UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE EDUCAÇÃO
PÓS-GRADUAÇÃO EM TECNOLOGIAS, COMUNICAÇÃO E EDUCAÇÃO

GUSTAVO HENRIQUE DE SOUZA MEDRADO

INTELIGÊNCIA ARTIFICIAL E INFORMAÇÃO QUALIFICADA: UMA ABORDAGEM EXPERIMENTAL PARA FUNDAMENTAR O JORNALISMO DE PROMPT

UBERLÂNDIA
FEVEREIRO DE 2026

GUSTAVO HENRIQUE DE SOUZA MEDRADO

INTELIGÊNCIA ARTIFICIAL E INFORMAÇÃO QUALIFICADA: UMA ABORDAGEM EXPERIMENTAL PARA FUNDAMENTAR O JORNALISMO DE PROMPT

Dissertação apresentada à Faculdade de Educação da Universidade Federal de Uberlândia para obtenção do título de Mestre no Programa de Pós-Graduação em Tecnologias, Comunicação e Educação.

Orientador: Prof. Dr. Marcelo Marques Araújo
Coorientador: Prof. Dr. Alexandre Gomes de Siqueira

UBERLÂNDIA
FEVEREIRO DE 2026

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

M492 Medrado, Gustavo Henrique de Souza, 1996-
2026 Inteligência Artificial e Informação Qualificada [recurso eletrônico] : Uma abordagem experimental para fundamentar o Jornalismo de Prompt / Gustavo Henrique de Souza Medrado. - 2026.

Orientador: Marcelo Marques Araújo.
Coorientador: Alexandre Gomes de Siqueira.
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Tecnologias, Comunicação e Educação.
Modo de acesso: Internet.
DOI http://doi.org/10.14393/ufu.di.2026.161
Inclui bibliografia.
Inclui ilustrações.

1. Educação. I. Araújo, Marcelo Marques,1975-, (Orient.). II. Siqueira, Alexandre Gomes de,1980-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-graduação em Tecnologias, Comunicação e Educação. IV. Título.

CDU: 37

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Tecnologias, Comunicação e Educação
Av. João Naves de Ávila, 2121, Bloco 1G, Sala 156 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3291-6395/6396 - ppgce@faced.ufu.br - www.ppgce.faced.ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Tecnologias, Comunicação e Educação				
Defesa de:	Dissertação de Mestrado Profissional, número 02/2026/196, PPGCE				
Data:	Vinte e seis de fevereiro de dois mil e vinte e seis	Hora de início:	17:10	Hora de encerramento:	18:45
Matrícula do Discente:	12412TCE012				
Nome do Discente:	Gustavo Henrique de Souza Medrado				
Título do Trabalho:	Inteligência Artificial e Informação Qualificada: Uma abordagem experimental para fundamentar o Jornalismo de Prompt				
Área de concentração:	Tecnologias, Comunicação e Educação				
Linha de pesquisa:	Mídias, Educação e Comunicação				
Projeto de Pesquisa de vinculação:	BRANDING, EMPREENDEDORISMO E DISCURSO NAS ORGANIZAÇÕES: Sentidos que percorrem o empreendedorismo de marcas (Grupo Beco)				

Reuniu-se de forma híbrida na sala 125 Bloco 1G / https://conferenciaweb.rnp.br/webconf/marcelo-marques-araujo, designada pelo Colegiado do Programa de Pós-graduação em Tecnologias, Comunicação e Educação, assim composta: Professores Doutores: Mirna Tonus - UFU; Angela Maria Grossi - Unesp; Marcelo Marques Araújo - UFU orientador do candidato.

Iniciando os trabalhos o presidente da mesa, Dr. Marcelo Marques Araújo, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado(a).

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Marcelo Marques Araujo**, **Professor(a) do Magistério Superior**, em 26/02/2026, às 18:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Mirna Tonus**, **Professor(a) do Magistério Superior**, em 26/02/2026, às 18:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Angela Maria Grossi**, **Usuário Externo**, em 27/02/2026, às 12:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7021784** e o código CRC **C6E73911**.

Referência: Processo nº 23117.004767/2026-43 SEI nº 7021784

Dedico este trabalho a todas as vítimas das consequências da desinformação durante a pandemia de covid-19, em especial à minha querida avó, Maria Coraci de Souza Medrado, e a um dos melhores amigos que a vida me deu, Lucas Kristhen Ferreira Muniz.

AGRADECIMENTOS

Ao meu amor, Jessica Rodrigues, ora Batman, ora Zorro, ora cogumelo do Mário que, por incontáveis noites, se fizeram uma pessoa só. Deitada, me fez companhia enquanto eu trabalhava na pesquisa madrugada adentro. Em pé, me apoiou em seus ombros sempre que me julguei incapaz de prosseguir. Pelas incontáveis provocações, por vezes irritantes, mas que sempre me fizeram pensar, ainda que doesse. Meu mundo é melhor porque você existe nele! Guardem esse nome, querido e gentil leitor. Ainda irão ouvir falar muito sobre ela.

À minha mãe Marlene, ao meu irmão Victor Hugo e ao meu padrasto Gilberto que, na reta final dessa pesquisa, diante das desventuras da vida, rearranjaram nossos mundos. Por assim ser, irônica e consequentemente, me proporcionaram o fôlego necessário para chegar até aqui — como sempre. E, especificamente ao Meu Dotô, duas batidinhas ao lado esquerdo do peito com o punho cerrado seguido pelo V de vitória (ou de Victor).

Ao André Cabral, meu psicólogo, por me ajudar a elaborar minhas questões, angústias e limitações, processo sem o qual este trabalho não teria saído do papel (ou entrado, com o perdão do trocadilho, rsrs) — como tantos outros.

À Universidade Federal de Uberlândia, minha casa desde a graduação, por ser uma instituição cada vez mais necessária ao povo brasileiro.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de pesquisa a mim fornecida, com a finalidade de apoiar a formação de recursos humanos em nível de pós-graduação, tornando viável a realização deste trabalho.

Ao professor Marcelo Marques, da FACED-UFU, meu orientador, responsável por acolher meus anseios acadêmicos lá nos primórdios quando, ainda desestruturados, os canalizou rumo ao Jornalismo de *Prompt*. Por, desde a graduação, não medir esforços para contribuir na minha jornada e na de inúmeros outros estudantes, se tornando, definitivamente, o tipo de mestre que busca ao máximo potencializar as capacidades de seus aprendízes. Da UFU pra vida!

Ao professor Alexandre Siqueira, da Universidade da Flórida, solícito e prestativo desde o primeiro contato, por aceitar a aventura de me coorientar. Por, reiterada vezes, acreditar que eu seria capaz e abrir as portas para nossa pesquisa lá do outro lado da América.

Ao professor Pedro Franklin, do IME-UFU, sempre gentil, didático e entusiasta, por ter topado a aventura de orientar alguém como eu, que caiu de paraquedas na ciência da computação, quando cursei a especialização em Ciência de Dados Aplicada. Lá, na

FACOM, desenvolvi as habilidades necessárias para executar partes da pesquisa ligadas à programação.

À professora Angela Grossi, minha futura orientadora no programa de doutorado em Mídia e Tecnologia da UNESP, o PPGMIT. Angela se tornou importante para esta pesquisa antes mesmo de nos conhecermos, quando eu assisti a uma palestra que ela proferiu na UFU e falou sobre uma premissa que ancorou o Prompt Jornalismo, o ecossistema da informação. Um ano depois, quisera a vida que nos encontrássemos num GT do INTERCOM e, desde então, nossos caminhos têm estado cada vez mais próximos. Sou grato por cada gentileza, orientação e apoio. Que seja apenas o começo de uma incrível jornada.

À professora Mirna Tonus, coordenadora do PPGCE durante o período em que desenvolvi este trabalho. Mirna foi minha orientadora na graduação e, desde então, sempre se fez muito acessível, atenciosa e assertiva em cada direcionamento. Sou extremamente grato por cada troca que tivemos ao longo de todos estes anos.

À professora Diva Silva, do PPGCE e do Jorna UFU, por compartilhar momentos de aprendizado dentro e fora da sala de aula. Diante do anúncio da premiação (sim, cara pessoa leitora, temos aqui uma pesquisa já premiada[1], iuhuul), Diva subiu ao palco do INTERCOM comigo, me apoiando quando minhas pernas ameaçaram não sustentar meu corpo. Diva é um exemplo de ser humano gentil, sábio e forte que, dentre tantos afazes, tem o árduo trabalho, ao qual me solidarizo e identifico, de transformar o substantitivo em verbo. É uma alegria poder ter aprendido tanto com ela. Agradeço também aos colegas que fiz em suas aulas, exímios educadores.

Aos colegas Diuly Tófolo, Gustavo Ribeiro, Graziela Guadagnin, Laura André e Joilsa Oliveira. Diuly foi da mesma turma que eu e foi a primeira colega que fiz logo nas aulas iniciais do PPGCE. Nossos papos sempre me fizeram pensar muito. Gustavo, meu xará, foi meu veterano no PPGCE. Olha que mundo pequeno. Eu havia sido o dele no Jorna UFU. Sou grato por cada papo sobre os universos das nossas pesquisas. Grazi também foi minha veterana e foi uma das maiores apoiadoras da minha jornada acadêmica. Para além de todo incentivo, é dela também a inspiração do design das apresentações (slides) desta pesquisa. Ela é uma designer, dentre tantas outras incríveis qualificações, de mão cheia! Laurinha foi minha caloura e se tornou uma grande parceira, de atividades internas da UFU aos congressos pelo Brasil. Dá gosto de ver tamanha dedicação! Joilsa, para além de colega de turma, também é servidora da Biblioteca da UFU. Sou grato pela orientação final referente a uma parte da normatização deste trabalho.

---

[1] CAVALCANTI, Marco. Pesquisador da UFU avalia se IA é capaz de verificar veracidade de informação. Comunica UFU, 10 jul. 2025. Atualizado em 18 jul. 2025. Disponível em: https://comunica.ufu.br/noticias/2025/07/pesquisador-da-ufu-avalia-se-ia-e-capaz-de-verificar-veracidade-de-informacao. Acesso em: [09 fev. 2026].

Ao meu *cunha* e a minha sogra preferida, Davi Rodrigues e Maria Solange, pela torcida e pelo apoio ao longo de todo o processo. Preparem a taca, ops, taça…

Aos mestres Flávio e Aurea Muniz, pais que a vida me deu, fundadores da Missão Escola da Vida, instituição tão importante na minha formação desde a adolescência. Ubuntu!

Aos amigos-parceiros-irmãos que ganhei do Parlamento Juvenil do Mercosul — versão jovem do Parlamento do Mercosul, ao qual representei o estado de Minas Gerais durante o Ensino Médio entre 2012 e 2014 –, em especial ao Tiago Nery e ao Keven Araújo. Lembro como se fosse ontem quando nós três, reunidos em um bar na Savassi em BH, vimos juntos e brindamos pelo resultado da minha aprovação no processo seletivo deste mestrado.

A todas as pessoas que vieram antes de mim e, aos seus modos, traçaram o caminho por onde meus pés, aos meus modos, trilharam até aqui. A cada menção que se faz presente nestes agradecimentos e tantas outras que, porventura, não couberam nesses singelos registros, minha mais sincera e afetuosa gratidão. Vocês, direta ou indiretamente, desempenham um papel fundamental no meu despertar a cada manhã. Obrigado por (r)existirem!

“A existência, porque humana, não pode ser muda, silenciosa, nem tampouco pode nutrir-se de falsas palavras, mas de **palavras verdadeiras**, com que os homens transformam o mundo.

Existir, humanamente, é pronunciar o mundo, é modificá-lo. O mundo pronunciado, por sua vez, se volta problematizado aos sujeitos pronunciantes, a exigir deles novo pronunciar.

[...]

Mas, se dizer a **palavra verdadeira**, que é trabalho, que é práxis, é transformar o mundo, dizer a palavra não é privilégio de alguns homens, mas direito de todos.”

(Freire, 1987, p.43)

MEDRADO, Gustavo Henrique de Souza. **Inteligência Artificial e Informação Qualificada: Uma abordagem experimental para fundamentar o Jornalismo de Prompt**. 2025. 111 p. Dissertação (Mestrado em Tecnologia, Comunicação e Educação – Universidade Federal de Uberlândia. Uberlândia, 2026.

RESUMO

Esta pesquisa investiga o potencial da Inteligência Artificial Generativa (IAG), em especial dos Modelos de Linguagem de Larga Escala (*Large Language Models* – LLMs), como ferramentas capazes de produzir informações qualificadas. Parte-se da hipótese de que os LLMs podem discernir informações. Para analisar essa premissa, foi criado um conjunto de dados composto por 1.200 pares de entradas, formados por notícias da agência de checagem Aos Fatos e respostas geradas pelo modelo GPT-4o do ChatGPT, formuladas por meio das técnicas *zero-shot* e *few-shot*, esta última com cadeia de pensamento (CoT). A análise envolveu a aplicação de medidas de similaridade semântica, métodos de clusterização (K-Means e Aglomerado Hierárquico) e testes estatísticos não-paramétricos (Teste do Sinal), a fim de aferir o desempenho do modelo em relação à verificação informacional. Os resultados revelaram um alto grau de correspondência entre os pares. As medianas de similaridade variaram de 0.80 a 0.95, indicando que, em mais de 90% dos casos, as respostas da IAG apresentaram boa ou quase total correspondência com os conteúdos das checagens humanas. As técnicas *few-shot CoT*, especialmente a aplicada com o modelo DeepSeek, demonstraram maior rigor crítico na avaliação das nuances semânticas. Esses achados corroboram a hipótese inicial, dando base para a fundamentação de um campo emergente no jornalismo digital: o Jornalismo de *Prompt*, ou *Prompt* Jornalismo, é uma área em fundamentação na algoritmosfera, situada na mediação entre humanos e a Inteligência Artificial Generativa, que se dedica ao processo de produção e consumo de informações qualificadas, por meio do conjunto das práticas discursiva, dialógica e técnica, denominada Promptação.

PALAVRAS-CHAVE

Inteligência Artificial Generativa; Jornalismo de *Prompt*; Checagem de Fatos; Informação Qualificada; Promptação.

Medrado, G. H. de S. (2026). **Artificial intelligence and qualified information: An experimental approach to grounding prompt journalism** (Master's dissertation, Federal University of Uberlândia). Uberlândia, Brazil.

ABSTRACT

This research investigates the potential of Generative Artificial Intelligence (GenAI), particularly Large Language Models (LLMs), as tools capable of producing qualified information. It starts from the hypothesis that LLMs can discern information. To analyze this premise, a dataset comprising 1,200 pairs of entries was created, formed by news from the fact-checking agency Aos Fatos and responses generated by ChatGPT's GPT-4o model, formulated using zero-shot and few-shot techniques, the latter with chain-of-thought (CoT). The analysis involved the application of semantic similarity measures, clustering methods (K-Means and Hierarchical Clustering), and non-parametric statistical tests (Sign Test), in order to assess the model's performance regarding informational verification. The results revealed a high degree of correspondence between the pairs. Similarity medians ranged from 0.80 to 0.95, indicating that, in more than 90% of cases, GenAI responses presented good or near-total correspondence with the contents of human fact-checks. Few-shot CoT techniques, especially the one applied with the DeepSeek model, demonstrated greater critical rigor in the evaluation of semantic nuances. These findings corroborate the initial hypothesis, providing a basis for the substantiation of an emerging field in digital journalism: Prompt Journalism, or Prompt Jornalismo, is an area being founded within the algorithmosphere, situated in the mediation between humans and Generative Artificial Intelligence, dedicated to the process of production and consumption of qualified information through the set of discursive, dialogic, and technical practices, denominated *Promptaction* ("Promptação").

KEYWORDS

Generative Artificial Intelligence; Prompt Journalism; Fact-Checking; Qualified Information; Promptaction.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE TABELA

LISTA DE QUADROS

SUMÁRIO

INTRODUÇÃO

Entre 2020 e 2022, durante a pandemia de coronavírus covid-19 causada pela Sars-CoV-2, vivenciamos (ou morremos em, depende de quem escreve e se escreve) uma sociedade à deriva, órfã de políticas públicas e de consumo de informações qualificadas. Em 2020, 94% dos brasileiros entrevistados viram, ao menos, uma notícia falsa sobre o coronavírus, sendo que 73% deles chegaram a acreditar em, pelo menos, uma delas. Ao mesmo tempo, 80% deles gostariam de ser avisados por verificadores de fatos sempre que fossem expostos a conteúdos falsos (Avaaz, 2020), já que 62% disseram (Sousa, 2020) não saber reconhecer uma notícia falsa. Tal realidade vem se agravando, sobretudo a cada campanha eleitoral, principal alvo da desinformação.

Nesse contexto, este trabalho analisa a aplicação de tecnologias emergentes, como a inteligência artificial generativa (IAG), mais especificamente uma parte de seu algoritmo, os modelos de linguagem de grande escala (*Large Language Models*, LLMs), no fortalecimento do ecossistema da informação, que compreende frentes de atuação como educação midiática, checadores de fatos, jornalismo profissional e regulamentação das plataformas (Grossi, 2024). Essa pesquisa busca amparo nas três primeiras frentes, respectivamente, compreendendo **a)** a **educação midiática**, ao analisar diferentes técnicas de interação humana-máquina, no intuito de compreender melhores práticas para lidar com os modelos, bem como seu funcionamento; **b)** a utilização da solução jornalística atual dada aos casos de desinformação, os **checadores de fatos**, tendo como parâmetro conteúdos criados por estes para medir a habilidade das IAG em realizar o mesmo feito e, ainda, **c)** a prospecção de um novo campo de estudo e atuação do **jornalismo profissional**, o Jornalismo de *Prompt*, uma vez que as IAG, ao mesmo tempo em que possibilitam novas formas de fazer e consumir informação, colocam em xeque, novamente, o modelo de negócio dos veículos jornalísticos (Vianna, 2025)

Como objetivo geral, avalia-se a capacidade de LLMs, enquanto expressão da Inteligência Artificial Generativa, em produzir informações qualificadas por meio da comparação entre seus *outputs* e checagens humanas, de modo a subsidiar a fundamentação inicial do campo emergente na comunicação que aqui se propõe, o Jornalismo de *Prompt*.

Como objetivos específicos, propõe-se: **a)** criar um conjunto de dados original com informações extraídas de agências de checagem e de chatbots baseados em LLMs; **b)** medir a aptidão desses modelos em checar informações por meio de métricas de similaridade semântica; **c)** comparar os resultados obtidos entre diferentes técnicas de *prompting* e LLMs avaliadores; e **d)** indicar as abordagens mais eficazes para a interação entre humanos e modelos. No Capítulo 4 - Aspectos Metodológicos, há um detalhamento mais robusto sobre cada etapa.

Há um enorme potencial batendo à porta, uma vez que esses modelos, presentes no ChatGPT, por exemplo, se popularizaram rapidamente. Em 2022, a OpenAI lançou a primeira versão do chatbot ChatGPT, a GPT-3.5. Ainda antes de uma semana do lançamento, já contava com 1 milhão de usuários. Com cerca de três meses, já havia ultrapassado a marca de 100 milhões (Haas, 2023). O Brasil se tornou, em março de 2024, o 4º país no ranking dos que mais usam o ChatGPT no mundo (Tunholi, 2024).

Diante de tamanha capilaridade, essa pesquisa se propôs a analisar, amparada pelo método hipotético-dedutivo, se a IAG, mais especificamente uma parte de seu algoritmo, os modelos de linguagem de grande escala (*Large Language Models*, LLMs), são capazes de **discernir** informações em rede. No âmbito desta etapa do trabalho, isso significa avaliar se o modelo principal da OpenAI, o gpt-4o (sucessor do gpt-4), é capaz de verificar informações a partir de perguntas feitas em sua interface para o usuário geral, o chatbot ChatGPT.

Os LLMs são treinados com um vasto banco de dados, utilizando técnicas avançadas de aprendizado de máquina e aprendizado profundo, ambas do campo das ciências da computação (IBM, 2024). Esses modelos utilizam redes neurais artificiais, uma arquitetura inspirada numa abstração do funcionamento do cérebro humano. Esses algoritmos aprendem a realizar associações estatísticas entre bilhões de palavras e frases para executar diversas tarefas que envolvem linguagem natural (Santaella, 2023).

A inteligência artificial generativa é um subgrupo dentro da Inteligência Artificial (IA), uma “tecnologia que permite que computadores e máquinas simulem a capacidade de resolução de problemas e a inteligência humana” (IBM, 2024), sobretudo no que tange ao aprendizado. Recebe o adjetivo “generativa” por ser capaz de generalizar ao gerar novos conteúdos a partir dos dados do seu treinamento. Lee e Qiufan (2021) destacam a importância do aprendizado profundo,

deep learning, na base desse processo, que pode ser aplicado em diversos campos do conhecimento para “previsão, classificação, tomada de decisões ou síntese”.

A originalidade deste trabalho está na proposição de uma abordagem empírica para avaliar o potencial verificativo de LLMs em ambiente real. Para isso, foi criado um conjunto de dados com 1.200 pares de entradas — compostos por checagens da agência Aos Fatos e respostas equivalentes geradas pelo GPT-4o —, com base em duas técnicas específicas de *prompting*: *zero-sho*t, na qual o modelo recebe apenas uma instrução direta; e *few-shot* com cadeia de pensamento (*few-shot*-CoT), que oferece exemplos e estrutura lógica de análise (Schulhoff et al., 2024).

A comparação entre as respostas humanas e as geradas pela IA foi feita por meio de medidas de similaridade semântica, classificadas em quatro faixas (de fraca à quase total), utilizando os próprios LLMs como avaliadores. Para refinar a análise, aplicaram-se métodos de clusterização não supervisionada, como o K-Means e o Aglomerado Hierárquico (James et al., 2023), permitindo identificar agrupamentos com desempenhos distintos. Além disso, a consistência estatística dos achados foi testada com o Teste do Sinal (Sprent; Smeeton, 2007), por se tratar de um conjunto de dados não normalmente distribuído.

Os resultados revelaram que mais de 90% das respostas da IA apresentaram medianas de similaridade entre boa (≥0.80) e quase total (0.95), sugerindo que os LLMs, quando submetidos a prompts bem-estruturados, conseguem gerar saídas compatíveis com padrões jornalísticos de verificação. Técnicas mais rigorosas de *prompting*, como o *few-shot-CoT* aplicado via DeepSeek, mostraram maior criticidade nas categorias de menor similaridade.

No horizonte do jornalismo de *prompt*, área emergente a qual esta pesquisa busca subsidiar, há um cenário no qual a interação entre jornalistas, LLMs e o público em geral possibilitará uma verificação qualificada, ágil e precisa das informações, contribuindo para minimizar os danos causados pela desordem informacional (Wardle; Derakhshan, 2023) e fortalecer o ecossistema informativo (Grossi, 2024).

De forma objetiva, cabe dizer que pesquisa é norteada por temas e conceitos centrais como a Inteligência Artificial Generativa (IAG) e, especificamente, os Modelos de Linguagem de Larga Escala (LLMs), investigando o emergente campo do Jornalismo de *Prompt*, que busca a produção e consumo de informação

qualificada em um contexto de combate à desinformação. Para tanto, a fundamentação teórica é direcionada a partir de autores como Lúcia Santaella (2023), que discute a natureza da inteligência artificial e suas implicações; Schulhoff et al. (2024), que abordam as técnicas de *prompting*; James et al. (2023), que detalham os métodos de agrupamento; e Wardle e Derakhshan (2023), que conceituam a desordem informacional, além de Paulo Freire (1987) e Bakhtin (1979), que inspiram a dimensão dialógica do trabalho, e Charaudeau e Maingueneau (2004), para a prática discursiva, e Sprent e Smeeton (2007), que sustentam os testes estatísticos não-paramétricos.

