Comunicação & Sociedade**, Braga, v. 43, n. 2, p. 129-153, 2021. DOI: <https://doi.org/10.15603/2176-0985/cs.v43n2p129-154>. Disponível em:

<https://revistas.metodista.br/index.php/comunicacaosociedade/article/view/378>. Acesso em: 18 Dec. 2024.

LUG, A. Lambda vs. ChatGPT – Similaridades e diferenças. **Comparações, Ferramentas**.

André Lug. 13 mar. 2023. Disponível em: <https://bit.ly/3KndrXI>. Acesso em: 26 jul. 2024.

MARTINS FILHO, E. L. **Manual de redação e estilo de O Estado de S. Paulo**. 3. ed. São Paulo: O Estado de S. Paulo, 1997.

PAVLIK, J. V. Collaborating with ChatGPT: considering the implications of generative artificial intelligence for journalism and media education. **Journalism & Mass**

Communication Educator, Londres, v. 78, p. 84-93, 2023. DOI: <https://us.sagepub.com/en-us/journals-permissions>. Disponível em: <https://bpw-us-w2.wpmucdn.com/hawksites.newpaltz.edu/dist/7/800/files/2023/02/collaborating.pdf>. Acesso em: 18 Dec. 2024.

PICHAU, S. An important next step on our AI journey. **Google Blog**, 6 fev. 2023. Disponível em: <https://bit.ly/3KMG0Ou>. Acesso em: 01 jul. 2023.

PIRES, R. *et al.* Sabiá: Portuguese large language models. **AirXiv**, 2023. DOI <https://doi.org/10.48550/arXiv.2304.07880>. Disponível em: <https://arxiv.org/abs/2304.07880>. Acesso em: 02 July 2023.

REITER, E.; BELZ, A. An investigation into the validity of some metrics for automatically evaluating natural language generation systems. **Computational Linguistics**, Cambridge, v. 35, n. 4, p. 529-558, 2009. DOI: 10.1162/coli.2009.35.4.35405. Disponível em: <https://aclanthology.org/J09-4008/>. Acesso em: 18 Dec. 2024.

SANTOS, M. C. What I learned interviewing a robot: notes on the experimental application of the EIAF methodology using the ChatGPT artificial intelligence tool. **Hipertext.net**, Barcelona, n. 26, p. 23-29, 2023. DOI: 10.31009/hipertext.net.2023.i26.04. Disponível em: https://www.researchgate.net/publication/371149895_What_I_learned_interviewing_a_robot_Notes_on_the_experimental_application_of_the_EIAF_methodology_using_the_ChatGPT_artificial_intelligence_tool. Acesso em: 18 Dec. 2024.

SANTOS, M. C. Narrativas automatizadas e a geração de textos jornalísticos: a estrutura de organização do lead traduzida em código. **Brazilian Journalism Research**, Brasília, v. 12, p. 160-185, 2016. DOI: <https://doi.org/10.25200/BJR.v12n1.2016.757>. Disponível em: <https://bjr.sbpjor.org.br/bjr/article/view/757>. Acesso em: 18 Dec. 2024.

SHANNON, C. E. Prediction and entropy of printed English. **Bell System Technical Journal**, New York, v. 30, p. 50-64, 1951. Disponível em: https://www.princeton.edu/~wbialek/rome/refs/shannon_51.pdf. Acesso em: 18 Dec. 2024.

SOMMERVILLE, I. **Engenharia de software**. 8. ed. São Paulo: Addison Wesley, 2007.

VASWANI, A. *et al.* Attention is all you need. *In*: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NIPS 2017), 31., 2017, Long Beach. **Anais [...]**. Long Beach: NIPS, 2017, p. 1-11. Disponível em: <https://bit.ly/483iwy7>. Acesso em: 05 July 2023.