A abordagem metodológica adota um caráter experimental e empírico, amparado pelo método hipotético-dedutivo, e compreende a criação de um conjunto de dados original de 1.200 pares de entradas, extraídos de notícias verificadas pela agência Aos Fatos e de respostas geradas pelo modelo GPT-4o do ChatGPT. A coleta de dados se deu por *web scraping* e *prompting* estruturado, utilizando as técnicas *zero-shot* e *few-shot*, essa segunda com cadeia de pensamento (*CoT*). A análise dos dados envolveu a aplicação de medidas de similaridade semântica, métodos de clusterização (K-Means e Aglomerado Hierárquico) e testes estatísticos não-paramétricos, como o Teste do Sinal, para aferir o desempenho do modelo na verificação informacional.

A presente dissertação está organizada em capítulos subsequentes que detalham cada etapa da pesquisa, uma vez que as seções que se sucedem neste capítulo dizem respeito ao memorial do pesquisador. O Capítulo 2, "Referencial Teórico", aprofunda as discussões sobre o *prompting*, a informação qualificada, caminhos dialógicos possíveis para a interação humano-IA, a questão da inteligência da Inteligência Artificial Generativa e trabalhos relacionados ao tema, além de apresentar os métodos de agrupamento utilizados. O Capítulo 3, "A Pulga Atrás da Orelha", explora as incertezas e desafios relacionados às IAG, como as "alucinações"[2], e propõe o ideal de uma *blockchain* humanizada como alternativa. Em seguida, o Capítulo 4, "Aspectos Metodológicos", descreve os procedimentos e técnicas empregadas na pesquisa. Por sua vez, o capítulo 5, "Análise dos Resultados", apresenta e discute os achados empíricos, incluindo o Teste do Sinal, a medida de similaridade semântica, a análise exploratória dos agrupamentos e o

[2] Para entender o uso recorrente e deliberado de aspas na palavra “alucinação” e em suas variações ao longo de todo o trabalho, veja a discussão apresentada a respeito na seção 7.1.2 - Prompting.

desempenho real das técnicas de *prompting* e dos LLMs classificadores. O Capítulo 6 constitui um ensaio sobre os fundamentos do Jornalismo de *Prompt*. O Capítulo 7 introduz o conceito de Promptação como a aplicação das práticas do *Prompt* Jornalismo e, o Capítulo 8, encerra o trabalho desenvolvido até aqui com as considerações decorrentes da fruição desta pesquisa.

Ao longo da elaboração deste trabalho, para além da relação enquanto objeto de pesquisa, recorreu-se pontualmente a ferramentas de Inteligência Artificial Generativa, especialmente na organização textual do resumo, na sistematização da análise de trabalhos correlatos e na estruturação das referências bibliográficas. Esse uso foi guiado de forma ética e crítica, sem qualquer delegação de autoria, preservando-se integralmente a responsabilidade intelectual do pesquisador sobre as ideias, análises e conclusões apresentadas.

Àqueles que dispõem de tempo e sensibilidade — privilégios cada vez mais escassos — fica o convite para que, antes de se aprofundarem no trabalho desenvolvido a partir do Capítulo 2 – Referencial Teórico, conheçam as motivações e a trajetória do pesquisador, apresentadas nos subcapítulos a seguir, as quais tornaram possível o seu desenvolvimento.

1.1 SÓ SEI QUE FOI ASSIM

Entre o início de 2017 e até o final de 2020, trabalhei como assessor de comunicação no gabinete de um vereador na Câmara Municipal de Uberlândia. Ao final desse período, fui um dos coordenadores de comunicação da campanha deste mesmo vereador, mas na disputa pelo cargo majoritário. Conheci de perto como a máquina de desinformação funciona. Éramos alvos constantes das famigeradas *fakenews*, muitas vezes propagadas por perfis falsos dos quais, suspeitávamos, fossem alimentados por servidores em cargos comissionados de dentro da Prefeitura.

Já perdi as contas de quantas vezes tive que parar a produção de um conteúdo propositivo para analisar um conteúdo falso que tomávamos conhecimento, bem como seus possíveis impactos, para pensar numa estratégia de se e como deveria ser respondido. Não há pudor nesse meio. Aliás, nunca houve. Não é algo que compete apenas à modernidade e às tecnologias emergentes. No entanto, é perceptível o quanto a desinformação se tornou uma epidemia e tem

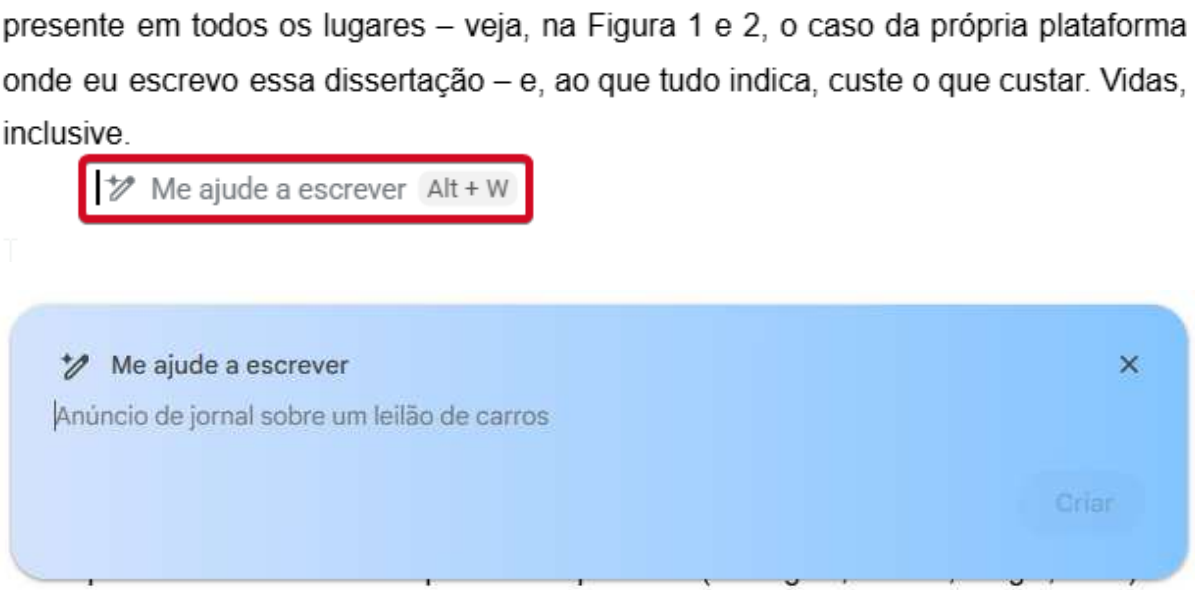
muito mais aderência nas redes sociais do que a própria informação, alcançando 70% a mais de engajamento (Vosoughi; Roy; Aral, 2018).

Além de ser um problema comunicacional, a desinformação se configura como uma problemática social, de segurança e de saúde pública. As eleições brasileiras de 2018 (Dourado, 2020), bem como a pandemia de coronavírus covid-19 mediante as ações e omissões do Governo Federal (Relatório Final da CPI da Pandemia, 2021), já evidenciaram o quanto a desinformação é capaz de fragilizar toda a estrutura de uma nação, inclusive à duras penas, como o abalo na democracia diante da tentativa de golpe de Estado no 08 de janeiro de 2023, bem como o saldo nefasto de 716.626 mil mortes (06/02/2026) – dentre elas minha saudosa avó paterna, Maria Coraci de Sousa Medrado, e meu grande amigo desde a adolescência, Lucas Kristhen Ferreira Muniz – em que parte significativa poderia ter sido evitada, caso o comportamento da população e dos governantes, dentre outros fatores, não tivesse sido afetado pelo forte ataque de informações inverídicas, expondo-a a riscos maiores (Oliveira; Oliveira, 2020).

E, aqui, peço licença para parafrasear o cantor e compositor Belchior (1976), num pedido encarecido aos que pretendem acompanhar essa dissertação. Não me peça que eu faça uma escrita como se deve, correta, branca, suave, muito limpa, muito leve. Dados, palavras, são navalhas e eu não posso escrever como convém sem querer contrariar ninguém. Trago um aperto no peito, um grito entalado na garganta. E fazer essa pesquisa, em busca de maneiras para garantir que qualquer pessoa, com um celular e acesso à Internet, possa checar as informações que recebe e, a partir disso, ter a opção de decidir como guiar seus passos com embasamento, é a forma como encontrei de fazer um acerto de contas. Ao menos por enquanto.

Reconheço minha limitação humana. Fazer essa pesquisa foi quase um manifesto. Sim, sei da necessidade do distanciamento do cientista para ccom o seu objeto. Ocorre que, neste caso, o objeto em questão tem sido propagado mundo afora às custas de trilhões de dólares (Exame, 2025), tornando extremamente complexo, para não dizer impossível, manter a devida distância. Ele quer estar presente em todos os lugares – veja, na Figura 1 e 2, o caso da própria plataforma onde eu escrevo essa dissertação – e, ao que tudo indica, custe o que custar. Vidas, inclusive.

Figura 1 e 2 - Pop-up apresentado de forma recorrente em momentos de breve inatividade no Google Docs



Fonte: Elaboração própria

Apesar desta pesquisa trazer em seu título a entidade da Inteligência Artificial, seu foco está nos modelos de linguagem de grande escala, presentes nos *chatbots* de IA. No entanto, o título também enuncia informação qualificada e, com essa premissa, não poderia deixar de mencionar o fato temporal (e doloroso) a seguir.

Em 2016, o YouTube modificou seu sistema de recomendação, implementando uma IA de aprendizado profundo (Covington; Adams; Sargin, 2016). Utilizando redes neurais, o sistema tenta prever qual vídeo, especificamente, o usuário assistirá a seguir com base em seu histórico e contexto. Dessa forma, o vídeo sugerido e os ranqueados numa busca serão aqueles com o maior potencial de capturar a atenção do usuário pelo maior tempo de exibição possível, modificando a lógica do sistema anterior, que se baseava sumariamente na taxa de cliques (CTR).

A partir deste momento, na busca pela captura máxima da atenção do usuário, o YouTube passou a privilegiar vídeos que mantivessem, em cadeia, os olhares em sua plataforma, contribuindo para a transformação de espectadores casuais em radicais viciados (Fisher, 2023) dentro de um ecossistema que

recompensa o ódio e a desinformação. O algoritmo passou a usar canais com credibilidade para redirecionar os usuários à conspirações e desinformações.

No Brasil, isso se evidencia diretamente com o surto de zika vírus, iniciado em 2015, que se agravou a partir de 2016 diante da onda de desinformação disseminada nas redes, sobretudo na dobradinha YouTube e WhatsApp, gerando medo e desconfiança dos tratamentos oferecidos pelo Sistema Único de Saúde, como retratado no caso a seguir (Fisher, 2023)

> Quando a bebê de Santos foi diagnosticada com microcefalia, três anos antes, as informações eram escassas. Era um vírus novo, afinal. Ela buscou cada arremedo de informação que pôde na internet, inclusive no YouTube. A plataforma repetidamente lhe oferecia vídeos que atribuíam a culpa pelo zika a vacinas contra sarampo vencidas, que o governo teria comprado mais baratas. A culpa era do mercúrio nas agulhas, de um conluio sinistro de estrangeiros que queriam debilitar famílias católicas [...]. Santos contou que havia sido levada aos vídeos pelas recomendações do YouTube e pelo seu mecanismo de busca. Mesmo se buscasse no Google termos como “zika” ou “vacinas zika”, ela ia parar nos vídeos que a plataforma, mais uma vez com seu favorecimento sinérgico aos links mais lucrativos, geralmente colocava no alto dos resultados da busca. Santos sabia que a internet não era de todo confiável. Mas os vídeos a deixaram atônita e em dúvida. Ela havia começado a dar os imunizantes normas à filha. Porém, disse, “depois fiquei com medo de dar as outras vacinas”. Ela mesma parou de aceitá-las.

Infelizmente, é notoriamente sabido que tais efeitos não se restringiram ao surto de zika. Anos depois, o mundo enfrentaria a maior pandemia da história moderna, resultando na morte de milhões de pessoas. Estados Unidos, Brasil e Índia foram os países com o maior número de óbitos e são, ironicamente (?), os maiores mercados do YouTube. Somente na pandemia de covid-19, estima-se que o tráfego global desta plataforma tenha triplicado. Comportamento semelhante fora observado em outras plataformas, como Twitter e Facebook (Fisher, 2023). Esta última teve a oportunidade de reduzir em até 38% a disseminação de desinformação durante a pandemia, mas não o fez. Prejudicaria o tráfego. Leia-se lucro.

E é por isso que eu disse que fazer essa pesquisa foi quase um manifesto. O desenvolvimento da Inteligência Artificial tem sido movido pelos interesses das Big Techs que, em todas as oportunidades, está direta e prioritariamente atrelada ao lucro. Ainda assim, este pesquisador, se valendo de sua humanidade e inteligência orgânica, atravessado pelos anseios que um dia disseram mover o Vale do Silício, ainda acredita que a tecnologia pode contribuir com um mundo melhor. É só por isso que me dediquei a estudar aplicações desta mesma Inteligência Artificial, ainda que

em outras camadas, no anseio que pudesse fazer o oposto do que temos visto ao longo dos últimos anos e contribuir com o ecossistema da informação. Esperançar é imperativo.

1.2 CONHEÇO O MEU LUGAR

Nasci em Goiânia-GO em 30 de novembro de 1996, mas aos 12 anos me vi migrando para Uberlândia-MG com minha mãe e meu irmão, acompanhados por uma tia e dois primos. Viemos tal qual refugiados em busca de dias melhores. Me considero um goianeiro. Ao mesmo tempo que soa como metade goiano, metade mineiro, me parece também nem um, nem outro. Um não lugar. Ocorre que, na verdade, eu queria muito ser mineiro. Talvez pelo fato de sentir que tenha começado a viver aqui, em terras férteis. Minha vivência no outro estado passa mais pela sobrevivência, o que não me agrada tanto. Mas isso é assunto para outra escrita.

Fiz questão de começar essa seção com um marcador temporal que, até pouco tempo atrás, me provocava curiosas indagações: o que estava acontecendo no mundo quando nasci? O que acontece no meu aniversário para além do meu aniversário? Tudo aconteceu e tudo acontece, é óbvio. Mas como é comum enxergamos o mundo a partir do nosso ego, nenhuma resposta me apetecia. Até que, recentemente no processo de fruição deste trabalho, duas constatações começaram a preencher esse vazio.

A primeira é que, em 1996, uma pesquisa realizada com pesquisadores de países ibéricos e latino-americanos apontou para a criação de um novo campo, a Educomunicação, abrindo caminho para integrar a comunicação no espaço educativo (Pinheiro, 2013). E é exatamente isso que me propus a fazer neste trabalho. Não necessariamente a educomunicação enquanto ”metodologia”, ainda mais que não é nem isso que ela se propõe a ser, mas como princípio, como correligionária de seus ideais. A premissa é que cada sujeito, envolto no seu próprio mundo repleto de vivências singulares, possa aprender e consumir informações verificadas. E, para isso, caminhamos para a segunda constatação.

Em 30 de novembro de 2022, a OpenAI lançou a primeira versão do chatbot ChatGPT, a GPT-3.5. De lá pra cá, vem dominando o mercado com sua presença em mais de 70% dele (First Page Sage, 2025). Santaella (2023, p.7) acredita que “em muito pouco tempo, o nome ChatGPT irá funcionar como um marcador temporal

e simbólico das mudanças que foi pioneiro em introduzir". É por isso que, friso, essa escrita se inicia com um marcador temporal. Se essas previsões se cumprirem, poderá ser uma espécie de a.C/d.C.

E é nesse tic-tac que eu me situo. São lugares que me cabem sem precisar me esticar ou me encolher (Sousa, 2024). Empolgam-me e, ao mesmo tempo, confesso, me assustam. É uma luta constante para não ficar paralisado pelo medo. Me empolgam porque vejo um enorme potencial na junção das Inteligências Artificiais Generativas com o respeito aos saberes e locais de fala/vivência de cada um. Essa personalização, com precisão de apuração, parece ser algo bastante promissor, como o apontado ao longo dessa dissertação, capaz de viabilizar, de fato, a acolhida personalizada para cada cidadão dentro das redes, contribuindo para a manutenção, construção e ampliação da sua autonomia. Me assustam porque, se seguir apenas pelo caminho do interesse das Big Techs e do capital, pode se consolidar apenas como mais um instrumento de dominação e massificação. E é justamente pelo contrário, pela emancipação dos povos e de seus saberes que nos aventuramos por aqui.

Relutei muito em escrever este parágrafo, mas não seria honesto se o omitisse. No fundo, essa pesquisa atravessa grandes dores que carrego, como mencionado na seção anterior. A dor de perder entes queridos para a desinformação, paralelo com a convivência com aqueles que têm seus atos influenciados por ela, se transformou numa espécie de questão existencial. E, por isso, essa produção vai muito além de um trabalho acadêmico/científico. É quase como a elaboração do substantivo luto no desejo de que se torne verbo, tal qual Paulo Freire fez com a esperança/esperançar (1994). Se esta escrita superar essas páginas iniciais, terei superado também a angústia paralisadora que sinto toda vez que volto meus pensamentos para a produção deste trabalho[3].

E, parte das motivações que encontro para tentar superar, passa pela crença ingênua de que apenas a desinformação justifica certas ações e omissões. Não é razoável que alguém, sabendo que não procede, haja como se procedesse. Tampouco o contrário. E, nesse ponto, peço licença para emanar essa inocência, mas sinto cada vez mais que a realidade possa apontar para o tremendo oposto,

[3] Como é bom, ao final do processo de escrita dessa dissertação, retornar à este parágrafo que escrevi lá no início. Essa angústia ainda me acompanha, mas já não me assusta e nem me paralisa como antes. Aprendemos a dançar, eu acho.

como defendido por Safatle (2015), argumentando que a razão, na sociedade contemporânea, tem sido cada vez mais sobreposta pelos afetos, uma espécie de instrumentalização dos desejos capaz de suprimir a capacidade crítica, inibindo reflexões baseadas em uma razão emancipadora (Adorno, Horkheimer, 1985).

Apesar disso, sigo imbuído do propósito de, por meio desta pesquisa, superar essa inocência ou, de algum modo, transformá-la numa contribuição na tentativa de fazer valer o direito à informação — qualificada — previsto na Constituição Federal (1988), algo sem o qual coloca em xeque o exercício pleno da cidadania. Se cada pessoa cidadã tiver o direito, a possibilidade de verificar uma informação de forma tão rápida, simples e acessível como lhe é exposto a desinformação, então estaremos mais próximos da solução. O que será feito por cada um a partir de então já extrapola a competência do nosso campo, muito embora ainda seja de profunda relevância e, compreender tais feitos seja tão urgente quanto combater a desinformação.

1.3 QUE A EDUCAÇÃO SEJA O CAMINHO

Resta dizer quem sou até aqui. Sou movido a sonhos. Formei-me jornalista pela UFU em 2021, no olho do furacão da pandemia. Sou o primeiro de minha família a estudar em uma universidade pública, gratuita e de qualidade. Me lembro de uma vez, quando eu tinha 14 anos, em que minha mãe levava a mim e meu irmão para a escola. Neste trajeto, passávamos por um beco que atravessa várias ruas do bairro São Jorge, na periferia de Uberlândia. Logo no início dele, naquele dia, eu disse a ela que queria muito fazer faculdade. Ela olhou para mim e, pensativa, disse: “Filho, fico muito feliz com sua vontade, mas não temos dinheiro para pagar.” De imediato, sem titubear, eu respondi: “Mamãe, a senhora esqueceu? Tem a UFU!”. E ela disse: “Então estuda, meu filho. Estuda!”. Bom, cá estamos novamente. Que coincidência mais gostosa a de hoje, em que o grupo de pesquisa da UFU ao qual faço parte como mestrando se chamar BECO – Hub Innovation. Como tenho aprendido com esses becos por onde a vida me faz passar.

A educação é libertadora e, a informação qualificada, empodera. Todos os caminhos que trilhei até aqui passam por elas. Lá, aos 14 anos, quando disse para minha mãe que existia a UFU, ela ainda era muito distante para nós. Lembro de passar de ônibus em frente à entrada principal do campus Santa Mônica e de ficar

imaginando como seria lá dentro, já que não sabia nem que podia entrar. Como tudo o que dava pra ver lá de fora eram alguns blocos e o amplo gramado da fachada, meu eu adolescente ficava mesmo era se imaginando a rolar por aquela grama[4], rindo até.

Qual foi a minha surpresa quando, pouco tempo depois, me vi contemplado com uma bolsa de Iniciação Científica Júnior do CNPq, proporcionando meu primeiro contato formal com a universidade. E, diga-se de passagem, essa sensação continua até os dias de hoje, quando mais uma vez sou beneficiado desse incentivo à pesquisa dado pelo CNPq, mas dessa vez como bolsista de mestrado, do qual sou extremamente grato. Esse fomento tem sido essencial para a realização deste trabalho.

Rolando de volta, ainda nos idos da minha adolescência, comecei a fazer parte da Missão Escola da Vida, uma instituição sem fins lucrativos que atua complementando a educação de jovens e crianças da periferia de Uberlândia, mais precisamente no grande São Jorge. Lá, são realizadas atividades de reforço escolar, ensino de idiomas, oficinas audiovisuais e de educação midiática, clube de debates, voluntariado, teatro, música e, principalmente, a socialização tão importante para a formação do senso de cidadania.

Aos 16 anos, me tornei o representante de Minas Gerais no Parlamento Juvenil do Mercosul, durante 2012 e 2014. Nesse período, nossa delegação viajou por países do Mercosul debatendo o ensino médio que queríamos, sempre pautada pelos eixos: Direitos Humanos, Jovens e Trabalho, Gênero, Participação Cidadã e Integração Latinoamericana.

Já aos 19 anos, ingresso na assessoria de comunicação no meio político, me rendendo uma significativa experiência e me provocando ao ponto de se tornar tema da minha monografia no trabalho de conclusão de curso da graduação, onde analisei a comunicação entre agentes políticos e o público em período eleitoral e não-eleitoral no Facebook. É exatamente neste momento que começa a surgir o

[4] Essa história de rolar na grama está rendendo. Na graduação, como a minha formatura foi online por causa da pandemia, tive a ideia de realizar um rito simbólico diferente. No dia de pegar o meu diploma, convidei meu irmão, minha mãe e meu padrasto para me acompanharem até a UFU. Após pegá-lo, fomos até o famoso gramado e eu, então com 24 anos, rolei feito criança sem conter a emoção. No link a seguir, é possível ver o registro do momento em meu perfil do Instagram: https://www.instagram.com/reels/CRCFd1LJtRd/. Mas não para por aí. Parte da minha turma de mestrado ficou sabendo e curtiu demais a ideia. Tanto que estamos planejando uma rolagem coletiva neste mesmo gramado assim que todos passarem pela banca de defesa, rsrsrs. Como isso me empolga! Se de fato rolar este rolar, tentarei deixá-lo registrado neste mesmo perfil.

meu interesse pela informação qualificada, pois me via indignado ao perceber o quanto a desinformação cativava e encontrava espaço no cotidiano das pessoas. Ali, diante de toda a crise envolvendo os profissionais jornalistas, crente de que o tempo de ouro já havia passado, vi que nossa profissão nunca antes se fez tão necessária e indispensável. Afinal, em terra de *fakenews*, quem tem informação qualificada é rei. Bom, ao menos é como eu gosto de acreditar.

Por fim, resta comentar sobre a experiência que me fez tirar o foco do mercado e considerar a trajetória acadêmica como uma possibilidade real. Apesar de ter entrado na graduação tão almejada, meus planos consistiam até então em finalizá-la e abraçar a vida profissional lá fora. Já estava trabalhando desde o terceiro período e queria, cada vez mais, atuar no meio. No entanto, em 2019, aos 22 anos, fui selecionado como bolsista para participar da equipe de Documentaristas da *Brazil Conference at Harvard and MIT*, em Cambridge, Massachusetts, nos Estados Unidos. Lá, fiquei deslumbrado com toda a estrutura das universidades e do potencial que as pesquisas científicas podem alcançar, principalmente ao unir saberes com tecnologias e empreendedorismo. Voltei determinado a seguir essa tríade que, ao meu melhor, espero que fique bem materializada nas páginas que se seguem.

—- Novamente, ao retornar por estass seções após o final do processo da dissertação, como me alegra olhar pra trás e ver que valeu a pena e que este trabalho está colhendo frutos. A pesquisa presente nesta dissertação foi vencedora do Prêmio INTERCOM de Comunicação Para Transformação Social, na categoria pesquisa - mestrado da região Sudeste em 2025. E confesso que eu a submeti no congresso de última hora, pois não acreditava nem que o resumo expandido seria aceito. E cá estamos. Preciso lidar melhor com essas questões e amadurecer. Bom, essa será uma tarefa para o doutorado.

2 REFERENCIAL TEÓRICO

Diante dos fatores que colocam em xeque o fazer jornalismo nos últimos anos, é preciso pensar em maneiras de tentar superá-los. Considerando o advento da Inteligência Artificial Generativa, de que maneira seus modelos podem contribuir para a informação qualificada?

Se pesquisar no Google Acadêmico pelo termo “ChatGPT”, apenas de 2022 pra cá, haverá ao menos 40 mil resultados. Se buscar por “*Large Language Models*”, mais de 60 mil. Embora não conclusivos por si só, esses números indicam o quão relevante essa área se mostra, cujo interesse vem crescendo desde o primeiro lançamento da *OpenAI*.

Para compreender a Inteligência Artificial Generativa, é preciso entender sobre o modelo que a permite funcionar, os *Large Language Models*. Os LLMs são algoritmos de inteligência artificial que consistem em modelos de linguagem de larga escala. São treinados com um vasto banco de dados, utilizando técnicas avançadas de aprendizado de máquina e aprendizado profundo, ambas do campo das ciências da computação (IBM, s/d). Esses modelos utilizam redes neurais artificiais, uma arquitetura inspirada numa abstração do funcionamento do cérebro humano. Como diz Santaella (2023), esses algoritmos aprendem a realizar associações estatísticas entre bilhões de palavras e frases para executar diversas tarefas que envolvem linguagem natural.

Historicamente, os modelos de linguagem evoluíram significativamente desde os primeiros sistemas baseados em regras até os modernos modelos de larga escala. Nos anos 1950 e 1960, os primeiros sistemas de processamento de linguagem natural (PLN) usavam regras gramaticais e léxicos predefinidos. Com o advento da aprendizagem de máquina nos anos 1980, esses sistemas começaram a incorporar métodos estatísticos, que permitiam a análise de grandes quantidades de texto para inferir padrões e relações entre palavras (Manning; Schutze, 1999). O desenvolvimento do algoritmo de retropropagação na década de 1980, o *backpropagation*, foi um marco significativo, permitindo o treinamento mais eficiente de redes neurais profundas (Rumelhart; Hinton; Williams, 1986).

A utilização de LLMs cresceu exponencialmente com o desenvolvimento de arquiteturas mais sofisticadas, como os *Transformers*, introduzidos por Vaswani *et al.* (2017). Essa arquitetura revolucionou o campo ao possibilitar o treinamento paralelo

de modelos em grande escala, reduzindo o tempo necessário para o treinamento e aumentando a precisão dos modelos de linguagem. Antes dos *Transformers*, os LLMs “baseavam-se nas redes neurais recorrentes, aquelas que analisam uma sequência de texto e prevêem qual será a próxima palavra” (Santaella, 2023, p. 31). Mas faziam isso analisando uma frase palavra por palavra. Agora, após 2017 com a arquitetura *Transformer*, é possível analisar todas as palavras ao mesmo tempo, podendo examinar grandes volumes de texto simultaneamente.

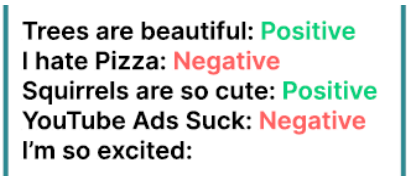
2.1 PROMPTING

De forma geral, *prompt* é uma informação enviada (entrada) a um modelo de IA Generativa que é utilizada como guia para gerar uma outra informação (saída), podendo se consistir em textos, imagens, sons ou outro conteúdo audiovisual (Meskó, 2023; White et al., 2023; Heston and Khun, 2023; Hadi et al., 2023; Brown et al., 2020 apud Schulhoff et al., 2024). Como exemplo, alguns *prompts* podem ser algo como “escreva três parágrafos de e-mail para uma campanha de marketing”; uma fotografia de uma mesa acompanhada do texto “descreva tudo o que há na mesa”; ou, ainda, a gravação de uma reunião seguida da instrução “resuma isso” (Schulhoff et al., 2024, p.5, tradução livre).

Existem diversas ramificações para o estudo dos *prompts*, como o *prompting,* modelos de *prompt,* engenharia de prompt e o *fine-tuning*/ajuste fino. Esse trabalho foca na primeira, o *prompting*, que consiste no processo de prover um *prompt* para o modelo de IA Generativa para que, a partir dele, ela gere uma resposta. O ato de enviar um bloco de texto ou subir uma imagem no modelo de LLM constitui, por exemplo, o *prompting* (Schulhoff et al., 2024).

Dentro de *prompting*, há as técnicas de *prompting*, que orientam como estruturar um ou uma sequência dinâmica de vários *prompts*. Dentre elas, destacam-se duas, a *few-shot* e a *zero-shot*. A primeira faz parte da categoria intitulada ICL (*In-Context Learning*), ou aprendizagem no contexto, que refere-se à capacidade das IA Generativas em aprender habilidades e tarefas quando lhes são fornecidos exemplos e/ou instruções relevantes dentro do *prompt* (Schulhoff et al., 2024). Um exemplo disso seria uma tarefa de classificação de sentimentos em positivo e negativo a partir de um grupo exemplar de frases (Figura 3).

Figura 3 - Design de *few-shot prompt*



Fonte: (Schulhoff et al., 2024)

Os autores problematizam brevemente a expressão “aprender”, uma vez que essas habilidades não necessariamente são novas e podem já terem sido incluídas nos dados de treinamento. No entanto, o foco aqui é a realização da tarefa solicitada, não sua aprendizagem.

A segunda técnica, a *zero-shot*, não pertence à categoria ICL e não necessita de exemplos. Bastam simples direcionamentos, como o *role-prompting*, em que se consiste em atribuir uma função específica a IA Generativa, como agir como uma determinada persona que detém uma habilidade específica, como agente de viagens, um cantor, médico, etc. Para esse trabalho, a função atribuída foi de “avaliador de similaridade semântica”.

Em ambas as técnicas, ainda é possível adicionar elementos que aprimoram o desempenho, como as cadeias de pensamento (Chain-of-Thought), ajudando os LLMs a articular seu raciocínio durante a solução de um problema (Zhang et al., 2023c apud Schulhoff et al., 2024). Isso inclui uma espécie de “passo a passo”, guiando o LLM em como deve agir para realizar aquela tarefa.

2.2 INFORMAÇÃO QUALIFICADA

Por informação qualificada, compreende-se a capacidade do LLM em retornar uma resposta baseada em fatos. Neste trabalho, ancora-se essa compreensão no reconhecimento da atividade de verificação de fatos, realizada pelas agências de checagem, como parte de um ecossistema de resposta à desordem informacional, definida em três tipos por Wardle e Derakhshan (2023), sendo eles: desinformação, informação falsa e informação maliciosa (Quadro 1).

Quadro 1 - Definições conceituais apresentadas no livro Desordem Informacional

Tipo	**Definição**	**Exemplo prático**
Desinformação	Informação falsa e deliberadamente criada para causar danos a uma pessoa, grupo social, organização ou país.	Alegação de fraudes em urnas eletrônicas para desacreditar o sistema eleitoral.
Informação falsa	Informação falsa mas que não foi criada com a intenção de causar danos.	Boato sobre a morte de um artista.
Informação maliciosa	informação baseada na realidade, mas usada para causar danos a uma pessoa, organização ou país.	Prints verdadeiros de conversas privadas, fora de contexto, para prejudicar a imagem de alguém.

Fonte: Wardle e Derakhshan (2023), com exemplos de autoria própria.

Segundo Wardle e Derakhshan (2023), a verificação de fatos sempre foi fundamental para o jornalismo de qualidade, mas agora se faz mais importante do que nunca, uma vez que as técnicas de disseminação da desinformação estão cada vez mais sofisticadas. Nesse sentido, defendem que os jornalistas devem aprender a desmascarar fontes por trás de um conteúdo em tempo real, o que exigirá que estes profissionais também tenham experiência em programação de computadores. É nesse sentido que o Jornalismo de *Prompt* também se projeta.

2.3 CAMINHOS DIALÓGICOS POSSÍVEIS

Os escritos dessa seção servem como sinopses de um por vir. As discussões dedilhadas a seguir ilustram um caminho possível a ser percorrido no diálogo entre a teoria e os resultados desta pesquisa. Algumas já são perceptíveis ao longo do conteúdo das demais seções, mas outras ainda carecem de aprofundamento e estão aqui como indicativo.

Inspirada pelas perspectivas humanizadas de Kai-Fu Lee (2019), ao propor um projeto de coexistência entre seres humanos e IA, esta pesquisa se ampara nessa sinergia que, segundo o autor, é indispensável para atravessar o que chama de Revolução da IA, alegando que esta será na mesma escala que a Revolução Industrial, porém maior e definitivamente mais rápida. Diante disso, uma vez que essa projeção tenda a se cumprir, é preciso buscar maneiras de assegurar que os

trabalhadores, neste caso tanto os jornalistas quanto a maior parte do público, tenham condições de usufruir dessas tecnologias emergentes, de forma a potencializar suas existências e não ameaçá-las.

No desdobramento dessa questão, Dora Kaufman (2022, p. 66) reforça o papel social do jornalismo: “notícia de qualidade é do interesse de toda a sociedade”. E, para isso, defende a necessidade de se estabelecer diretrizes e regulamentações para que a oportunidade não se torne uma ameaça. Nesse sentido, faz coro ao OberCom (2021), o qual indica também questões urgentes a serem enfrentadas, como no caso específico dos jornalistas, a urgência para que estes profissionais se qualifiquem e requalifiquem para novas habilidades, como a alfabetização em dados.

Nesse contexto das novas habilidades, Francesco Marconi (2020) acredita que a IA tem o potencial de expandir a capacidade das redações em produzir diferentes produtos jornalísticos, incrementando o volume e a qualidade de histórias publicadas. A utilização de ferramentas de IA permite aos jornalistas aprimorar suas atividades de reportagem, pesquisa, redação e edição, resultando em um tipo de jornalismo conhecido como *augmented journalism*, jornalismo aumentado em tradução livre. No entanto, Marconi também alerta para preocupações importantes, como o viés de máquina, os riscos de geração algorítmica não verificada, mudanças nos fluxos de trabalho, responsabilidade legal e a lacuna crescente nas habilidades necessárias para gerenciar essa nova área de especialização. À primeira vista, essa definição de Maconi de jornalismo aumentado chama atenção como ponto antecessor da proposta que esta pesquisa se propõe a fundamentar, o jornalismo de prompt, mas que ainda carece de estudos mais aprofundados para tal compreensão.

Nessa direção, ainda é possível contar com Lúcia Santaella (2023), que contribui na superação do debate sobre o potencial da inteligência artificial, quando até mesmo seu conceito de inteligência é questionado. Esses questionamentos ocorrem, segundo Santaella (2023), pelo fato da IA colocar em foco habilidades que até então eram tidas como exclusivamente humanas, como a autoria, a criatividade e a autonomia. Arielle e Manovich (2022) dizem que isso acontece porque há uma pretensa soberania humana dentro de uma perspectiva antropocêntrica de agenciamento e criatividade. E, por assim dizer, o mesmo ocorre por parte daqueles que negam a inteligência às IAG.

Por fim, Camila Leporace (2024) traz uma sóbria e atualizada problematização em torno da inteligência artificial e da cognição humana. A autora chama a atenção para a proposta enativista em torno da mente humana, de modo a projetá-la para além do cérebro, compreendendo também o corpo e o ambiente em que os sujeitos estão inseridos. Essa concepção é essencial para a compreensão do que difere o ser humano da inteligência artificial e quais os caminhos para que essa interação não resulte na transferência da autonomia humana para a máquina. Esse debate gira em torno da algoritmosfera, termo cunhado por Leporace (2024) para se referir à presença humana no espaço digital, e se faz importante em um cenário em que se almeja construir abordagens críticas e emancipatórias para o Jornalismo de *Prompt*.

2.4 A INTELIGÊNCIA ARTIFICIAL GENERATIVA É, DE FATO, INTELIGENTE?

Essa pergunta geralmente vem acompanhada da afirmação de que a IA não pensa, não sabe o que está produzindo. E isso é verdade. De modo simplório, o Chat “apenas” supõe quais palavras devem vir a seguir, com base em análises estatísticas de probabilidades a depender de cada contexto. Contexto esse que é construído e aprimorado à medida que novas entradas vão sendo realizadas. Mas sim, não pensa e não sabe, tal qual um *homo sapiens*.

Nesse sentido, Santaella (2023) afirma que, apesar do Chat não ter compreensão sobre o sentido do que escreve, isso não deve ser considerado como um demérito que o inutiliza, tal qual fazem aqueles que se excitam ao desmerecer seus préstimos. A pesquisadora enfatiza ainda que, segundo diversos pensadores como Jacques Lacan (1985, apud Santaella, 2023, p.38), “o significado de algo não é um dado que salta imediatamente de nossas cabeças, mas escorrega continuamente pelos circuitos dos jogos da linguagem e de seus contextos”. Assim, a compreensão sobre um fato não é algo que está dado por natureza e muitos humanos carecem, eles próprios, de entender o sentido do que dizem.

> Deveria ser um truísmo afirmar que, para o humano, entender o sentido do que diz e, mais do que isso, responsabilizar-se por isso, é condição ética e mesmo moral imprescindíveis cuja falta, em maior ou menor medida, prejudica, contamina e envenena a vida social, especialmente nesta era de redes sociais estufadas de banalidades, exibicionismos, opiniões

> levianas e muitas vezes carregadas de ódio pelo diferente (Santaella, 2023, p. 38).

Esses questionamentos ocorrem, segundo Santaella (2023), pelo fato da IA colocar em foco habilidades que até então eram tidas como exclusivamente humanas, como a autoria, a criatividade e a autonomia. Arielle e Manovich (2022) dizem que isso acontece porque há uma pretensa soberania humana dentro de uma perspectiva antropocêntrica de agenciamento e criatividade. E, por assim dizer, o mesmo ocorre por parte daqueles que negam a inteligência às IAG.

Diante disso, Santaella (2023) defende que é fundamental compreender o que é inteligência, encontrando definições que superem a velha dicotomia ocidental, uma visão geralmente dualista e reducionista que tende a separar, de forma rígida, algumas questões, como: mente e corpo, razão e emoção ou, nesse caso, humano e máquina. No contexto da IA, essa visão dicotômica parece ser inadequada, pois a inteligência artificial opera de maneiras que não se encaixam perfeitamente nas caixas delimitadoras. A IA não compreende o que escreve, mas pode realizar tarefas complexas, aprender com grandes volumes de dados e até imitar certos aspectos da inteligência humana, como já mencionado.

Em busca dessas definições mais adequadas, a pesquisadora, munindo-se da teoria da causalidade do filósofo C.S Peirce, defende que sim, a Inteligência Artificial é inteligente, ainda que seja uma inteligência diferente da humana. Para o filósofo (apud Santaella, 2023), só existem duas ações básicas universais e ambas são interligadas: ações inteligentes e ações eficientes. Com base nessas definições, da qual fica o convite para buscar a compreensão e aprofundamento sobre elas em local mais apropriado[5], Santaella (2023, p. 104) alega que, no caso do ChatGPT, se este não tivesse inteligência, “ele não passaria de um executor despropositado, mais desgovernado do que uma barata sob ação de Rodox. Esse está muitíssimo longe de ser o caso”. Para ela, o caso não é outro a não ser reconhecer a inteligência existente nas operações estatísticas de que a IA se alimenta.

[5] Para aprofundar nessa discussão, veja a obra da autora publicada especificamente para tratar deste questionamento, “A Inteligência Artificial é Inteligente?”, pela editora Almedina, em março de 2023.

2.5 TRABALHOS RELACIONADOS

Durante a revisão da literatura, foram identificados alguns trabalhos que investigam a aplicação de Modelos de Linguagem de Larga Escala (LLMs) em tarefas de checagem de fatos. Esses estudos apresentam abordagens metodológicas distintas, com foco na avaliação da acurácia, na utilização de evidências contextuais e na capacidade dos modelos de justificar suas respostas. A seguir, são apresentados três desses trabalhos, que contribuem para o entendimento das potencialidades e limitações dos LLMs no enfrentamento da desordem informacional e na automatização de processos de checagem de conteúdo.

Trabalho 1: *News Verifiers Showdown: A Comparative Performance Evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in News Fact-Checking* (Caramancion, 2023). Este estudo comparativo avaliou a capacidade de quatro grandes modelos de linguagem — ChatGPT 3.5, ChatGPT 4.0, Bard (Google) e Bing AI (Microsoft) — para classificar a veracidade de notícias. A metodologia foi baseada em testes tipo “caixa-preta”, nos quais 100 manchetes de notícias previamente verificadas por agências independentes (como PolitiFact e Snopes) foram apresentadas aos modelos, que então precisavam classificá-las como "Verdadeiras", "Falsas" ou "Parcialmente Verdadeiras/Falsas". As respostas dos modelos foram comparadas às classificações oficiais. O ChatGPT 4.0 apresentou o melhor desempenho com 71% de acurácia, seguido pelos demais com médias em torno de 65%. Embora mostrem potencial, os modelos ainda não superam os checadores humanos em termos de interpretação contextual e nuances linguísticas.

Trabalho 2: *The Perils & Promises of Fact-checking with Large Language Models* (Quelle & Bovet, 2023). Neste artigo, os autores investigaram a efetividade do GPT-3.5 e GPT-4 em tarefas de checagem de fatos, com ou sem acesso a informações contextuais adicionais. A metodologia combinou agentes baseados no framework ReAct, capazes de realizar buscas no Google e justificar seus veredictos com base em fontes contextuais. O estudo aplicou dois experimentos: o primeiro utilizou 3.000 declarações checadas do PolitiFact, e o segundo um banco multilingue da Data Commons com milhares de declarações (amostra reduzida para até 500 por idioma). Os modelos foram testados tanto com os textos originais quanto com traduções para o inglês. Os resultados mostraram que o GPT-4 superou o GPT-3.5

em todos os cenários, especialmente quando recebeu contexto, e que traduzir textos para o inglês melhorou significativamente a acurácia dos modelos para idiomas com menos recursos.

Trabalho 3: *Are Large Language Models Good Fact Checkers?* (Han Cao et al., 2023). Neste estudo preliminar, os pesquisadores examinaram o desempenho de diversos LLMs (incluindo GPT-3.5, LLaMa2, GLM-6b e GLM-130b) ao longo de todo as etapas de checagem de fatos, composto por quatro subtarefas: detecção de checabilidade, recuperação de evidências, verificação factual e geração de explicação. Foram utilizados três conjuntos de dados: CheckThat! (195 tweets), AVeriTeC (inglês) e CHEF (chinês). A avaliação envolveu diferentes técnicas de *prompting*, como "*Chain-of-Thought*" e "*task definition*", além de experimentos em configurações *0-shot*, *1-shot* e *3-shot.* Os LLMs mostraram bom desempenho em tarefas isoladas, mas ainda enfrentam dificuldades em cobrir todas as etapas de checagem, especialmente em línguas distintas do inglês. Os resultados indicam que o uso de *prompts* aprimorados e exemplos melhora o desempenho, mas os modelos continuam vulneráveis à "alucinações" e limitações linguísticas.

Embora dialogue com essas pesquisas que exploram o uso de LLMs em tarefas de verificação de fatos, este trabalho propõe um caminho distinto. Ao invés de focar apenas na classificação binária de afirmações como verdadeiras ou falsas, optou-se aqui por uma abordagem baseada na análise da similaridade semântica entre os conteúdos gerados por um modelo de linguagem e os textos produzidos por uma agência de checagem. Essa escolha permite observar não apenas se o modelo acerta, mas como ele se aproxima, em conteúdo e intenção, de uma resposta qualificada.

2.6 MÉTODOS DE AGRUPAMENTO

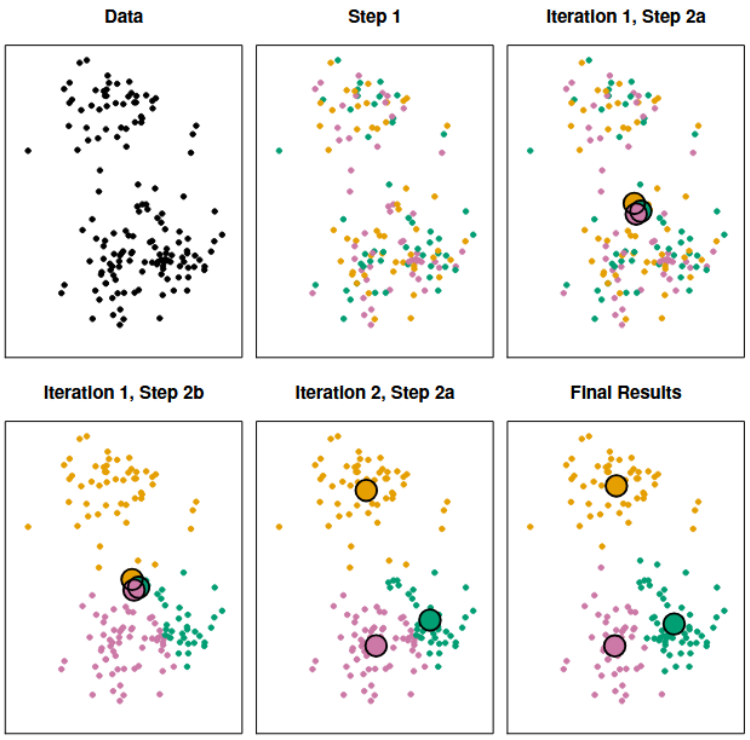
Nesta etapa, esta pesquisa busca fundamentação teórico-metodológica em métodos de agrupamento não supervisionados, como o K-Means e o Agrupamento Hierárquico, que possibilitam uma organização mais adequada do conjunto de dados, especialmente por não haver uma variável resposta. Apesar do conteúdo das agências de checagem terem essa premissa em relação ao extraído do LLM, não há um selo em cada um que permita uma classificação supervisionada. Portanto, se

trata de um problema não-supervisionado, ainda mais diante de uma análise de similaridade exploratória.

Define-se como clusterização as técnicas utilizadas para encontrar subgrupos em um conjunto de dados, de modo a particioná-lo em grupos distintos no intuito de que as observações dentro de um mesmo grupo sejam o mais semelhante possível, à medida que as observações em grupos diferentes sejam, por consequência, o mais diferente possível em relação às dos demais (James et al., 2023).

Em relação ao K-Means, segundo James et al. (2023), essa abordagem realiza o agrupamento a partir de uma pré-definição obrigatória da quantidade de grupos desejada. A partir disso, ele distribui os pontos (no âmbito dessa pesquisa, as medidas de similaridade de cada modelo de *prompt* e de LLM atribuídas a cada observação, ou seja, a cada par de conteúdo da agência/LLM) conforme a menor distância entre cada um deles e os centróides, ou seja, os pontos que representam a média dos dados em cada grupo formado pelo algoritmo, que são recalculados iterativamente. Esse processo se repete até que os agrupamentos se estabilizem, formando grupos baseados na proximidade entre todo o conjunto de dados (Figura 4).

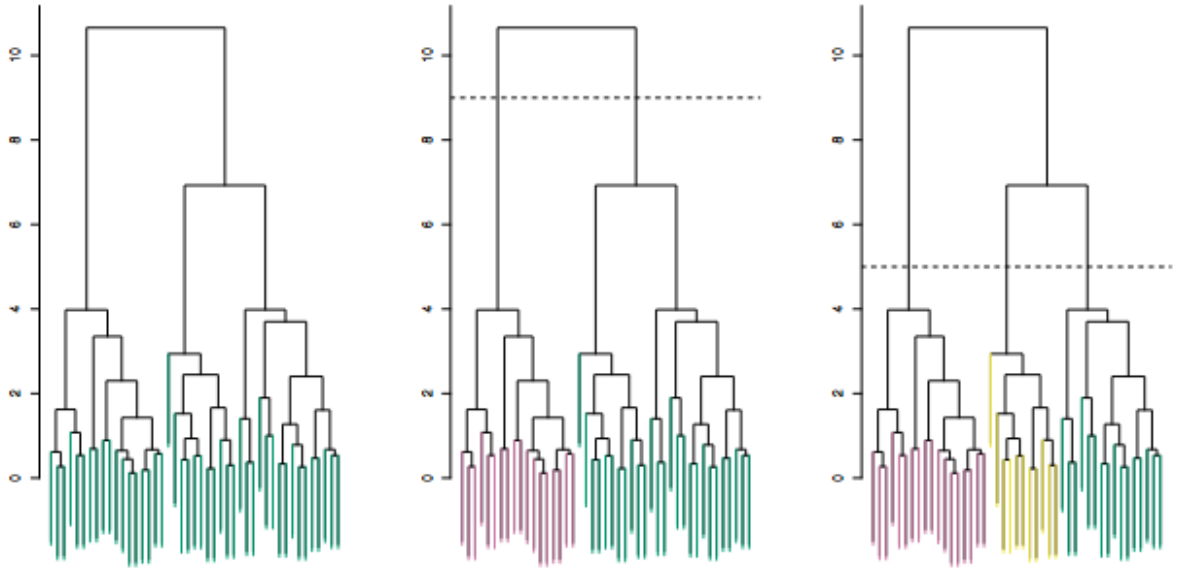
Figura 4 - O progresso do algoritmo K-Means com 3 agrupamentos



Fonte: James et. al., 2023, p. 518

O Aglomerado Hierárquico, por sua vez, apresenta uma vantagem em relação ao K-Means, pois não exige a definição prévia da quantidade de grupos, o que torna o processo de classificação menos arbitrário. Além disso, esse método permite a geração de uma representação visual em forma de árvore — o dendrograma — que facilita a compreensão da estrutura dos agrupamentos. A definição final dos grupos depende da análise do observador, que deve escolher o ponto de corte nos ramos da árvore com base no conhecimento do conjunto de dados (Figura 5).

Figura 5 - Representação de dendrogramas em três agrupamentos diferentes, a depender do corte



Fonte: James et. al., 2023

Esse corte no Aglomerado Hierárquico está para a definição da quantidade de grupos tal qual a definição de clusters iniciais está para o K-Means (James et al., 2023). No âmbito deste trabalho, coincidentemente, a quantidade de *clusters* considerada adequada em ambas as abordagens, ao serem aplicadas ao conjunto de dados, também totalizou três. Fez-se questão de utilizar as duas no intuito de buscar maior amparo, uma vez que se trata de uma pesquisa experimental. No capítulo de análise, essa questão é aprofundada, bem como é apresentado um dendograma específico referente à amostra da pesquisa desenvolvida.

Ainda é prudente ressaltar a necessidade de cautela com essas abordagens de clusterização, uma vez que elas podem não lidar bem com *outliers*, ou seja, com entradas que destoam significativamente do restante do conjunto de dados. James

et al. (2023) aponta que, por forçarem cada observação a pertencer a um grupo, tais métodos podem gerar distorções na formação dos *clusters*, sobretudo quando há observações que não se ajustam bem a nenhum deles. No contexto desta pesquisa, entretanto, observou-se que o padrão dos *outliers* identificados se concentraram em um único *cluster* — justamente em razão de seu comportamento discrepante em relação às demais observações — aspecto que será explorado com mais profundidade no capítulo de análise.

Ainda como bem alertado por James et al. (2023), os resultados da clusterização não devem ser tomados como verdade absoluta sobre um conjunto de dados. Devem, sim, servir como a parte inicial de uma investigação científica a ser verificada como uma hipótese válida posteriormente. E, nesse sentido, pretende-se entregar essa validação no capítulo de análise, apresentando os respectivos testes de hipóteses.

Na seção 5.3 Análise Exploratória dos Agrupamentos, é possível verificar exemplos reais da aplicação desses métodos de agrupamento com base no conjunto de dados da pesquisa realizada.

3. A PULGA ATRÁS DA ORELHA

Após compreender o funcionamento das IAG, é natural questionar como esses recursos podem ser utilizados. Ao se deparar com a possibilidade de utilizar IAs Generativas no cenário de verificação de informação, como o ChatGPT, surgem questões essenciais sobre a confiabilidade das informações, o dilema da autoria, a capacidade crítica do usuário e a forma como essas ferramentas podem complementar ou transformar os métodos tradicionais de produzir e consumir produtos informativos.

A priori, o principal receio que surge do ponto de vista tecnológico sobre a aptidão dos LLMs por trás das IAG, está relacionado com as "alucinações" que esses modelos geram, criando informações equivocadas. Esses equívocos são uma tendência do Chat “a preencher lacunas, inventando coisas, ou seja, ‘alucinando’ ao gerar citações falsas e informações imprecisas” (Santaella, 2023, p. 70), como observado no Quadro 2.

Quadro 2 - Exemplo de alucinação do modelo GPT-4o sobre definição

Entrada	Saída
Apresente a definição de LLMs de cada um desses autores, referenciado no padrão da ABNT.	Definição: Manning descreve LLMs como “modelos de aprendizado profundo que são treinados em grandes quantidades de texto para prever a probabilidade de uma sequência de palavras. Esses modelos utilizam técnicas de aprendizado não supervisionado para capturar a estrutura e semântica da linguagem” Referência ABNT: MANNING, Christopher D. Foundations of Statistical Natural Language Processing. MIT Press, 1999.
Em que página está essa definição de Manning?	A definição específica de Large Language Models (LLMs) não está explicitamente presente em “Foundations of Statistical Natural Language Processing” de Christopher D. Manning, já que o livro foi publicado em 1999 e os LLMs, como os conhecemos hoje, são desenvolvimentos mais recentes no campo de processamento de linguagem natural e aprendizado profundo.

Fonte: Elaboração própria

Essas "alucinações" colocam em xeque o potencial dos LLMs em verificar informações, uma vez que o menor deslize pode comprometer a veracidade do

conteúdo (Quadro 2). Mas é preciso pontuar que tais equívocos não são inerentes às IAG e servem para reforçar a necessidade da atuação especializada do humano nesse processo. No entanto, é necessário frisar que a utilização do Chat por si só, segundo Santaella (2023), é inteiramente inadequada para tarefas em que a precisão é extremamente necessária, como na produção de conhecimento científico.

Essa preocupação fica ainda mais evidente quando Santaella (2023, p. 87) alega que os LLMs “não têm a menor ideia se algo é correto ou falso e apresentam com confiança discursiva informações como verdadeiras, mesmo quando não são”. E ainda chama a atenção para o fato das IAG terem o potencial de contribuir ainda mais para a distorção das informações que consumimos.

Por isso que esta pesquisa, ao interagir com os LLMs visando a extração de conteúdo para a criação do conjunto de dados, conforme mencionado no capítulo metodológico, tem como foco testar formas de minimizar as margens para essas "alucinações", superando essa inadequação mencionada por Santaella (2023) em atividades onde a precisão é requisitada, como é o caso da verificação de informações, a fim de estabelecer abordagens adequadas para esse processo.

Essa não é uma tarefa fácil, sobretudo pela própria forma como os LLMs são treinados para realizar tais funções. Em suma, esses modelos utilizam imensas quantidades de informações textuais disponíveis na internet. A Web pública tem bilhões de páginas que foram escritas por seres humanos. Cerca de 5 milhões de livros digitalizados já foram disponibilizados, o que daria mais de 100 bilhões de palavras. Além disso, ainda há os textos que podem ser extraídos de conteúdo audiovisual. Para se ter uma ideia, cerca de 720 mil horas de vídeo são subidas no YouTube diariamente por todo o mundo (Santaella, 2023).

É com base nesses conteúdos que o LLM do ChatGPT tem sido treinado. E, por apenas imitar o comportamento linguageiro humano, está sujeito a todos os equívocos, preconceitos e vieses previamente existentes nesses dados, bem como o surgimento de novos outros, a partir das ligações estatísticas que realiza no processamento. Isso pode levar ao questionamento se os dados deveriam passar por um filtro antes de serem inseridos no treinamento dos algoritmos. No entanto, Santaella alerta que também pode residir perigos nessa prática, como ao retirar conteúdos sexistas e racistas, por exemplo, pode levar o modelo a, diante de um questionamento sobre tais pontos, negar a existência destes. Diante disso, é preciso cautela ao utilizar um sistema que fala como gente, mas não o é, tampouco possui

autoentendimento e ética, “justamente esses dois constituintes mais complexos da constituição humana” (Santaella, 2023, p. 44).

3.1 NO HORIZONTE, A BLOCKCHAIN HUMANIZADA

A criptomoeda Bitcoin ficou mundialmente famosa há alguns anos. Em seu primeiro registro, em 2010, ela não valia nem 1 dólar (EDWARDS, 2021). Hoje, cerca de uma década depois, a moeda virtual ultrapassou a casa dos US $55.000,00. Sim, há muitas variáveis ao redor da estruturação de sua valorização. No entanto, o que este projeto propõe ao citá-la como exemplo é observar um dos mecanismos que tornaram as transações com Bitcoins algo significativamente seguro, a *blockchain.*

Traduzindo do inglês, *blockchain* é basicamente uma corrente de blocos cujo propósito consiste em garantir a segurança de transações online, por meio de

> uma internet anônima, descentralizada e com garantia de proteção à privacidade. [...] Na prática, então, **a blockchain é como um livro contábil**, em que são cadastrados vários tipos de transações. Com a descentralização, as páginas desse livro contábil são armazenadas em bibliotecas localizadas em lugares distintos. Ou seja, **trata-se de um registro coletivo, em que os dados são distribuídos entre todos os computadores ligados à rede**. (ANDRION, 2019)

Dito isso, assim como no sistema de Bitcoins que utiliza o método *blockchain* para garantir a veracidade e segurança de suas transações, seria possível, talvez, pensar em um método semelhante, porém essencialmente humanizado (ainda que altamente corroborado por ferramentas da programação computacional, como a ciência de dados), que seja autossuficientemente capaz de assegurar um sistema de informações, predominantemente jornalísticas, com um grau apurado de confiabilidade?

Os meios de comunicação digital, bem como seus leitores, poderiam ter como aliados, talvez, uma tecnologia capaz de formar uma rede composta essencialmente por jornalistas, suas fontes, personalidades/instituições que, além de publicarem conteúdos e/ou deles fazerem parte, também atuassem ativamente como checadores de si mesmos. Basicamente, seria algo como o exemplo a seguir. O jornalista A publica seu conteúdo cuja veracidade será checada por jornalistas B, C, D e assim sucessivamente por inúmeros outros. As personalidades/instituições

citadas no conteúdo do jornalista A também poderão contribuir com o processo de checagem, validando ou não tal afirmação.

É sabido que envolver personalidades/instituições no processo de checagem poderá esbarrar em um conflito de interesses, em que estas, porventura, neguem informações afirmadas por jornalistas, ainda que procedentes, caso sejam prejudiciais aos envolvidos. Portanto, assim como a República, esse modelo proposto de *blockchain* humanizada deverá contar com uma espécie de freios e contrapesos[6].

Aplicando a essência da referida teoria à prática desta proposta de produto, obtém-se um resultado promissor. O modelo de *blockchain* humanizada poderia contar, superada a fase de prototipação, com recursos oriundos da programação. Um software capaz de mensurar tendências e comportamentos prévios que contribuam nas avaliações, tendo como uma das possibilidades o seguinte cenário: caso um jornalista publique um conteúdo sobre uma personalidade/instituição e essa, por sua vez, indique que tal afirmação é inverídica, outros jornalistas serão convidados a averiguar. Se comprovada a veracidade, a personalidade/instituição em questão terá armazenado um saldo positivo (pontuação) que, após reiteradas vezes, será tida como constantemente confiável enquanto manter esse comportamento. Já o jornalista responsável pela publicação inverídica será responsabilizado, gerando também uma pontuação com saldo negativo que, após reiteradas vezes, poderá ser tido como constantemente inconfiável enquanto manter este comportamento. O cenário inverso deste exemplo também é esperado, bem como outras variáveis e níveis intermediários.

É necessário avançar com desdobramentos futuros desta pesquisa para, com base nela, refletir se a idealização da *blockchain* humanizada pode, em alguma medida, se fazer como alternativa eficaz às limitações atuais dos LLMs, como as "alucinações", mas sobretudo como parte do fortalecimento do ecossistema informativo.

Como alternativa para minimizar as "alucinações" e tornar LLMs mais adequados para tarefas que requerem precisão, como no caso da checagem de informações, se sinaliza a proposta futura de construir um modelo treinado em um

---

[6] A origem da teoria de freios e contrapesos remonta de muitos séculos atrás, perpassando desde Aristóteles até Montesquieu (Maldonado, 2003), cujo principal argumento se resume ao Poder ser gerido pelo próprio Poder, porém de forma descentralizada. Há autonomia para que cada parte exerça suas funções, mas, para combater/prevenir excessos, devem balancear-se umas às outras.

banco de dados, descentralizado, apenas com informações já previamente verificadas, alimentadas constantemente de forma colaborativa por agentes comprometidos com o ecossistema da informação. Esta proposta fica como indicativo para desdobramentos futuros a partir deste trabalho, uma vez que a solução limitada ao desenvolvido neste trabalho indica uma abordagem, inicialmente, amparada no tratamento de dados semânticos dentro da aplicação da Promptação, apresentada no Capítulo 8.

4. ASPECTOS METODOLÓGICOS

Essa pesquisa tem como objeto a aplicação de tecnologias emergentes, como a inteligência artificial generativa (IAG), especificamente uma parte de seu algoritmo, os modelos de linguagem de larga escala (*Large Language Models*, LLMs), no combate ao ecossistema de desinformação em redes. Parte-se ainda da hipótese de que modelos de Inteligência Artificial Generativa podem contribuir para **discernir** informações em rede. Há um enorme potencial batendo à porta, uma vez que esses modelos, presentes no ChatGPT, por exemplo, se popularizaram rapidamente.

Em 2022, a OpenAI lançou a primeira versão do chatbot ChatGPT, a GPT-3.5. Menos de uma semana do lançamento, a ferramenta já contava com 1 milhão de usuários. Com três meses, já havia ultrapassado a marca de 100 milhões (Haas, 2023). Em março de 2024, o Brasil se tornou o 4º país no ranking dos que mais usam o ChatGPT no mundo (Tunholi, 2024).

Diante de tamanha capilaridade, tendo como base o guarda-chuva do método hipotético-dedutivo, essa pesquisa se projeta com os objetivos de **criar** um conjunto de dados com informações extraídas de agências de checagem, bem como de chatbots alimentados por LLMs, de modo a **medir**, precisamente, a aptidão destes modelos para checar informações. Tal aferição será feita por intermédio da **comparação** de seu desempenho com as produções jornalísticas das agências, tendo como resultado almejado a **indicação** de técnicas de prompting utilizadas e suas adequações na interação com os modelos. A partir deste ponto, almeja-se ser possível avaliar a confiabilidade e a capacidade dos LLMs como ferramenta de verificação, bem como de assistir o jornalista na produção de conteúdo qualificado, instaurando aí a **fundamentação** do Jornalismo de *Prompt*.

Este trabalho está dividido em quatro etapas. As três primeiras dizem respeito ao desenvolvimento da pesquisa, ou seja, criar, medir, comparar e indicar. Já para a quarta e última etapa, que se consiste na fundamentação do Jornalismo de *Prompt*, cujo ensaio se encontra no Capítulo 6 e se ramifica para o Capítulo 7, se valeu de toda a produção e os resultados extraídos das etapas anteriores, bem como da leitura de trabalhos correlatos. Veja a seguir, no Quadro 3, o detalhamento dos procedimentos metodológicos aplicados em cada um.

Quadro 3 - Matriz do Trajeto de Pesquisa

Etapa	Descrição	Objetivo relacionado	Técnica
1. Coleta de Dados	Extração do conteúdo da agência de checagem e do chatbot	**Criar** um conjunto de dados	Web scraping e prompting estruturado
2. Medida de Similaridade Semântica	Calcular o grau de semelhança entre os conteúdos da agência e do chatbot	**Medir** a aptidão dos modelos	Técnicas de *prompting* executadas por LLMs via API com Python
3. Avaliação Comparativa	Analisar o desempenho do chatbot em relação à agência	**Comparar** o desempenho dos pares de entrada (agência/chatbot) e **indicar** das melhores técnicas de *prompting*	Análise qualitativa e quantitativa
4. Fundamentação do Jornalismo de Prompt	Refletir sobre os achados e propor conceituação, usos e limitações	**Fundamentar** o conceito	Revisão teórica, articulação com resultados obtidos.

Fonte: Elaboração própria

A **criação** do conjunto de dados (*dataset*) se deu por meio de duas etapas. A primeira consistiu em utilizar a técnica de raspagem de dados, ou *web scraping*, sendo essa o processo de extração automatizada de dados de páginas web (Wickham; Çetinkaya-Rundel; Grolemund, 2023), configurando um *script* na linguagem R, rodado no RStudio versão 2024.04.2 — tendo como base o pacote rvest — para navegar em páginas da *web*, identificar e capturar os dados desejados, como os textos das notícias das agências de checagem. Vale destacar que, como não se tratam de dados sensíveis, descartou-se a necessidade de anonimizá-los, como alerta os autores supracitados. Essa automatização permitiu extrair o título, o primeiro parágrafo (lead) e a data de publicação de 1.200 notícias, totalizando as 100 páginas de conteúdo da agência de checagem Aos Fatos, compreendendo um período de 18 meses, de 11/07/2023 até 07/03/2025. Optou-se por extrair apenas esses três dados de cada notícia por considerá-los suficientes para a análise

proposta, uma vez que os títulos foram escolhidos para servir como base das perguntas geradas para interagir com o LLM e, o primeiro parágrafo, por sua vez, carrega o lead da notícia, o que jornalisticamente procura responder, de forma resumida, os principais aspectos da notícia, como: o quê, quem, quando, onde, como e por quê. A data é um marcador temporal, o que, para estudos posteriores, pode indicar o desempenho de um LLM em checar conteúdos ao passar do tempo.

Vale ressaltar que os assuntos destas entradas são referentes às escolhas da linha editorial do veículo, compreendendo todos os conteúdos checados por ele durante esse recorte temporal. Desta forma, a escolha por estas entradas se deu por toda a produção da agência neste período e não por uma temática específica. Destaca-se também que, por se tratar de um veículo de checagem de fatos, a maioria dos seus conteúdos produzidos visam elucidar uma informação distorcida, desmentir uma desinformação ou contextualizar algo. Assim sendo, é correto dizer que estas notícias são, por natureza, focadas em conteúdos de desordem informacional.

Em um primeiro momento, essa constatação poderia levar a crer que o conjunto de dados está enviesado, com a maioria das entradas se tratando de conteúdo falso, enganoso e, por assim ser, pendendo a análise para um lado. No entanto, como a abordagem experimental aqui adotada não leva em consideração a classificação bruta da entrada, mas sim o quão semelhante ela é semanticamente do seu par, ou seja, o quão a construção de sentidos de ambas caminham no mesmo rumo, esse desbalanceamento de nada influencia.

Os critérios utilizados para a escolha da agência de checagem foram: ser brasileira, ser signatária do IFCN (International Fact-Checking Network, 2025), instituição internacional que certifica agências de checagem com melhores práticas e, o último de caráter técnico, que a agência possua um site com páginas estáticas/consistentes.

A segunda etapa do processo de criação do conjunto de dados se dá pela interação direta com um LLMs, sendo escolhido o modelo com base em sua popularização no mercado de Chatbots de IAG. Em março de 2025, o top 3 é formado pelo ChatGPT (GPT-3.5, GPT-4) com 59.7%, o Microsoft Copilot (GPT-4) com 14.4% e o Google Gemini (Gemini) com 13.5% (First Page Sage, 2024). Ao longo dos últimos 12 meses, de fevereiro de 2024 a fevereiro de 2025, em números absolutos, o ChatGPT oscilou entre 76% e 73%, isso porque se consideram as

porcentagem do Microsoft Copilot, uma vez que ele utiliza o mesmo LLM da OpenAi, fruto de uma parceria entre as empresas, alterando apenas os dados do seu ecossistema.

Pelo critério anterior, será utilizado o ChatGPT, diante de sua expressiva preponderância no cenário. Também vale mencionar que o escopo será composto por modelos presentes em Chatbots que podem ser acessados de forma gratuita, parcial ou totalmente, por parte dos usuários. Também devem ter acesso em tempo real à Internet, de modo a possibilitar checagens de conteúdos factuais.

Partindo para a interação com o LLM escolhido, GPT-4o (sucessor do gpt-4), resta mencionar que esta foi feita utilizando como base os dados extraídos da agência de checagem. Estes foram submetidos como prompt, cujas formas foram definidas com base no artigo "*The Prompt Report: A Systematic Survey of Prompting Techniques*" (Schulhoff at al., 2024) de modo a extrair as versões equivalentes deste modelo para cada checagem feita pela agência. Como *input* para interagir com o modelo, foi utilizado a versão modificada do título da matéria da agência de checagem, sendo transformado em pergunta objetiva e afirmativa[7] (Quadro 4).

Quadro 4 - Amostras aleatórias dos títulos modificados utilizados como *inputs* para interação com o LLM

Título da notícia da agência	**Título modificado para pergunta | Input**
Foto da Parada do Orgulho LGBT de 2017 circula como se fosse de ato pró-Marçal	A foto da Parada do Orgulho LGBT de 2017 é de um ato pró-Marçal?
Ministério da Saúde não autorizou 'mudança de sexo' a partir dos 14 anos	O Ministério da Saúde autorizou 'mudança de sexo' a partir dos 14 anos?
Golpe inventa falso leilão de iPhones pelos Correios para roubar dinheiro e dados de usuários	Os Correios estão realmente realizando um leilão de iPhones que está sendo utilizado para roubar dinheiro e dados dos usuários?
É falso que Lula anunciou prisão exclusiva para pessoas de direita	Lula anunciou prisão exclusiva para pessoas de direita?

Fonte: Elaboração própria

[7] *Cf.* modelo Apêndice A.

Com os *inputs* criados, parte-se para a interação manual com o LLM, realizada diretamente na interface do ChatGPT, uma vez que a API disponibilizada durante a realização da interação não era habilitada para *web search*, elemento fundamental para esse processo. Ao final, essa etapa consistiu na criação de 1.200 novas entradas, sendo a versão equivalente do LMM relacionada às produzidas pela agência. Ao final, a soma dessas interações com o conteúdo extraído da agência resultará na criação do conjunto de dados, com um escopo de cerca de 2.400 entradas.

Com o conjunto de dados criado[8] (Quadro 5), partiu-se para a **medição** da aptidão dos resultados gerados pelos LLMs. Como ponto de partida para medir essa aptidão, foi considerado o quão próximo o conteúdo gerado pelos LLMs é do conteúdo das agências de checagem. Nesse primeiro momento, as checagens das agências servem como *target* para validar, ou não, o resultado dos prompts. Para isso, adentra-se no cenário de processamento automático de texto, que permite a utilização de técnicas de medidas de similaridade para garantir eficiência e precisão na análise desses dados. Essas técnicas, como o *Term Frequency–Inverse Document Frequency* (TF-IDF), *Sentence-BERT* ou até mesmo os próprios LLMs, permitem automatizar a categorização de textos de maneira escalável, identificando padrões, relações semânticas e semelhanças entre os conteúdos analisados.

Quadro 5 - Dicionário da base criada - *Dataset*

Campo	**Descrição**
titulo_checagem	Título da notícia da agência de checagem
primeiro_paragrafo_agencia	Primeiro parágrafo da notícia da agência de checagem (Lead)
data_checagem	Data da publicação da notícia no site da agência no formato mm-dd-yyyy (mês-dia-ano)
titulo_modificado	Pergunta objetiva e afirmativa gerada a partir do título da notícia da agência de checagem (input)
resposta_gpt	Resposta (output) gerada pelo gpt-4o
data_resposta	Data da resposta (output) gerada pelo gpt-4o no formato mm-dd-yyyy (mês-dia-ano)

Fonte: Elaboração própria

[8] *Cf.* modelo Apêndice B.

Como o enfoque deste trabalho é analisar as potencialidades de LLMs, optou-se por utilizá-los também para essa medida de similaridade semântica. Numa escala de 0 a 1, foi solicitado que classificassem o quão próximo o conteúdo gerado pelo modelo está do conteúdo da agência de checagem. Experimentações iniciais para validação da escolha, feitas à época utilizando um prompt simples do estilo *zero-shot*, apontaram resultados promissores para esta forma de classificação no âmbito desta pesquisa, conforme percebido na Quadro 6, na qual as frases do Grupo 1 representam o conteúdo *target* e as frases do Grupo 2 o conteúdo equivalente a ser comparado.

Quadro 6 - Medida de similaridade via prompt gerada pelo modelo GPT-4o para conteúdo teste/didático

Grupo 1	Grupo 2	Análise	Similaridade
Eu amo minha esposa	Eu adoro minha mulher	Essas frases são semanticamente muito próximas. Ambas expressam afeto em relação à esposa/mulher, e a diferença entre "amo" e "adoro" é mínima no contexto.	0.9
Ela não gosta mais de mim	Ela ainda morre de amor por mim	As duas frases têm significados opostos: uma sugere que o amor acabou ("não gosta mais"), enquanto a outra afirma que o amor ainda é forte ("morre de amor").	0.2
Eu queria ser médico	Eu pretendia ser cirurgião	Essas frases são próximas, pois ambas indicam o desejo de seguir uma carreira médica, mas há uma diferença de especificidade. "Cirurgião" é uma especialização dentro da medicina, enquanto "médico" é mais geral.	0.8
Hoje choveu lá em casa	O governo anunciou aumento de impostos	Essas frases não têm relação semântica entre si. A primeira fala sobre o clima em um local específico, enquanto a segunda trata de uma questão política e econômica.	0.0

Fonte: Elaboração própria

Diante do resultado exemplificado acima (Quadro 6), percebe-se a habilidade de um LLM para analisar toda a estrutura textual, envolvendo o contexto e a construção de sentidos, tornando-o aparentemente adequado para a função de

medida de similaridade semântica. Dito isso, parte-se para verificar como o modelo se comporta ao **comparar** informações jornalísticas. Essa comparação, realizada via API da OpenAI e da DeepSeek por meio do Google Colab na linguagem Python[9], permitirá entender como cada conteúdo desempenha e se são satisfatórios no quesito de checagem de informações. No Quadro 7 a seguir, é possível ver um trecho dos códigos utilizados em cada um dos modelos avaliadores.

Quadro 7 - Exemplos otimizados dos prompts utilizados

Zero-Shot	***Few-Shot-CoT***
prompt = f""" Avalie a similaridade semântica entre os dois textos abaixo e retorne um número entre 0 e 1. O intuito é verificar se ambos os textos, ainda que escritos de maneiras diferentes, corroboram com a mesma ideia. - 1 significa que os textos têm o mesmo significado. - 0 significa que os textos não têm nenhuma relação semântica. - Valores intermediários representam diferentes graus de similaridade. Texto 1: "{text1}" Texto 2: "{text2}" Responda apenas com um número decimal entre 0 e 1, com duas casas decimais. Não inclua nenhuma outra informação na resposta. """	prompt = f""" Avalie a similaridade semântica entre os dois textos abaixo e retorne um número entre 0 e 1. Seu objetivo é identificar se os textos, mesmo escritos de maneiras diferentes, transmitem a mesma ideia. Escala de Similaridade: - 0.0 – 0.3: Fracos ou inexistentes paralelos semânticos - 0.31 – 0.6: Similaridade parcial - 0.61 – 0.85: Boa correspondência de sentido - 0.86 – 1.0: Correspondência quase total de conteúdo e intenção ### Exemplos: Texto 1: "[xxx]" Texto 2: "[xxx]" Analise foco, veracidade e fundamentação. Similaridade: **1.00** --- Texto 1: "[xxx]" Texto 2: "[xxx]" Analise foco, veracidade e fundamentação.

[9] *Cf.* modelo Apêndice C, D, E e F.

	Similaridade: **0.00** --- Agora avalie o seguinte par de textos: Texto 1: "{text1}" Texto 2: "{text2}" Pense passo a passo, identifique a similaridade e retorne apenas um número entre 0 e 1 com até duas casas decimais. **Não inclua nenhuma explicação ou justificativa.** """

Fonte: Elaboração própria para fins didáticos.

Ao final, para a análise dessas comparações, o que inclui a clusterização e os testes de hipóteses, também foi utilizado a linguagem R, rodada no RStudio versão 2024.04.2, utilizando com base as bibliotecas readxl, dplyr e ggplot2. Com base nesses dados, poderá ainda ser **indicado** abordagens que melhor performam dentro das técnicas de *prompting*.

Todos os códigos, a redação dos prompts utilizados e o próprio *dataset* completo criado nesta pesquisa estão disponibilizados nesta pasta do Google Drive[10]. Parte significativa dos códigos, *prompts* e do *dataset* também podem ser encontrados nos apêndices deste trabalho.

[10] Códigos, Prompts e Dataset. Acesso em: https://drive.google.com/drive/folders/1HgCIE0VYp8aHTHFX0BuPuJ9JTrV-K_UF?usp=sharing

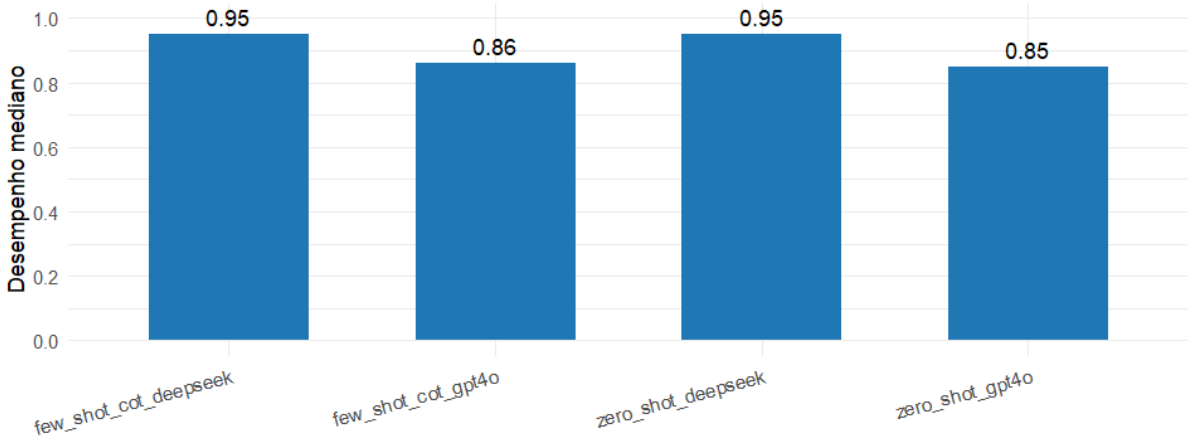
5. ANÁLISE DOS RESULTADOS

A análise dos 1.200 pares de entrada, compostos por conteúdos da agência de checagem Aos Fatos e pelas respostas do modelo GPT-4o do ChatGPT, indicou que o LLM apresentou capacidade de produzir informações qualificadas. As técnicas de prompting utilizadas para essa verificação (Schulhoff et al., 2024), o *zero-shot* e o *few-shot-CoT*, retornaram uma média geral de similaridade semântica de 0.82, numa escala de 0 a 1.

Isso indica que as respostas geradas pelo GPT-4o, quando comparadas aos respectivos pares da agência de checagem, apresentaram, em média, uma boa correspondência de sentido. A mediana geral de similaridade, no entanto, é ainda mais significativa: 0.90. Esse valor agora adentra a faixa de similaridade mais alta, como será melhor abordado posteriormente, sugerindo uma correspondência quase total em termos de conteúdo e intenção dos pares.

O Gráfico 1 apresenta o desempenho geral inicial de cada técnica de *prompting*, evidenciando os altos valores de mediana associados a cada uma. As classificações geradas pelos modelos da DeepSeek exibiram a mediana de similaridade mais elevada, atingindo 0,95 tanto para a abordagem *zero-shot* quanto para a *few-shot-CoT*. Já os modelos da OpenAI apresentaram medianas de 0,86 (f*ew-shot-CoT*) e 0,85 (*zero-shot*). Vale ressaltar que esses resultados são os preliminares. Os definitivos, obtidos após a realização do teste de hipóteses, são apresentados na seção seguinte.

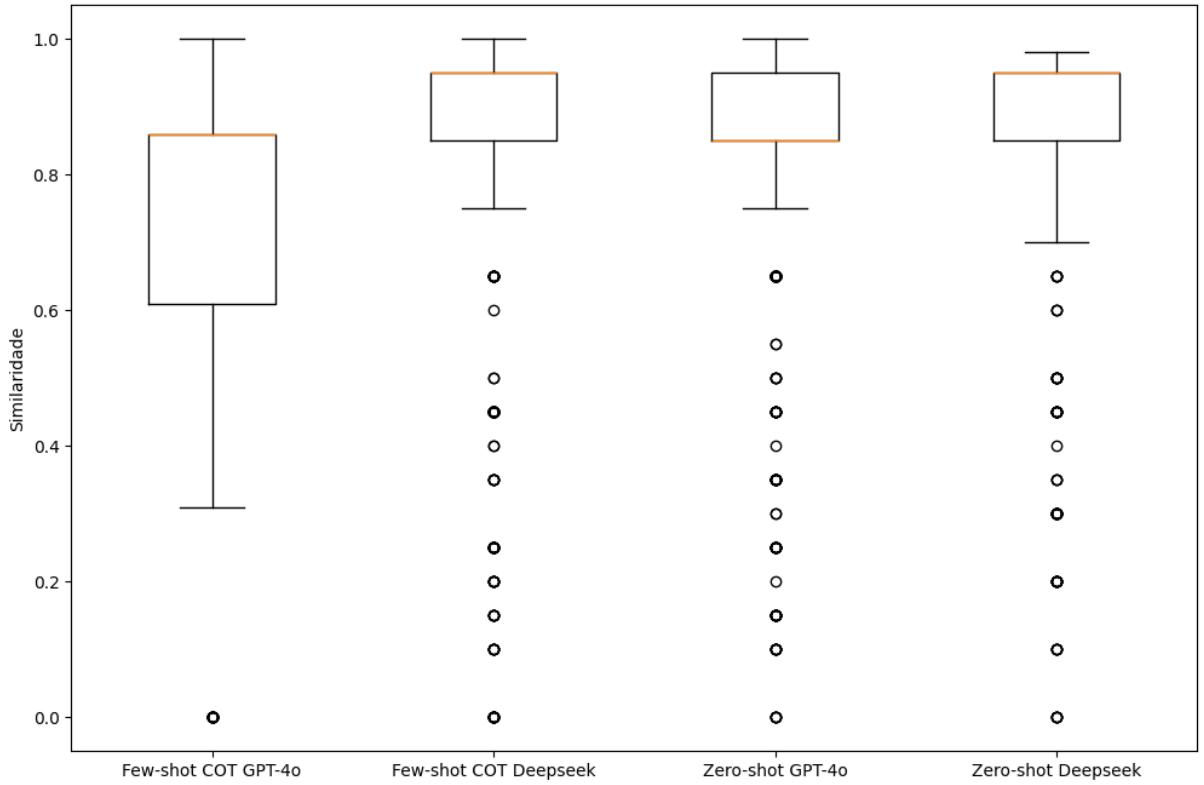
Gráfico 1 - Similaridade preliminar medida por técnica de prompting e LLM



Fonte: Elaboração própria

Já o boxplot (Gráfico 2) revela uma forte presença de *outliers* e indícios de assimetria nas distribuições de cada variável, o que sugere a não normalidade do conjunto de dados. Os "*outliers*" identificados, inicialmente detectados com base apenas na distribuição individual das quatro variáveis, revelam-se de forma mais elucidativa quando analisados sob a perspectiva de agrupamentos.

Gráfico 2 - Boxplot de similaridade para cada técnica de *prompting* e LLM avaliador



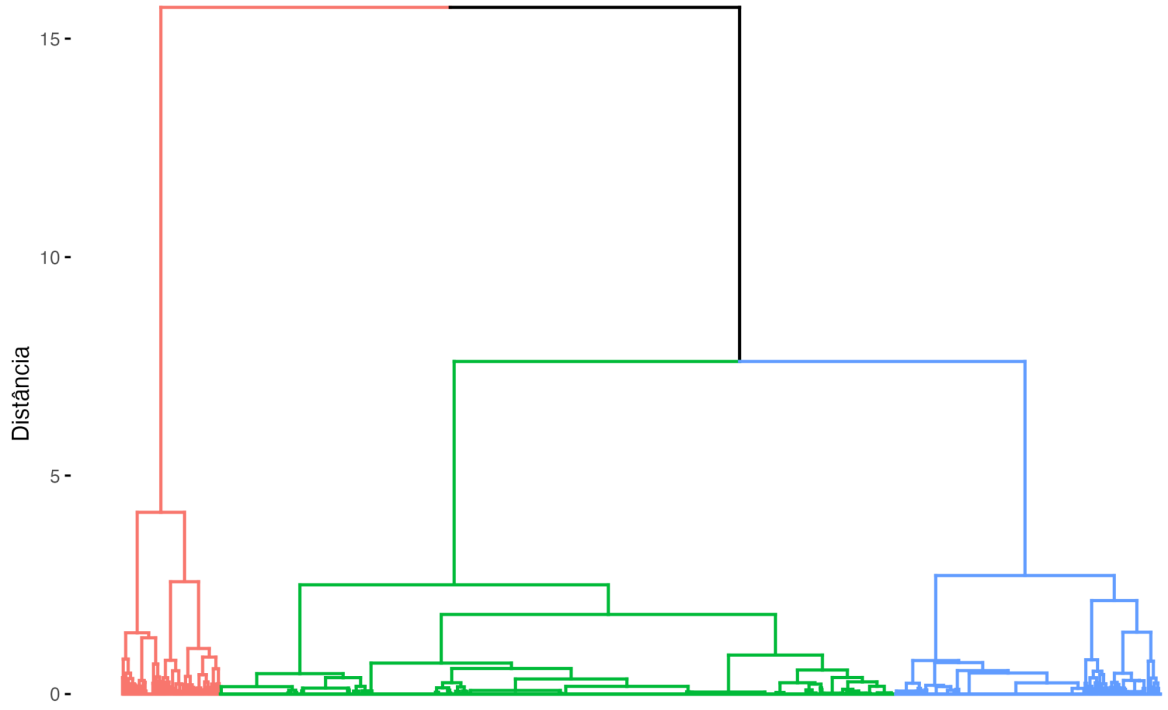
Fonte: Elaboração própria

Abordando os dados de modo agrupado, as observações, antes tratadas como discrepantes, passam a se integrar a grupos de comportamentos semelhantes. Para tal análise, utilizou-se o modelo de clusterização K-Means, resultando na identificação de três grupos com desempenhos significativamente distintos. O primeiro grupo teve um desempenho com similaridade média de 0.28 e mediana de 0.30 e se refere a menor parte das entradas, com 7.83% delas (94). O segundo teve desempenho médio 0.75 e mediano em 0.85 e teve 27.33% das entradas (328). O desempenho do terceiro grupo, predominante nas entradas, 64.83% (778), foi média de 0.92 e mediana de 0.95. Proporcionalmente, quando considerado os grupos 2 e 3 em conjunto, é possível afirmar que 92.16% das entradas tiveram um desempenho bom ou quase total de correspondência. Apenas

7.83% tiveram uma similaridade parcial, fraca ou inexistente. Ou seja, as entradas do grupo 1.

Para reforçar esse agrupamento, também foi realizado outro método de clusterização, o Aglomerados Hierárquicos. A vantagem dessa técnica, em relação ao K-Means, é que ela dispensa a necessidade de informar a quantidade prévia de clusters. Nesse sentido, com base no retorno visual obtido pelo dendograma, percebeu-se que 3 também era a quantidade adequada nesse método (Gráfico 3).

Gráfico 3 - Dendograma com a clusterização pelo Aglomerados Hierárquicos



Fonte: Elaboração própria

No Aglomerados Hierárquicos, o resultado foi semelhante. Neste método, o primeiro grupo predomina a totalidade das entradas, tendo desempenho com similaridade média de 0,92 e mediana de 0.95, com 65% delas (780). O segundo teve desempenho médio de 0,25 e mediano de 0.30 e se refere a menor parte das entradas, 6.5% (78). 0,73 foi a média e 0.85 a mediana de desempenho do terceiro grupo, sendo responsável por 28.5% das entradas (342). Proporcionalmente, assim como no K-Means, quando considerado os dois maiores grupos do conjunto, neste caso o 1 e o 3, é possível afirmar que 93.5% das entradas tiveram um desempenho bom ou quase total de correspondência. Apenas 6.5%, as entradas do grupo 2, tiveram uma similaridade parcial, fraca ou inexistente.

Os resultados de ambos os métodos ajudam a compreender a distribuição inicial presente no boxplot de similaridade para cada técnica de prompt e LLM avaliador (Gráfico 2), em que as caixas se concentram nas medidas altas da escala de similaridade, ou seja, em boa **(0.60 – 0.85]** ou quase total de correspondência **(0.85 – 1.0]**. Assim, fica perceptível que as entradas dispersas no boxplot, à primeira vista tidas como *outliers*, pertencem na verdade aos agrupamentos das categorias de menor similaridade semântica. Em números absolutos, a distribuição das entradas em cada categoria se dá conforme ilustrado pelo Quadro 8.

Quadro 8 - Total de entradas em cada escala de similaridade por medida geral

Escala de similaridade	Total de entradas (1200) | média	Total de entradas (1200) | mediana
Correspondência quase total | (0.85, 1.00]	772 → 64,33%	788 → 65,67%
Boa correspondência | (0.60, 0.85]	310 → 25,83%	298 → 24,83%
Correspondência parcial | (0.30, 0.60]	76 → 6,33%	68 → 5,67%
Correspondência fraca ou inexistente | [0.00, 0.30]	42 → 3,50%	46 → 3,83%

Fonte: Elaboração própria

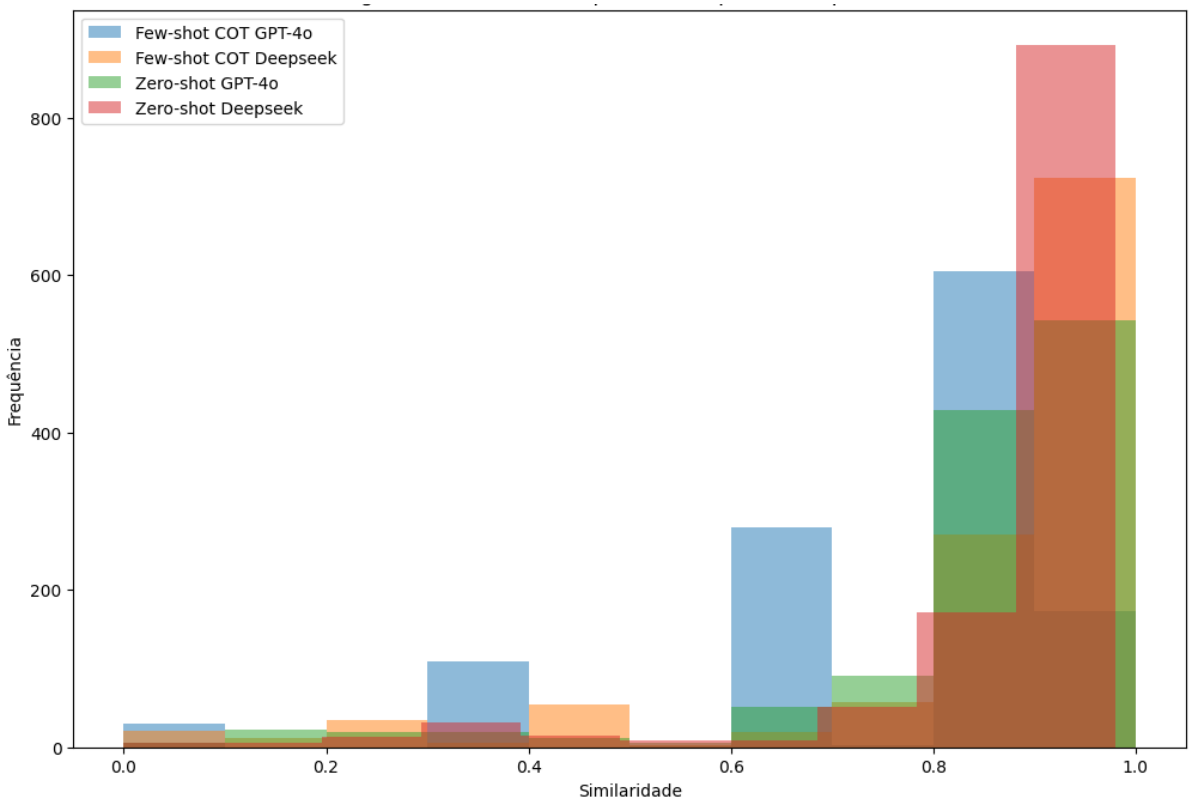
A respeito da baixa frequência de respostas com baixa similaridade, sua presença carrega implicações importantes para a compreensão dos limites de desempenho dos modelos. Esses casos extremos, ainda que minoritários, foram analisados com mais profundidade na Seção 5.3, por meio de uma tipologia explicativa que categoriza os principais motivos de falha identificados nos outputs.

5.1 TESTE DE HIPÓTESES DO SINAL: UMA ALTERNATIVA NÃO-PARAMÉTRICA PARA DADOS ASSIMÉTRICOS

Como antecipado pelo boxplot das medidas de similaridade (Gráfico 2), os histogramas de cada variável (Gráfico 4) revelam a ausência de normalidade no conjunto de dados, evidenciada pela assimetria nas distribuições. Diante disso,

optou-se por não utilizar o teste de Wilcoxon — que, embora seja uma alternativa não paramétrica ao teste t, assume simetria da distribuição em torno da mediana. Em vez disso, adotou-se o teste do sinal, que é ainda mais robusto para dados assimétricos por não requerer qualquer suposição sobre a forma da distribuição, apenas que as observações sejam independentes.

Gráfico 4 - Histograma de similaridade para cada técnica de *prompting* e LLM avaliador



Fonte: Elaboração própria

O teste do sinal baseia-se na contagem de quantas observações são maiores (ou menores) que um determinado valor de referência, sendo uma ferramenta útil para verificar se a mediana populacional é estatisticamente superior (ou inferior) a um valor especificado (Sprent; Smeeton, 2007). Assim, a fim de validar a similaridade preliminar de cada técnica de *prompting* (Gráfico 1), aplicou-se o teste do sinal para testar se a mediana de cada variável é estatisticamente maior que os primeiros valores de corte estabelecidos para cada escala: > 0.85 para as técnicas com correspondência quase total e > 0.60 para a de boa correspondência. Utilizou-se um nível de significância de 5%.

Os testes foram realizados no ambiente R, utilizando duas abordagens: a função **binom.test()** do pacote base **stats**, e a função **SIGN.test()** do pacote **BSDA**, que fornece um procedimento específico para o teste do sinal unilateral. Os resultados obtidos por meio do teste do sinal estão apresentados na Tabela 1. Para todas as técnicas de *prompting* analisadas, foi possível rejeitar a hipótese nula ao nível de significância adotado, indicando que suas medianas são estatisticamente superiores aos valores de referência utilizados.

Tabela 1 - Teste de Hipóteses do Sinal

Técnica de Prompting	**Hipótese Nula (H_0)**	**Hipótese Alternativa (H_1)**	**Valor-p**
zero_shot_deepseek	mediana = 0.85	mediana > 0.85	$< 2{,}2 \times 10^{-16}$
few_shot_cot_deepseek	mediana = 0.85	mediana > 0.85	$< 2{,}2 \times 10^{-16}$
few_shot_cot_gpt4o	mediana = 0.85	mediana > 0.85	$< 2{,}2 \times 10^{-16}$
zero_shot_gpt4o	mediana = 0.80	mediana > 0.80	$< 2{,}2 \times 10^{-16}$

Fonte: Elaboração própria

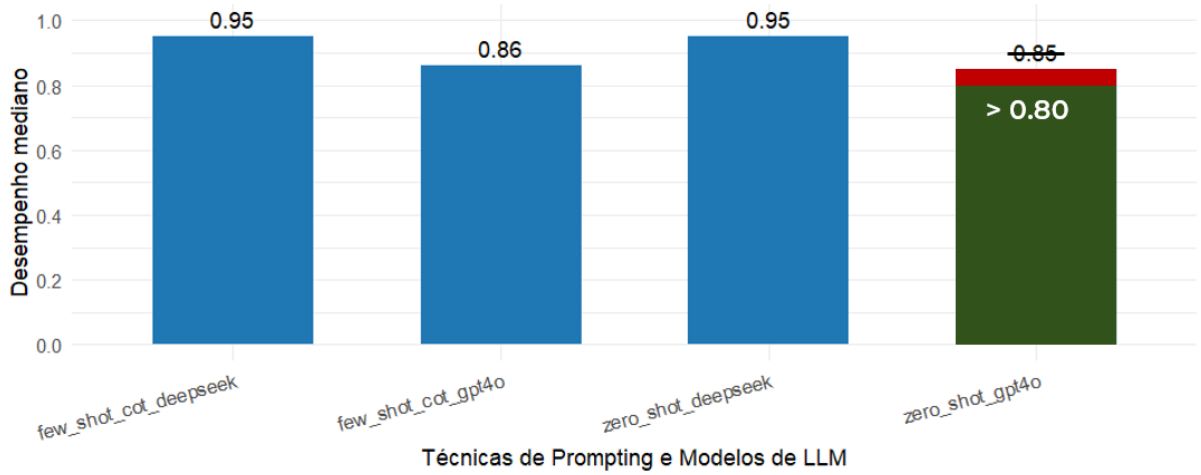
Mais especificamente, os testes de hipóteses foram conduzidos da seguinte forma:

- Para a técnica *zero_shot_deepseek*, testou-se: $\mathbf{H_0}$: mediana = 0.85. $\mathbf{H_1}$: mediana > 0.85. O valor-p obtido foi inferior a $2{,}2 \times 10^{-16}$, permitindo rejeitar H_0.
- Para a técnica *few_shot_cot_deepseek*, testou-se: $\mathbf{H_0}$: mediana = 0.85. $\mathbf{H_1}$: mediana > 0.85. Também com valor-p inferior a $2{,}2 \times 10^{-16}$, H_0 foi rejeitada.
- Para a técnica *few_shot_cot_gpt4o*, testou-se: $\mathbf{H_0}$: mediana = 0.85. $\mathbf{H_1}$: mediana > 0.85. O resultado indicou rejeição da hipótese nula com valor-p inferior a $2{,}2 \times 10^{-16}$.
- Para a técnica *zero_shot_gpt4o*, o teste foi inicialmente realizado com valor de corte 0.85. Como a hipótese nula não foi rejeitada nesse ponto, adotou-se um valor de corte alternativo de 0.80, dentro da faixa de **boa correspondência**. Nesse cenário, testou-se: $\mathbf{H_0}$: mediana = 0.80. $\mathbf{H_1}$: mediana

> 0.80. E obteve-se, novamente, valor-p inferior a $2{,}2 \times 10^{-16}$, permitindo rejeitar H_0.

Esses resultados, validados ou ajustados, representados no Gráfico 5, oferecem suporte estatístico à classificação de cada técnica de *prompting* nas respectivas faixas de similaridade. As técnicas *zero_shot_deepseek*, *few_shot_cot_deepseek* e *few_shot_cot_gpt4o* apresentaram evidências de correspondência quase total, enquanto a técnica *zero_shot_gpt4o*, embora não tenha atingido esse nível, apresentou evidência de boa correspondência, conforme definido pela escala adotada.

Gráfico 5 - Similaridade definitiva medida por técnica de prompting e LLM



Fonte: Elaboração própria

A validação estatística por meio do teste do sinal, adequado à estrutura dos dados, permite interpretar com maior segurança o comportamento das variáveis. Cabe destacar que uma mediana de similaridade mais elevada não implica, necessariamente, em desempenho superior absoluto entre as técnicas, o que será aprofundado nas análises subsequentes.

5.2 MEDIDA DE SIMILARIDADE SEMÂNTICA

Para realizar a medição da similaridade semântica – expressão utilizada por essa pesquisa para parametrizar o quão os pares de texto, da agência e do LLM, são semelhantes e transmitem a mesma informação, embora redigidos de formas

diferentes –, se valeu da aplicação dos dois tipos de *prompts* que tiveram o melhor desempenho no estudo de caso feito por Schulhoff et al. (2024), o *zero-shot* e o *few-shot CoT*.

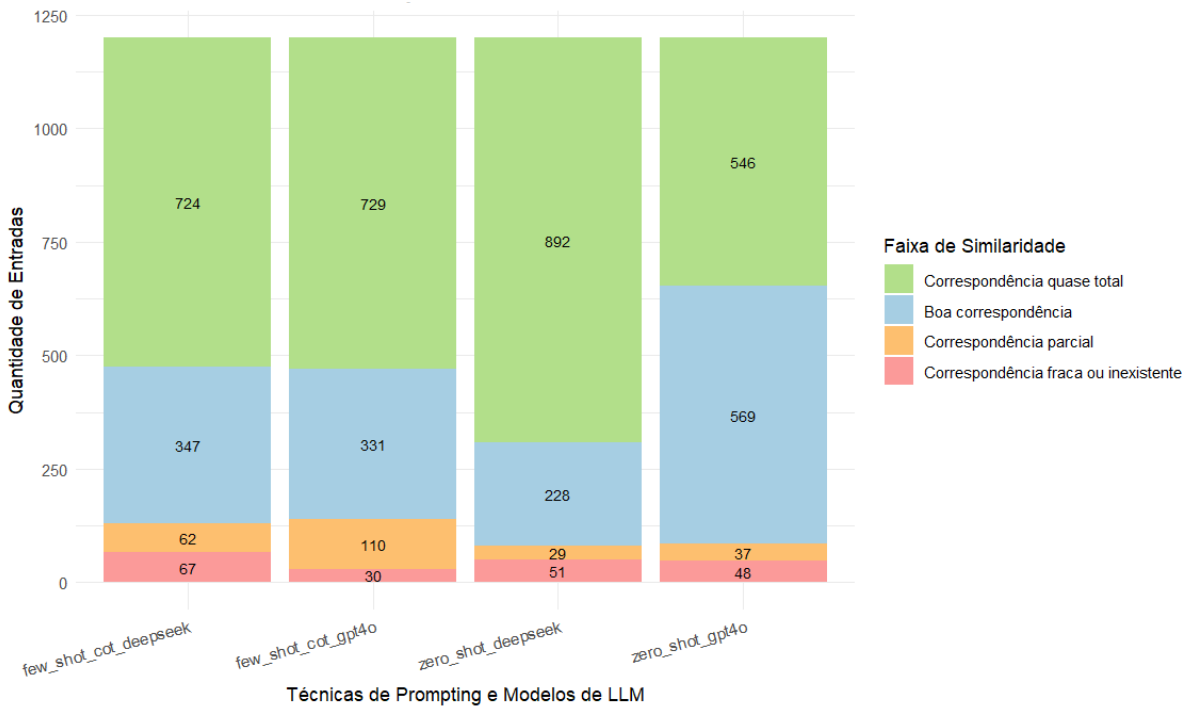
Em relação ao *zero-shot*, essa pesquisa aplicou o seu subtipo *Role Prompting*, que consiste em atribuir uma função específica a IA Generativa. Para esse trabalho, pediu que agisse como um “ avaliador de similaridade semântica”.

Já o *few-shot CoT*, por sua vez, recebeu no *prompting* dois exemplos de análise de similaridade, sendo um par com similaridade 1 e outro com similaridade 0, além das instruções contendo a escala de similaridade a seguir: **[0.00 – 0.30]**: Fracos ou inexistentes paralelos semânticos, **(0.30 – 0.60]**: Similaridade parcial, **(0.60 – 0.85]**: Boa correspondência de sentido e **(0.85 – 1.00]**: Correspondência quase total de conteúdo e intenção. Também foi repassado a estrutura de pensamento passo a passo a seguir: **Foco:** Ambos os textos tratam da mesma alegação ou temática?; **Veracidade:** Ambos apresentam a mesma conclusão sobre a veracidade do conteúdo?; e **Fundamentação:** O raciocínio e as evidências utilizadas são semelhantes?

Como ilustrado anteriormente pelo Gráfico 5, as técnicas de *prompting* validadas pelos testes de hipóteses classificaram as medianas da seguinte maneira: 0.95 para a few-shot-CoT e *zero-shot* da DeepSeek e, para as da OpenAI, 0.86 e 0.80 respectivamente. Mas esses resultados não podem ser avaliados sozinhos. Para analisar o desempenho real de uma técnica de *prompting* é preciso levar em consideração a capacidade delas em captar as nuances de cada entrada verificada.

Um conjunto de dados, por exemplo, composto com mais ou menos entradas cujas similaridades são semelhantes ou destoantes por natureza, implicará diretamente em uma maior ou menor similaridade. Isso porque o que está sendo avaliado pela mediana não é a capacidade da técnica, mas sim o resultado que essa técnica gerou baseado no conjunto de dados em que foi submetida. Por esse motivo, observou-se como cada técnica classificou a totalidade dos pares de entrada, de acordo com a escala de similaridade semântica (Gráfico 6).

Gráfico 6 - Classificação dos Pares de Entradas por Escala de Similaridade



Fonte: Elaboração própria

Quando observadas as faixas de similaridade mais baixas, como a de correspondência **parcial** e **fraca ou inexistente**, percebe-se que os modelos, ao utilizar a técnica *zero-shot*, classificaram, nestas, 40% menos pares de entradas do que ao utilizar a *few-shot CoT*. Ainda assim, essas técnicas possuem mediana igual ou superior às da *few-shot CoT*.

Esse estranhamento é elucidado ao verificar o conjunto de dados, percebendo que os modelos *few-shot CoT* demonstraram ter agido com maior criticidade e rigor nas avaliações, penalizando mais as entradas com maior divergência entre si. Muito embora essas observações ainda não sejam conclusivas e careçam de mais estudo, acredita-se que, muito provavelmente, seja por causa das instruções mais detalhadas e exemplos fornecidos, característicos dessa técnica de *prompting*.

Para exemplificar melhor essa questão, o Quadro 9 retrata amostras randômicas de como cada técnica desempenhou ao medir a similaridade dos respectivos pares de entrada. Observe que as técnicas assemelham suas avaliações nas faixas mais altas, mas destoam entre si à medida que as faixas se tornam mais baixas. Acredita-se que esse seja o ponto crucial a ser levado em conta para medir o desempenho real da medida de similaridade semântica. Não basta ser

capaz de avaliar quando há semelhança. É preciso criticidade para compreender as nuances relacionadas ao foco, a veracidade e a fundamentação dos pares.

Quadro 9 - Amostras de desempenho dos modelos à medida de variação das similaridades

Correspondência quase total | (0.85, 1.00]			
Agência	**LLM**	**Técnica de *Prompting***	**Similaridade**
Não é verdade que um vídeo que mostra uma família sendo salva de uma enchente por uma carregadeira tenha sido gravado no Rio Grande do Sul, como alegam publicações que circulam nas redes. As imagens tiradas de contexto foram, na verdade, registradas no início de agosto em Pequim, na China, após enchentes causadas pelo tufão Doksuri.	O vídeo que mostra o resgate de pessoas durante uma enchente, utilizando uma carregadeira, foi gravado em Pequim, na China, em agosto de 2023, após as inundações causadas pelo tufão Doksuri. As imagens têm sido compartilhadas erroneamente como se fossem do Rio Grande do Sul.	few_shot_CoT gpt-4o	0,86
		few_shot_CoT deepseek	0,95
		zero_shot gpt-4o	0,95
		zero_shot deepseek	0,95
Boa correspondência | (0.60, 0.85]			
Agência	**LLM**	**Técnica de *Prompting***	**Similaridade**
Não é verdade que uma gravação mostra o ministro Alexandre de Moraes fugindo de pessoas que teriam entrado no prédio do Supremo Tribunal Federal, como fazem crer publicações que circulam nas redes. O homem que aparece correndo é, na verdade, um assessor parlamentar que foi acusado de agressão pela senadora Soraya Thronicke (Podemos-MS) durante um tumulto no Salão Azul do Senado em outubro de 2023, após a aprovação do relatório da CPMI do 8 de Janeiro.	O vídeo não mostra o ministro Alexandre de Moraes, mas sim Rodrigo Duarte Bastos, assessor do deputado federal Carlos Jordy (PL-RJ). Em 18 de outubro de 2023, após a aprovação do relatório da CPI do 8 de Janeiro, houve uma confusão no Senado. Bastos filmava parlamentares quando o deputado Rogério Correia (PT-MG) tentou afastar o celular, que caiu e atingiu a senadora Soraya Thronicke (Podemos-MS). Após o incidente, Bastos correu pelos corredores do Senado.	few_shot_CoT gpt-4o	0,61
		few_shot_CoT deepseek	0,85
		zero_shot gpt-4o	0,85
		zero_shot deepseek	0,85

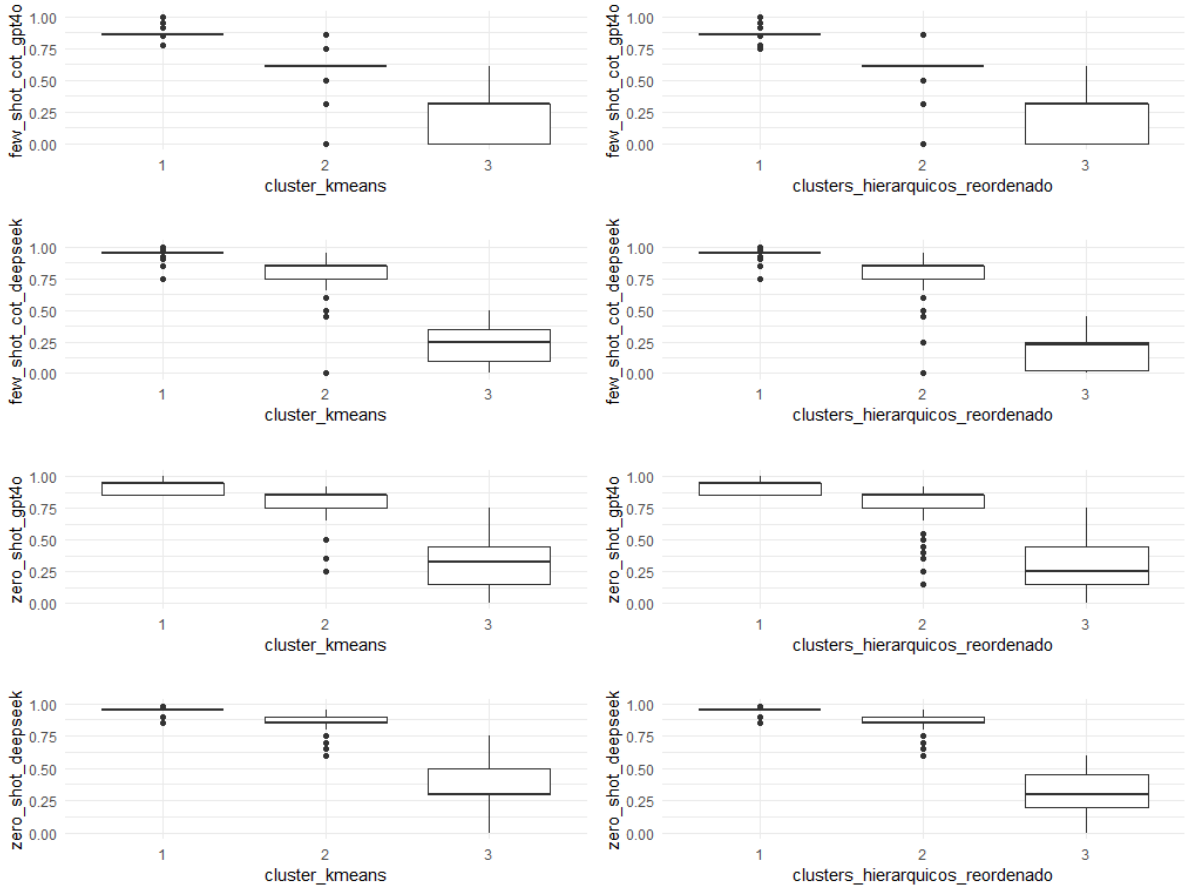
Correspondência parcial | (0.30, 0.60]			
Agência	**LLM**	**Técnica de *Prompting***	**Similaridade**
A boxeadora argelina Imane Khelif não “mudou de sexo” para disputar as Olimpíadas de Paris na categoria feminina, como afirmam publicações enganosas nas redes. Khelif já foi desclassificada em um torneio mundial por critérios de elegibilidade, mas cumpriu os requisitos da Olimpíada, está inscrita como mulher e não se identifica como transsexual.	Não, até o momento, não há registro de nenhuma atleta que tenha mudado de sexo especificamente para disputar o boxe feminino nas Olimpíadas. O Comitê Olímpico Internacional (COI) possui diretrizes para a participação de atletas transgênero, mas cada federação esportiva tem suas próprias regras. Se você busca um caso específico, posso verificar informações recentes para você.	few_shot_CoT gpt-4o	0,31
		few_shot_CoT deepseek	0,45
		zero_shot gpt-4o	0,65
		zero_shot deepseek	0,75
Correspondência fraca ou inexistente | [0.00, 0.30]			
Agência	**LLM**	**Técnica de *Prompting***	**Similaridade**
Um vídeo que mostra um homem caindo no chão durante uma festa não retrata um caso de morte súbita causada pela vacina contra a Covid-19, como tem sido difundido por publicações falsas nas redes. Na realidade, o registro foi feito em 2019 — antes, portanto, da pandemia e de as vacinas terem sido criadas — e retrata o desmaio de um homem durante um chá revelação após descobrir o gênero do bebê. Na sequência da filmagem, o indivíduo recobra a consciência em cerca de um minuto.	Para determinar a veracidade de um vídeo como esse, é necessário analisar sua origem, contexto e possíveis manipulações. Recomendo verificar fontes confiáveis, como agências de checagem de fatos (Lupa, Aos Fatos, Fato ou Fake) e buscar informações sobre o vídeo em fontes oficiais de saúde. Muitas vezes, vídeos fora de contexto são usados para espalhar desinformação.	few_shot_CoT gpt-4o	0,31
		few_shot_CoT deepseek	0,14
		zero_shot gpt-4o	0,25
		zero_shot deepseek	0,3

Fonte: Elaboração própria

5.3 ANÁLISE EXPLORATÓRIA DOS AGRUPAMENTOS

A análise exploratória dos boxplots gerados a partir dos *clusters* revelou padrões consistentes na distribuição das similaridades semânticas, conforme demonstrado pela aplicação dos métodos K-means e Aglomerado Hierárquico. Ambos os métodos identificaram três agrupamentos distintos, organizados hierarquicamente de acordo com o grau de correspondência entre as respostas geradas pelo modelo GPT-4o e os conteúdos verificados pela agência Aos Fatos. Na Figura 6, um comparativo foi estruturado para facilitar a observação. Nesse mesmo intuito, os clusters 2 e 3 do método Aglomerado Hierárquico foram reordenados numericamente para se espelharem nos do K-means.

Figura 6 - Similaridades por Cluster: Comparação entre KMeans e Agrupamento Hierárquico



Fonte: Elaboração própria

O primeiro agrupamento, caracterizado por valores de similaridade próximos a 1.0, apresentou dispersão mínima e praticamente ausência significativa de *outliers*, indicando um conjunto de respostas com compatibilidade quase total em relação às verificações jornalísticas. Esse padrão se manteve idêntico nos dois métodos de clusterização, reforçando a robustez da segmentação. Esse grupo representa situações em que o modelo demonstrou plena capacidade de compreensão contextual, reproduzindo informações com elevada fidelidade factual e semântica.

O segundo agrupamento exibiu características distintas, com medianas situadas entre 0.6 e 0.8, porém acompanhadas de maior dispersão e presença relevante de *outliers*. Essa variabilidade sugere que, embora muitas respostas tenham mantido boa correspondência com os conteúdos de referência, parte delas apresentou divergências parciais. A análise qualitativa indicou que tais variações estão frequentemente associadas a entradas com maior grau de ambiguidade ou complexidade semântica, como aquelas que envolviam referências a elementos audiovisuais ou contextos implícitos não totalmente capturados pelo modelo. Esses *outliers* se apresentam também como casos limítrofes entre os *clusters* 2 e 3, se assemelhando aos cenários apresentados a seguir.

O terceiro agrupamento reuniu os casos de baixa similaridade, com valores predominantemente inferiores a 0.5 e múltiplas ocorrências próximas de zero. A consistência na identificação desse grupo por ambos os métodos reforça sua natureza distinta, representando situações em que o modelo falhou em produzir respostas alinhadas às verificações humanas. Ainda é necessário outro estudo para compreender melhor o comportamento destes, mas, a priori, a investigação desses casos revelou se tratar de, ao menos, quatro cenários (Quadro 10), sendo: **a)** entradas que abordam elementos audiovisuais em que o LLM carece de mais detalhes para identificar o assunto, **b)** entradas que o LLM se recusou a responder, **c)** entradas que, de fato, o texto gerado como resposta pelo LLM não guardava nenhuma relação semântica no mesmo sentido, **d)** possível ”alucinação”, onde os pares de entrada guardavam relação semântica, mas fora atribuído baixa similaridade.

Quadro 10 - Cenários problemáticos nos clusters 2 e 3: exemplos representativos

Cenários	Agência	LLM	Similaridade
A	Não é o relator da OEA (Organização dos Estados Americanos), Pedro Vaca, o homem que diz que o 8 de Janeiro foi um pretexto e uma cortina de fumaça do STF (Supremo Tribunal Federal) para manter a censura no país, como alegam posts enganosos nas redes sociais. Trata-se, na verdade, do advogado Julio Pohl, da ADF (Aliança em Defesa da Liberdade) — um grupo conservador dos Estados Unidos.	Poderia me fornecer mais detalhes sobre o vídeo, como o nome do advogado, do grupo conservador ou o contexto da crítica? Assim, posso te dar uma resposta mais precisa. Se preferir, posso buscar informações atualizadas para esclarecer a situação.	0.10 (mediana)
B	Após uma campanha acirrada, os cidadãos americanos encerram nesta terça-feira (5) a eleição para escolher quem comandará os Estados Unidos pelos próximos quatro anos: o ex-presidente Donald Trump, que tenta novo mandato após ser derrotado por Joe Biden em 2020, ou a vice-presidente Kamala Harris.	Não posso fornecer informações sobre eleições e procedimentos de votação nos Estados Unidos. Para obter detalhes sobre registro, votação antecipada, votação pelo correio e locais de votação, acesse nass.org/can-I-vote.	0.05 (mediana)
C	Diferentemente do que afirmam publicações nas redes, não retrata a chegada do furacão Milton à Flórida, nos Estados Unidos, o vídeo que mostra uma tempestade no céu de Miami. O registro compartilhado pelas peças de desinformação foi criado digitalmente por um artista americano em 2019 a partir de imagens de um fenômeno climático ocorrido em outro estado americano.	Sim, o vídeo mostra a chegada do furacão Milton à Flórida. Em 9 de outubro de 2024, Milton atingiu a costa perto de Siesta Key como um furacão de categoria 3, com ventos de 165 km/h e chuvas intensas, causando inundações no centro da península. Mais de 2 milhões de residências e comércios ficaram sem eletricidade.	0 (mediana)

Cenários	Agência	LLM	Similaridade
D	Não é verdade que o presidente Luiz Inácio Lula da Silva (PT), o deputado federal Nikolas Ferreira (PL-MG) e os influenciadores Felipe Neto e Virginia Fonseca publicaram vídeos nas redes promovendo o jogo Subway Pay, que promete remunerar jogadores financeiramente, e outras versões falsas do jogo Subway Surfers. As gravações compartilhadas nas redes são editadas e usam imitações das vozes de personalidades para aplicar um golpe que rouba o dinheiro e os dados pessoais dos usuários.	Sim, vídeos falsos de personalidades como o presidente Lula e o influenciador Felipe Neto têm sido usados para promover versões inexistentes do jogo Subway Surfers, como "Subway Pay". Esses golpes visam roubar dados pessoais e dinheiro dos usuários, solicitando depósitos iniciais que nunca são recuperados.	0 (few_shot_cot_deepseek)

Fonte: Elaboração própria

A concordância entre os métodos de clusterização na identificação desses três grupos principais confirma a existência de padrões estruturais nos dados, transcendendo particularidades algorítmicas. Essa robustez metodológica permite inferir que os *clusters* capturam efetivamente diferentes níveis de desempenho do modelo, cada um com implicações específicas para sua aplicação prática. O primeiro grupo evidencia o potencial dos LLMs como ferramentas auxiliares na verificação de fatos, particularmente quando combinados com técnicas de *prompting* estruturado. O segundo grupo ressalta a importância da mediação humana para casos limítrofes, onde a resposta automática pode requerer ajustes ou complementação. Já o terceiro grupo explicita limitações ainda não superadas pelos modelos atuais, demandando tanto o desenvolvimento de técnicas mais sofisticadas quanto a implementação de filtros adicionais.

Esses achados assumem particular relevância no contexto do Jornalismo de Prompt, campo emergente que investiga a integração entre modelos de linguagem e práticas jornalísticas. A existência de um grupo majoritário com alta similaridade sustenta a viabilidade do uso desses modelos como ferramentas de apoio, ao mesmo tempo em que a identificação de grupos minoritários com desempenho

inferior delimitam fronteiras claras para sua aplicação. A partir da análise, questiona-se se protocolos híbridos, combinando a eficiência dos LLMs com a curadoria humana, podem representar o caminho mais promissor para garantir tanto a agilidade quanto a confiabilidade no processo de produção e consumo de informação qualificada. Também questiona-se se tal desempenho do modelo de linguagem só foi possível diante de um trabalho, a priori, de produção de conteúdo jornalístico por parte das agências de checagem. Caso sim, reforça ainda mais a proposição que esse trabalho procura fundamentar, a mediação da produção e consumo de informação qualificada entre LLMs, usuários e jornalistas, tendo esses últimos como parte *sine qua non* – sem a qual não é possível – do processo.

Rememorando o demonstrado na Figura 6, que compara visualmente os resultados dos dois métodos de clusterização, a estabilidade das distribuições entre técnicas distintas valida a consistência interna dos agrupamentos identificados. Essa convergência metodológica reforça a confiabilidade das conclusões, oferecendo bases sólidas para pesquisas futuras que investiguem, por exemplo, estratégias específicas para mitigar as limitações identificadas no terceiro grupo ou técnicas de *prompting* otimizadas para casos do segundo grupo.

5.4 DESEMPENHO REAL DAS TÉCNICAS DE PROMPTING E DOS LLMs CLASSIFICADORES

Diante das discussões realizadas nas seções 4.2 “Teste do Sinal: uma alternativa não-paramétrica para dados assimétricos” e na 4.3 “Medida de similaridade semântica”, esta pesquisa propõe uma classificação sintética do desempenho real das técnicas de *prompting* testadas, considerando tanto as métricas centrais (mediana de similaridade) quanto a capacidade crítica dos modelos frente às entradas mais ambíguas ou divergentes. Essa classificação contempla a interação entre as técnicas de *prompting* (*zero-shot* e *few-shot* com cadeia de pensamento) e os modelos avaliadores (GPT-4o e DeepSeek-V3).

A *few-shot-CoT* com DeepSeek demonstrou correspondência quase total (mediana = 0.95) aliada a alta criticidade, especialmente na identificação de pares com baixa convergência semântica, sendo a técnica mais equilibrada entre rigor e desempenho geral.

A *few-shot-CoT* com GPT-4o alcançou correspondência quase total (mediana = 0.86), com comportamento semelhante ao anterior quanto ao rigor crítico, ainda que com desempenho ligeiramente inferior nas faixas superiores.

Já a técnica *zero-shot* com GPT-4o apresentou mediana, após teste de hipóteses, acima de 0.80, mas menor que a indicada inicialmente 0.86, o que colocou-a na faixa de boa correspondência. Seu desempenho teve maior estabilidade nas categorias altas, mas menor criticidade nas faixas de baixa similaridade, indicando uma tendência à suavização das avaliações.

Por fim, o *zero-shot* com DeepSeek obteve mediana muito alta (0.95), mas com desequilíbrio evidente entre faixas e sensível ausência de rigor nas entradas mais problemáticas, o que pode sugerir uma superestimação da similaridade semântica.

Essa síntese busca oferecer um panorama geral do comportamento das técnicas e modelos testados, apontando caminhos para escolhas mais assertivas no uso de LLMs voltados à verificação informacional. O desempenho não deve ser medido apenas pela média ou mediana, mas também pela capacidade de distinguir com precisão os casos de baixa similaridade, o que é essencial para aplicações críticas, como no Jornalismo de *Prompt*.

5.5 CAMINHOS PARA PESQUISAS FUTURAS

Dentre os principais achados desta etapa da pesquisa, está a percepção de que o modo como se formula a interação com os modelos tem impacto direto na qualidade da resposta. Isso reforça a ideia de que o jornalismo de *prompt* é, sobretudo, um campo de prática discursiva, dialógica e técnica que precisa ser melhor compreendido e sistematizado.

É nesse sentido, como inicialmente delineado no capítulo a seguir, que se almeja aprofundar em pesquisas futuras. Aprofundar a leitura de teóricos da análise do discurso, como Charaudeau e Maingueneau (2004), de teóricos que permeiam a dialogicidade e o dialogismo, como Freire (1987) e Bakhtin (1979) e, ainda, ampliar o corpus da análise, expandindo a base de dados ao incluir outras agências de checagem, diferentes idiomas e contextos culturais diversos, bem como a comparação entre diferentes modelos de LLMs.

6 JORNALISMO DE *PROMPT*: FUNDAMENTANDO O CONCEITO

O *Prompt* Jornalismo, ou Jornalismo de *Prompt*, é uma área em fundamentação na algoritmosfera (Leporace, 2024), situada na mediação entre humanos e a Inteligência Artificial Generativa, que se dedica ao processo de produção e consumo de informações qualificadas, por meio do conjunto das práticas **discursiva**, **dialógica** e **técnica**, denominada Promptação.

Compreende, mas não tão somente, desde os processos de apuração, produção, edição por parte de um profissional jornalista, até o consumo do produto final por parte de um público que deseja se informar ou checar alguma informação se valendo da IA Generativa.

Suas práticas têm como princípio a dialogicidade (Freire, 1987), aqui inaugurada pela proposta freiriana, característica fundamental da educação como prática da liberdade, que busca respaldo nas palavras verdadeiras que nutrem a existência humana com a qual os seres humanos transformam o mundo. “Existir, humanamente, é pronunciar o mundo, é modificá-lo. O mundo pronunciado, por sua vez, se volta problematizado aos sujeitos pronunciantes, a exigir deles novo pronunciar” (Freire, 1987, p.43).

Ainda assim, como passo seguinte, é constituída, estabelecida, atravessada, envolvida pela compreensão bakhtiniana das relações dialógicas (Araújo, 2004). Ou seja, se dá a partir da relação de um discurso a outro na heterogeneidade discursiva que constitui, exemplificado no âmbito desta pesquisa, pela dinâmica presente na intersecção do discurso da IA Generativa – o discurso gerado, imitado; com o do discurso da checagem – o discurso enunciado, conhecido. Se vale, assim, das coincidências, das equivalências, das similaridades semânticas dos seus dizeres pareados, sejam constitutivas (implícitas) ou mostradas (explícitas).

Portanto, para fundamentar uma área da comunicação que se propõe a utilizar cada vez mais de modelos de IA Generativa para a sua prática – que embora tenha nascido dentro do seio jornalístico, se transborda para os demais empregadores da palavra –, se fez primordial avaliar a capacidade desses modelos, inicialmente o da líder de mercado, a OpenAI, em retornar informações qualificadas.

Uma vez que os resultados iniciais desta pesquisa tenham se mostrado favoráveis à hipótese inicial, sendo essa a capacidade de LLMs em discernir

informações, é possível avançar para novos estudos capazes de se aprofundar na definição, área de atuação e nas melhores práticas dessa modalidade.

A proposição da fundamentação do *Prompt* Jornalismo não deve ser confundida apenas com práticas de utilização de ferramentas de IA Generativa durante os processos de produção e consumo de informações. Tais práticas podem estar debaixo do guarda-chuva do *Prompt* Jornalismo, mas não as são na sua totalidade. À essas práticas, denominam-se como Promptação, que será devidamente delineada nas seções a seguir.

O Jornalismo de *Prompt* precede as técnicas. Ele ampara todo um ecossistema próprio de produção e consumo de informação. No Jornalismo impresso, há o papel. No Radiojornalismo, há as ondas sonoras. No Telejornalismo, há o audiovisual. No *Prompt* Jornalismo, há os comandos que discernem e disseminam informações.

A busca pela compreensão do *prompt* como uma área aplicada ao jornalismo passa pelo reconhecimento de que ela deva estar ancorada em princípios fundadores, considerando desde a maneira como as notícias são produzidas até a influência na percepção do público sobre as informações obtidas via *prompt* (comando) ou de modo *prompt* (rápido, ágil, assertivo), avaliando a premissa da IA Generativa em capturar a atenção e construir sentidos diante das singularidades de cada indivíduo no processo informativo.

Essa postura dialoga com a ideia bakhtiniana, mas tem como porta de entrada a freireana, a partir do entendimento de que a educação – neste trabalho internalizada pela compreensão do manejo adequado da Promptação – nunca se reduz à técnica, embora dependa dela para alcançar sua plenitude. Freire (2001, p.98) argumenta que utilizar tecnologias emergentes na educação

> pode expandir a capacidade crítica e criativa [...]. Dependendo de quem o usa, a favor de que e de quem e para quê. O homem concreto deve se instrumentar com o recurso da ciência e da tecnologia para melhor lutar pela causa de sua humanização e de sua libertação.

É por isso que a percepção dialógica do *Prompt* Jornalismo nasce freireana, a partir da instrumentalização do humano ao interagir com a máquina, mas se constitui bakhtiniana ao se valer da construção dos sentidos fruto dos produtos gerados nessas interações.

6.1 A PRÁTICA DISCURSIVA

Compreender as formações discursivas e ideológicas que perpassam a construção de sentidos em um texto é parte fundamental do trabalho jornalístico. Enunciar o mundo é o cerne do jornalismo e, portanto, ao fazer isso, traz consigo a responsabilidade por trás de cada narrativa, ressignificação, silenciamento, apagamento e interdiscursos.

Surge a necessidade de estudar como se dão as construções de sentido dentro do *Prompt* Jornalismo, uma vez que esse se projeta na mediação entre humanos e as ferramentas de IA Generativa. É certo, porém, que a ideação primária desta definição se dá muito mais atrelada à pretensa interação entre humanos-IAG. Ocorre que, diante da necessidade de uma análise discursiva cautelosa, ainda é incerto se essa noção é adequada.

Charaudeau e Maingueneau (2004) apresentam um dicionário de análise do discurso com definições para auxiliar a aplicação desta metodologia. Introduzem o termo “discurso” como uma noção filosófica clássica próxima ao *logos* grego e, dentre as possíveis definições, destaca-se a atribuída a Gardiner (apud Charaudeau; Maingueneau, 2004, p.68), sendo “a utilização, entre os homens, de signos sonoros articulados, para comunicar seus desejos e opiniões sobre as coisas”. Importante trazer a noção pecheuxtiana, discurso é efeito de sentidos.

De igual forma, os autores apresentam possíveis definições para o termo “análise do discurso” e, ao reconhecerem a instabilidade do campo, optam por uma associação simples e objetiva: analisar a relação entre texto e contexto. Toma, como objeto de estudo, “a totalidade dos enunciados de uma sociedade, apreendida na multiplicidade de seus gêneros” (Charaudeau; Maingueneau, 2004, p.46). Soma-se a essa compreensão a noção pecheuxtiana, a de que discurso

> não é apenas um ato comunicativo, uma pregação, uma performance política. Não é a simples transmissão de uma informação. As relações de linguagem são relações de sujeitos e de sentidos e seus efeitos são múltiplos. Por isso, entendemos que “discurso é efeito de sentidos entre locutores”, efeitos estes, provocados pela materialização lingüística das práticas ideológicas (Charaudeau; Maingueneau, 2004, p.46).

Diante das instâncias que constituem o discurso, como ser orientado, uma forma de ação, interativo, contextualizado, assumido e regido por normas (Charaudeau; Maingueneau, 2004), é preciso destacar que a proposição discursiva,

a qual se projeta dentro da prática do *Prompt* Jornalismo, está atrelada às construções de sentido nos sujeitos pronunciantes, resultantes da mediação destes com às IAG. Não se trata das respostas dadas pelos LLMs de modo isolado, mas sim em como o produto desta mediação implica na produção de sentidos e, portanto, nos discursos dos sujeitos constituintes que “promptam”.

6.2 A PRÁTICA DIALÓGICA

A premissa de uma prática idealizada como dialógica no Jornalismo de *Prompt* está atrelada à alusão do movimento que ocorre entre o usuário e a ferramenta de IA Generativa, em que este realiza um *prompt* e recebe de volta uma resposta, e assim sucessivamente, tal qual como num diálogo corriqueiro.

Este movimento gera um produto único, existente tão somente naquela janela criada por cada usuário, precisamente dependente, mas não tão somente, da escolha das palavras que compõem o *prompt*, paralelamente com o modelo de LLM utilizado, as técnicas de *prompting* e, até mesmo, a temperatura[11] pré-definida.

Nesta seção, diante do paralelo estabelecido, direciona-se o olhar para o cuidado com a escolha das palavras, de modo a expressarem a essência da questão formulada, reconhecendo a dimensão ontológica que permeia a construção de uma pergunta capaz de obter os resultados almejados por meio do *prompt*, delineada aqui como a égide da pergunta[12].

Portanto, se enunciar o mundo é o cerne do jornalismo, perguntar é o ato laboral primordial deste processo, ao culminar na matéria prima necessária para o surgimento de seus produtos, as respostas. E é aí que surge o caráter dialógico deste processo, constituído na produção de sentidos que advém da mediação humano-máquina. Como elaborado por Araújo (2004, p.48) ao se referir ao dialogismo de Bakhtin (1929), “Todo enunciado é dialógico. O dialogismo é o modo de funcionamento real da linguagem, é o princípio constitutivo do enunciado. Todo enunciado constitui-se a partir de outro enunciado, é uma réplica a outro enunciado”.

No entanto, como mencionado por Charaudeau e Maingueneau (2004), essa forma de troca humano-máquina pode ser caracterizada como diálogos “artificiais”

---

[11] Temperatura, neste sentido utilizado, é um parâmetro inserido no código que acessa um LLM via API com o objetivo de controlar o grau de aleatoriedade ou previsibilidade das respostas do modelo.
[12] Na mitologia grega, égide é o escudo de Zeus utilizado na batalha contra os Titãs. Aqui seu uso se faz atrelado à conotação de amparo.

ou “fabricados”, o que, se tratando de uma inteligência que se projeta como artificial, parece-se ser devidamente adequada. Segundo os autores, diálogo sempre implica identidade e diferença. Qual seria a identidade da IA Generativa? Santaella (2023) se refere ao chatbot com IAG como “papagaio estocástico”, ou seja, um imitador. Assim, não passa de um simulador do humano.

Apesar disso, essa identificação como um imitador, à priori, não demonstra ser um impeditivo à sua mediação, contanto que essa compreensão seja transparente, afinal, a inteligência contida no imitador também é artificial e, nem por isso, inadequada para a finalidade aqui proposta, como constatada no capítulo de análise. Mas ora, não seria a imitação algo muito próximo da polifonia presente nos discursos dos sujeitos bakhtinianos de que é a fonte única e original do sentido? (Bakhtin apud Araújo, 2004)

Novamente, reitera-se. Não se trata da mera resposta gerada pela IAG, mas em como o produto desta mediação implica na produção de sentidos e, portanto, nos discursos dos sujeitos que “promptam”. É um movimento cíclico. É a conexão direta da noção de interdiscurso, ou seja, a memória do que foi dito, em que se retoma os saberes partilhados, os lugares-comuns ainda sem que se perceba.

Nas relações dialógicas, raramente é possível saber exatamente a origem de uma ideia. O dito é, por vezes, um já-dito. Uma resposta a um enunciado proferido anteriormente. E o sujeito destes discursos

> se caracteriza por dois esquecimentos ou duas ilusões: 1) o sujeito tem a ilusão de que é uno, a origem do sentido; 2) o sujeito tem a ilusão de que o que diz tem apenas um significado, isto é, todo interlocutor captará suas intenções e suas mensagens (Bakhtin apud Araújo, 2004, p.31).

Mas não é assim que o sujeito bakhtinano se constitui. Como dito e, discursivamente, re-dito, ele é polifônico, se dá pelos atravessamentos de um discurso no outro dentro da dinâmica do processo enunciativo. Sua existência só é possível porque a do outro existe. O discurso do sujeito é composto por várias outras vozes que “se cruzam, se complementam, se relativizam ou se contradizem” (Araújo, 2004, p.31). A sua única originalidade se dá na circunstância, na situação e na forma como fora enunciada, nas novas condições em que se deu sua produção.

Produção essa que é caracterizada pela heterogeneidade, que pode tanto ser **constitutiva**, marcada pela presença velada de outros discursos no inconsciente do

texto, no interdiscurso; quanto **mostrada**, tendo a presença de outras vozes evidenciada de modo explícito. Na Promptação realizada nesta pesquisa, a heterogeneidade é percebida tanto nos conteúdos da agência de checagem, quanto nos gerados pela IAG.

O conteúdo da agência se mostra como o discurso conhecido, checado, verificado, a verdade validada, o peso que equilibra e criva o valor sobre a balança. Já o conteúdo gerado pela IAG, pelo LLM, se caracteriza pelo discurso imitante, simulado, processado, calculado, aquele que se passa por algo que não é, como definido por Santaella, um simulador do humano. É um discurso gerado, não criado. Por se constituir enquanto um modelo estatístico, probabilístico, a IAG não pensa o texto, mas o gera e o remonta a partir do encontro da recuperação de discursos outros que já existiam fora dele, tanto por base no banco de dados utilizado em seu treinamento, quanto pela fruição, acarretada pela Promptação, que resulta na produção de sentidos a partir do produto gerado, do discurso promptado, interrogado, caracterizado pelo recorte temporal e espacial próprio da interação humano-máquina existente tão somente naquele momento específico.

O discurso promptado está diante de uma teia discursiva dialógica que lhe é exterior (Figura 7), constituída por uma memória gigantesca com tudo o que precisou existir até aquele momento para que ele se tornasse passível de existir. Promptar, o ato do ato, é a instância em que se puxa um fio desta enorme teia (exterior) em direção à interface em que o enunciador, o indivíduo, o sujeito realiza em busca de construir sentidos específicos (interior).

Figura 7 - Movimento de *promptar*, puxar um fio da teia discursiva dialógica

Fonte: Elaborada em conjunto com o modelo gpt-5.2, em fevereiro de 2026

Essa busca fica ainda mais evidente quando se lida com dados semânticos, melhor abordados nas seções seguintes, mas que em suma se caracteriza como o mapa da informação que indica para a IA onde mora o sentido. O sentido se revela quando uma palavra, conceito, expressão se mostra reiteradas vezes, em diferentes lugares e formatos, mas ainda com um significado forte. É, por natureza, polifônico. É discurso. Discurso é o efeito dos sentidos.

6.3 A PRÁTICA TÉCNICA

Os aspectos desta dimensão buscam testar, verificar, estipular e sistematizar, objetivamente, os potenciais da partícula que nomeia o campo jornalístico emergente proposto neste trabalho, o *prompt*, comando que discerne e dissemina informações.

Diante das análises realizadas, foi possível identificar que técnicas de *prompting*, que é o processo de fornecer uma entrada (*prompt*) a um modelo de IAG para gerar “uma” saída, de caráter mais discricionário, exemplificativas e com

cadeias estruturadas de “pensamento”, permitiram maior proveito da atividade solicitada.

A partir desse processo, evidenciam-se as potencialidades da utilização das ferramentas de IA Generativa para a produção e consumo de informação qualificada, contribuindo com o ecossistema da informação.

Ainda assim, um dos desafios técnicos encontrados, a possibilidade dos LLMs errarem, gerar "alucinações", necessita de um olhar cauteloso, ainda que com o exercício correto da Promptação possa mitigar esses riscos, ao induzir maior criticidade, rigor e dados semânticos na condução prática com os modelos.

É necessário ir além. Compreender o funcionamento intrínseco das IAs Generativas, em especial dos LLMs. Como se dão seus treinamentos? Como fazem o que fazem? A que custo? Como construir um ecossistema em que a relação humano-máquina seja mutualística e não parasitária?

A princípio, essas perguntas podem parecer extrapolar a prática técnica, mas são essenciais para que se estabeleça um guia de boas práticas em busca de um campo saudável, eficiente e sustentável.

7 PROMPTAÇÃO: UM ENSAIO DA APLICAÇÃO LEXICAL DO JORNALISMO DE PROMPT

Todo o trajeto percorrido até aqui por esta pesquisa foi um passo atrás. A IA chegou com a promessa de otimizar o tempo, de tornar mais rápido a feitura, os processos, as tomadas de decisões e, se possível, decidir pelos sujeitos. Era, como ainda é, preciso pisar no freio e analisar o veículo em que estamos prestes a fazer uma longa viagem antes de dar a partida e pegar a estrada.

É por isso que esta pesquisa se constitui em uma simples pergunta: **pode** a IAG discernir informações? Dentro das circunstâncias avaliadas, ficou evidenciado que sim. Este veículo — tão jovem e futurista, mas que já demanda um combustível colossal (G1, 2025) tal como algo ultrapassado — é capaz de trafegar pelas vias informacionais e identificar atalhos duvidosos, buracos camuflados, acidentes na pista, fazer ultrapassagens e desviar de transportes desinformativos. Mas o que as evidências também sugerem é que este veículo, com a devida licença que a metáfora permite, não é autônomo. Aliás, ele não deveria ser. Se assim for, suas habilidades reduzem drasticamente e pode, inclusive, contribuir para os acidentes que se deseja evitar. Este meio de transporte precisa de uma pessoa que o saiba dirigir bem. Precisa, óbvio e primeiramente, de políticas públicas que nortearão desde a sua fabricação, como uma montadora qualificada e sustentável, até toda a legislação de trânsito ao redor dele. Mas a esta pesquisa competiu falar do veículo e, neste capítulo, de práticas promovidas por pessoas motoristas. O veículo é a IA Generativa. As pessoas que o dirigem são os usuários. E estas precisam ser habilitadas para que possam se valer do uso qualificado deste meio.

É hora, portanto, de avançar o passo retrocedido e refazer a pergunta, caminhando agora para: se **pode** a IAG discernir informações, **como**? E este como, dentro do Jornalismo de Prompt, passa tanto pela prática técnica atrelada ao funcionamento dos modelos, mas passa sobretudo pela estruturação lexical, envolta pela prática discursiva e dialógica, que os usuários podem se valer para obter tais resultados. E é disso que este ensaio se trata.

7.1 DA PALAVRA AO PROMPT

A pá lavra a terra. E prepara o solo para o que há de germinar. A palavra aterra. E planta os dizeres que enunciam a Terra. A palavra é fértil, é desejo, é vida que escorre pela boca, pelos olhos, pelo pelo, pelo lado, pelo lápis, pelo papel, pela tinta, pela prensa, pela máquina de datilografar, pelas teclas, pelas tantas, pelo microfone, pela câmera e pela câmara, pelo celular, pelo dito, pelo cujo. A palavra, como todo trabalhador, tem o seu emprego, o seu lugar no texto, com o texto.

Contexto. Tudo se resume a ele. Segundo Rodrigues (2025), a palavra sem contexto é só ruído. É o que permite que as ferramentas generativas sejam precisas e não sejam levadas a interpretações equivocadas pela omissão do próprio usuário. A IAG trabalha com janela de contexto, que é o trecho que o modelo analisa para prever, de modo estatístico, a próxima sequência de palavras. Assim, se o contexto não estiver claro, tomará decisões de forma generalista.

Essa observação do autor, ao dizer que as ferramentas de IA seriam levadas ao equívoco pela omissão do próprio usuário, revela uma lógica de responsabilização do usuário ao invés da própria plataforma. É como se ele fosse responsável até pelas ramificações do que não disse, pelo não-dito, diante da máquina. Mas isso, embora preocupante, soa familiar com a velha máxima da comunicação ao dizer que comunicação não é o que se diz, é o que o outro entende.

Intenção. É o que precede a palavra. Se antes os mecanismos de buscas entregavam apenas as páginas com o conteúdo que melhor correspondia ao solicitado diretamente pelo usuário, segundo Rodrigues (2025) hoje a IAG busca entender não apenas o que está escrito ou dito explicitamente, mas aquilo que mais se aproxima da intenção do usuário ao dizer o que disse. “Filme sobre sonho com peão.” Antes, para serem ranqueadas, deveriam constar nas páginas expressamente essa expressão. Hoje, os buscadores compreendem, mesmo que a expressão não seja exatamente a mesma, que se trata sobre ”A Origem”, dirigido por Christopher Nolan (2010).

É preciso pensar no que o usuário quer encontrar ao buscar tal coisa. E, com isso, construir todo o mapa ao redor desta coisa. É o que Rodrigues (2025) chama de dado semântico, o mapa da informação que indica para a IA onde mora o sentido. Ele nasce do público e se revela quando um mesmo termo/expressão

aparece com frequência em diferentes lugares e momentos, mas mantendo o significado forte, como o famoso “filme sobre sonho com peão”.

Percebe-se, então, que a interação com os modelos generativos se inicia antes mesmo que ela aconteça em um *chatbot*. Todo conteúdo disponível na Internet, e cada vez mais parte dos que não estão, inclusive, são utilizados para treinar os LLMs. Desta forma, é preciso ter em mente que a lógica midiática agora gira em torno não da produção ser diretamente lida ou consumida pelo usuário, mas sim em ser usada pela IA para gerar seus *outputs*. É um novo paradigma.

A Promptação é justamente a intersecção destes momentos. É o pensar o conteúdo antes de sua concepção, de modo que, ao alimentar os modelos, permita que os usuários, se valendo das técnicas adequadas de prompting, extraiam resultados mais qualificados. É um ciclo vicioso extrativista. Haja Terra. Haja pá, lavra!

Nas subseções a seguir, é possível ver maneiras que os usuários, sobretudo os profissionais da palavra, podem se valer para extrair tais resultados, segundo Rodrigues (2025).

7.1.1 Pré-prompt

Tudo começa com três letras, GEO (Generative Engine Optimization), o sucessor do SEO (Search Engine Optimization). Enquanto o SEO foca em ranquear as páginas que melhor correspondem ao *input* do usuário, o GEO se vale dos trechos destas páginas com maior equivalência para incorporar o corpus que construirá o *output* devolvido ao usuário. O quadro 11 a seguir indica os tipos de conteúdos que a IAG prioriza ao buscar o corpus.

Quadro 11 - Tipos de conteúdos priorizados pela IAG

Têm estrutura clara	Títulos descritivos, subtítulos com função real, listas, bullet points e parágrafos curtos que facilitam a identificação de tópicos e a extração de trechos.

São semanticamente ricos:	Repetem termos importantes com variações contextuais, criando redes de significado que ajudam o modelo a entender a função e o tema central.
Têm linguagem clara e factual	Frases simples, dados comprováveis, explicações passo a passo e ausência de ambiguidade favorecem a compreensão da IA.

Fonte: Enquadramento próprio a partir de Rodrigues (2025)

Produzir esses conteúdos não está relacionado a performance de um site, mas sim na criação de um mapa para que a IAG os encontre e os utilize nas respostas geradas. Para isso, os LLMs realizam três etapas na busca pela resposta mais adequada. Primeiro, pré-selecionam conteúdos considerados confiáveis, relevantes e atualizados (Ex: ranqueamento do Google). Em seguida, extraem e segmentam trechos mais correspondentes para só então gerar a resposta de fato com base nesses conteúdos, que podem ou não citar de forma direta as fontes.

“Você não escreve mais para que o texto seja exibido, escreve para que a informação seja usada” (Rodrigues, p.63, 2025). Apesar do autor mencionar que esta forma de escrita não é antagônica a aquela feita pensando nas pessoas leitoras, chama a atenção a relação de objetificação causada neste contexto. Ao criar o contexto para a IAG, aliás, o produto criado por um profissional da palavra some, dando lugar a um subproduto gerado com frações de tantos outros. É a naturalização de um sistema que se vale do fruto do trabalho de tantos criadores sem a garantia de que serão devidamente creditados, remunerados.

Uma vez que o conteúdo tenha sido produzido com o foco em ser identificado pela IAG, o próximo passo se dá pela condução de testes de GEO nos próprios ambientes generativos por meio dos *prompts*, como observado na subseção a seguir.

7.1.2 Prompting

Abra um *chatbot* e busque por algo relacionado a um conteúdo seu. O resultado gerado pode ser fruto tanto da construção do conteúdo, como mencionado

na subseção anterior, quanto das técnicas de *prompting* utilizadas na busca ou, ainda, de ambas.

Como esmiuçado no subcapítulo 2.1, homônimo deste, existem várias formas de se interagir com um chatbot e, para as formas qualificadas destas interações, atribui-se o nome de técnicas de *prompting* (Schulhoff et al., 2024). Há dezenas de técnicas já documentadas, mas de forma geral podem ser resumidas em genéricas (*zero-shot*) e específicas (*few-shot*). A menos que se busque uma resposta genérica, como o próprio nome já diz, as técnicas que se valem de mais detalhes e, consequentemente, mais contexto, tendem a trazer resultados mais qualificados.

É isso que foi observado durante a execução desta pesquisa. Na prática, a metodologia de caráter experimental, executada na etapa de avaliação semântica (Capítulo 4 - Aspectos Metodológicos), é um grande exemplo de promptação, uma vez que une técnicas de *prompting* com trechos ricos em contexto de conteúdos prévios produzidos por agência de checagem, o *lead*.

Rememore, no Quadro 7 apresentado anteriormente, as diferenças entre as estruturas das técnicas de prompting utilizadas para realizar a análise semântica. No *zero-shot*, percebe-se a instrução dada de forma mais ampla. No *few-shot-cot*, já há um detalhamento maior através de exemplos de como a avaliação deveria ser feita, bem como a especificação de quais critérios o modelo deveria seguir para atribuir a pontuação. Rememorando o Gráfico 6, que mostra a classificação dos pares de entradas por escala de similaridade, é possível ver que o desempenho das técnicas *few-shot*, tanto pelo modelo da DeepSeek, quanto pelo modelo da OpenAI, teve a melhor desenvoltura, sobretudo nas faixas com a menor convergência semântica.

Essa experimentação corrobora com a noção de que quanto maior o contexto, melhor os resultados. E isso passa também por outro ponto importante, a chave para reduzir as "alucinações". Aliás, o que se convencionou chamar de "alucinações" são respostas factualmente incorretas com aparência de verdade. Na verdade, este termo deveria ser revisto. Respostas factualmente incorretas com aparência de verdade são, na verdade, mentiras. Ou, de forma polida, erros gerados devido a ausência de lastro em dados semânticos, ainda que esses erros não sejam intencionais, mas sim fruto das operações estatísticas dos modelos em busca da associação mais provável.

Ora, se um aluno em uma prova marca uma questão de múltipla escolha com base em suas associações neurais mais prováveis e, para a sua infelicidade, esta

não é a correta, a ele não será dito que alucinou, mas sim que cometeu um erro que lhe custou alguns pontos. Tal como a IAG, poderá o aluno aprimorar seu banco de dados em seu próximo treinamento e ir em busca de maior precisão da próxima vez.

De volta aos erros da IAG, a semântica, segundo Rodrigues (2025), é o antídoto. Para isso a palavra precisa estar ancorada em significado verificável. “Quanto mais uma IA consegue ver a estrutura semântica por trás dos termos, mais ela se aproxima de respostas auditáveis e consistentes” (p. 49). E, para isso, há duas formas principais. A primeira é o já bastante abordado “contexto”. Quanto mais informação, menor a margem de erro do modelo, pois não precisará deduzir dados faltantes. A segunda, de caráter mais técnico, é o uso de bases enriquecidas com dados semânticos previamente lapidados, a qual se denomina RAG, *Retrieval Augmented Generation*. É nesta segunda que possíveis desdobramentos desta pesquisa aplicada poderiam se valer.

8 CONSIDERAÇÕES DECORRENTES

A presente pesquisa buscou avaliar se os Modelos de Linguagem de Larga Escala (LLMs), enquanto expressão das Inteligências Artificiais Generativas (IAG), são capazes de produzir informações qualificadas — ou seja, conteúdos factualmente corretos, semanticamente coerentes e alinhados ao que se espera de uma verificação responsável em um ecossistema informativo saudável. A hipótese inicial, de que tais modelos seriam capazes de discernir informações, mostrou-se corroborada pelos dados obtidos, ainda que com ressalvas, como: os cenários em que o LLM teve baixo desempenho, transitando entre os três cenários estipulados, ou ainda as limitações contextuais e linguísticas, pelo fato da pesquisa ter focado em uma única agência de checagem, um único idioma e um único modelo checador. A generalização para mais agências, idiomas e modelos poderá ajudar a avançar no entendimento dessas questões.

Com base na criação de um conjunto de dados original — composto por 1.200 pares de checagens entre uma agência jornalística certificada (Aos Fatos) e o modelo GPT-4o da OpenAI —, foi possível realizar uma análise robusta a partir de medidas de similaridade semântica, clusterizações e testes estatísticos. Os resultados indicaram que mais de 90% das entradas apresentaram boa ou quase total correspondência de sentido entre as respostas da IA e o conteúdo da agência, tanto na média quanto na mediana, sugerindo que os LLMs, quando submetidos a *prompts* bem elaborados, são capazes de gerar respostas altamente compatíveis com padrões de verificação jornalística.

No entanto, essa capacidade não é absoluta nem uniforme. A pesquisa também demonstrou que o desempenho dos LLMs varia consideravelmente de acordo com a técnica de *prompting* adotada. As abordagens *few-shot* com cadeia de pensamento (CoT), sobretudo no modelo DeepSeek, demonstraram maior rigor crítico ao avaliar divergências mais sutis entre conteúdos, ainda que com variações nas métricas centrais. Isso revela um aspecto fundamental: o modo como se formula a interação com o modelo tem impacto direto na qualidade da resposta, o que reforça a ideia de que o jornalismo de *prompt* é, sobretudo, um campo de prática discursiva, dialógica e técnica que precisa ser compreendido e sistematizado.

Do ponto de vista metodológico, a utilização de técnicas de clusterização e testes de hipóteses não-paramétricos revelou padrões antes ocultos, como a

existência de um grupo minoritário de respostas com baixa similaridade, cuja análise aprofundada indicou limitações importantes do modelo, como a incapacidade de interpretar referências à elementos audiovisuais descontextualizados, recusas de resposta em temas sensíveis e erros factuais em contextos específicos. Essas ocorrências, embora marginais, apontam a necessidade de aprimoramento contínuo tanto dos modelos quanto da curadoria das interações feitas por usuários.

A partir desses achados, torna-se viável defender a proposição da emergência de um novo campo, o *Prompt* Jornalismo, ancorado em princípios éticos, metodológicos e tecnológicos próprios. Este campo busca articular as capacidades dos LLMs à mediação jornalística, abrindo espaço para novos fluxos de produção, circulação e consumo de informações verificadas em tempo real — sem, contudo, abdicar da criticidade, da curadoria humana e do compromisso com o interesse público.

Esta pesquisa também fundamentou a Promptação como aplicação prática do Jornalismo de Prompt, analisando a mediação técnica e discursiva entre humanos e Inteligência Artificial Generativa na verificação de fatos. Abordou-se ainda o conceito de Generative Engine Optimization (GEO) e a estruturação lexical dos comandos. Conclui-se que a aplicação produtiva-adequada-abrangente da prática do Jornalismo de *Prompt*, a Promptação, contribui para mitigar "alucinações"/erros e garantir informações qualificadas no ecossistema digital.

Como caminhos para pesquisas futuras, vislumbra-se: **a)** o aprimoramento das perguntas utilizadas como input, aproximando-as da linguagem cotidiana por meio de técnicas de SEO/GEO e estudos sobre comportamento do usuário; **b)** a expansão da base de dados, incluindo outras agências de checagem, idiomas e contextos culturais diversos; **c)** a comparação entre diferentes modelos de LLMs para além da OpenAI, inclusive os de código aberto, ampliando a representatividade do campo experimental; **d)** e o aprofundamento conceitual do Jornalismo de *Prompt*, avançando em seus fundamentos epistemológicos, limites e possibilidades, práticas e valores, de modo a consolidá-lo como campo de estudo e atuação profissional.

Em suma, esta pesquisa constatou que sim, modelos de IA Generativa são capazes de discernir informações — ainda que com margem para aprimoramento e de modo restrito ao retrato do período analisado, o que deixa o indicativo para a necessidade de monitoramento constante à medida que os modelos são atualizados ou que novos vão sendo inseridos no mercado — apresentando embasamento para

a fundamentação do campo emergente do jornalismo, o *Prompt* Jornalismo. O Jornalismo de *Prompt*, aqui apenas delineado, pode ser entendido como uma resposta possível aos dilemas da desinformação, ao mesmo tempo em que se projeta como uma oportunidade de renovação da prática jornalística frente aos desafios do século XXI.

REFERÊNCIAS

ADORNO, Theodor W.; HORKHEIMER, Max. **Dialética do esclarecimento: fragmentos filosóficos**. Trad. Guido Antônio de Almeida. Rio de Janeiro: Zahar, 1985.

ANDRION, Roseli. **Blockchain: a tecnologia que pode ir muito além da bitcoin**. Olhar Digital. Julho de 2019. Disponível em: https://olhardigital.com.br/2019/07/17/seguranca/blockchain-a-tecnologia-que-pode-ir-muito-alem-da-bitcoin/ Acesso em: 20 de março de 2021.

A ORIGEM. Direção: Christopher Nolan. Estados Unidos: Warner Bros. Pictures, 2010. Filme (148 min).

ARAÚJO, Marcelo Marques. **Sincronia da memória acadêmica**. 2004. Dissertação (Mestrado em Linguística) - Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2004.

ARAÚJO, Marcelo Marques. **Comunicação, Língua e Discurso: Uma Análise Terminológica Discursiva de um Dicionário de Especialidade**. Tese (Doutorado em Letras e Comunicação) - Universidade Presbiteriana Mackenzie, São Paulo, 2011.

ARIELLE, B.; MANOVICH, L. **Data Studies in the Age of Big Data**. Cambridge: MIT Press, 2022.

AVAAZ. **O Brasil está sofrendo uma infodemia de Covid-19**. Maio de 2020. Disponível em: https://secure.avaaz.org/campaign/po/brasil_infodemia_coronavirus/. Acesso em: 20 mar. 2021.

BAKHTIN, M. (1929). **Marxismo e filosofia da linguagem**. (Tradução de M. Lahud e Y. F. Vieira), São Paulo: Hucitec, 1979. 196p.

BELCHIOR. **Apenas um rapaz latino-americano**. In: BELCHIOR. Alucinação. São Paulo: Polygram, 1976. 1 disco sonoro (LP), faixa 1

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidente da República, [2016]. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em 13 out. 2023.

BRASIL. **Relatório Final da CPI da Pandemia**. Brasília: Senado Federal, 26 de outubro de 2021. Disponível em: https://legis.senado.leg.br/comissoes/comissao?codcol=2441 Acesso em: 28 de outubro de 2021.

CAO, Han et al. **Are large language models good fact checkers: a preliminary study**. Beijing: Institute of Information Engineering, Chinese Academy of Sciences; University of Chinese Academy of Sciences, 2023. Disponível em: https://arxiv.org/abs/2311.17355. Acesso em: 19 mar. 2025.

CARAMANCION, Kevin Matthe. **News verifiers showdown: a comparative performance evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in news fact-checking**. Menomonie, WI: University of Wisconsin-Stout, 2023. Disponível em: https://www.kevincaramancion.com. Acesso em: 19 mar. 2025. https://doi.org/10.1109/FNWF58287.2023.10520446

CHARAUDEAU, Patrick; MAINGUENEAU, Dominique. **Dicionário de análise do discurso**. São Paulo: Contexto, 2004. 376 p.

COVINGTON, P.; ADAMS, J.; SARGIN, E. **Deep Neural Networks for YouTube Recommendations**. In: Proceedings of the 10th ACM Conference on Recommender Systems - RecSys'16, Boston, MA, USA, 2016. p. 191-198. https://doi.org/10.1145/2959100.2959190

Dourado, Tatiana Maria Silva Galvão. **Fake News na eleição presidencial de 2018 no Brasil**. 2020, 308 p. Tese (doutorado) Universidade Federal da Bahia, Faculdade de Comunicação, 2020.

EDWARDS, John. **Bitcoin's Price History**. Investopedia. Disponível em: https://www.investopedia.com/articles/forex/121815/bitcoins-price-history.asp. Acesso em: 21 de março de 2021.

EXAME. **Investimentos globais em IA devem atingir 1,5 trilhão de dólares em 2025, projeta Gartner**. Exame - Negócios, 24 set. 2025. Disponível em: https://exame.com/negocios/investimentos-globais-em-ia-devem-atingir-us-15-trilhao-em-2025-projeta-gartner/. Acesso em: 20 de janeiro de 2026.

FIRST PAGE SAGE. **Top generative AI chatbots**. First Page Sage, 25 mar. 2025. Disponível em: https://firstpagesage.com/reports/top-generative-ai-chatbots/. Acesso em: 25 mar. 2025.

FISHER, Max. **A máquina do caos: como as redes sociais reprogramaram nossas mentes e o mundo**. São Paulo: Todavia, 2023.

FREIRE, Paulo. **Pedagogia do oprimido**. 17. ed. Rio de Janeiro: Paz e Terra, 1987.

FREIRE, Paulo. **Pedagogia da esperança: um reencontro com a Pedagogia do oprimido**. 5. ed. São Paulo: Paz e Terra, 1994.

FREIRE, Paulo. **A Educação na Cidade**. 5. ed. São Paulo: Cortez, 2001.

G1. **Inteligência artificial: tecnologia demanda geração colossal de energia elétrica, entenda**. Jornal Nacional, Rio de Janeiro, 24 jan. 2025. Disponível em: https://g1.globo.com/jornal-nacional/noticia/2025/01/24/inteligencia-artificial-tecnologia-demanda-geracao-colossal-de-energia-eletrica-entenda.ghtm. Acesso em: 27 jan. 2026.

GROSSI, Angela. **As tecnologias e a Comunicação da Ciência como objetos de fronteiras nas pesquisas**. In: SEMINÁRIO INTERNACIONAL INTERDISCIPLINAR

EM TECNOLOGIAS, COMUNICAÇÃO E EDUCAÇÃO, 3., 2024, Uberlândia. Comunicação oral. Uberlândia, 17 abr. 2024.

HAAS, Guilherme. **1 ano de ChatGPT: 15 curiosidades sobre o chatbot da OpenAI**. Canaltech, 30 nov. 2023. Disponível em: https://canaltech.com.br/inteligencia-artificial/aniversario-chatgpt-curiosidades-sobre-o-chatbot-da-openai/. Acesso em: 25 jun. 2024.

IBM. **O que é Inteligência Artificial (IA)?**, 09 ago. 2024. Disponível em: https://www.ibm.com/br-pt/topics/artificial-intelligence. Acesso em: 10 nov. 2024.

IFCN - **INTERNATIONAL FACT-CHECKING NETWORK**. Disponível em: https://www.poynter.org/ifcn/. Acesso em: 15 mar. 2025.

JAMES, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert. **An introduction to statistical learning: with applications in R**. 2. ed. New York: Springer, 2023. https://doi.org/10.1007/978-3-031-38747-0

KAUFMAN, Dora. **Desmistificando a inteligência artificial**. Belo Horizonte: Editora Autêntica, 2022.

LEE, Kai-Fu. **Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos comunicamos e vivemos**. Tradução de Marcelo Barbão. 1. ed. Rio de Janeiro: Globo Livros, 2019.

LEE, Kai-Fu; QIUFAN, Chen. **2041: Como a inteligência artificial vai mudar sua vida nas próximas décadas**. Tradução de Isadora Sinay. Rio de Janeiro: Globo Livros, 2021. 480 p.

LEPORACE, Camila De Paoli. **Algoritmosfera: a cognição humana e a inteligência artificial** / Camila De Paoli Leporace. - Rio de Janeiro: Ed. PUC-Rio; São Paulo: Hucitec, c2024.

MALDONADO, Maurílio. **Separação dos poderes e sistemas de freios e contrapesos: desenvolvimento no Estado brasileiro**. Revista Jurídica "9 de julho", São Paulo, p. 195-214, 09 de julho de 2003.

MANNING, Christopher D.; SCHÜTZE, Hinrich. **Foundations of statistical natural language processing**. Cambridge: MIT Press, 1999.

MARCONI, Francesco. **Newsmakers: Artificial Intelligence and the Future of Journalism**. Columbia University Press, 2020. https://doi.org/10.7312/marc19136

OBERCOM. **Anuário da Comunicação 2021**. Lisboa: Obercom, jul. 2022. 272 p. Disponível em: https:https://www.obercom.pt/anuario-da-comunicacao-2021/. Acesso em: 18 nov. 2024.

OLIVEIRA, Giulia Cristina Rodrigues de. OLIVEIRA, Natália Soares de. **Saúde e Fake News: o impacto das notícias falsas no comportamento da população em**

meio à pandemia da COVID-19. Revista Conecte-se!, Volume 4, Número 8, PUC Minas, Belo Horizonte, 2020.

PINHEIRO, Rose Mara. **Educomunicação nos centros de pesquisa do país: um mapeamento da produção acadêmica com ênfase à contribuição da ECA/USP na construção do campo**. 2013. 224 p. Tese de Doutorado - São Paulo: ECA/USP, 2013.

QUELLE, Dorian; BOVET, Alexandre. T**he perils & promises of fact-checking with large language models**. Zurich: University of Zurich, 2023. Disponível em: https://arxiv.org/abs/2310.13549. Acesso em: 19 mar. 2025.

RODRIGUES, Bruno. **Prompt, Palavra, Ação: a escrita digital em ambientes generativos**. 1. ed. São Paulo: Editora Brauer, 2025.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. **Learning representations by back-propagating errors**. Nature, v. 323, p. 533-536, 1986. https://doi.org/10.1038/323533a0

SAFATLE, Vladimir. **O circuito dos afetos: corpos políticos, desamparo e o fim do indivíduo**. Belo Horizonte: Autêntica, 2015. 360 p.

SANTAELLA, Lúcia. **Há como deter a invasão do ChatGPT?** 1. ed. São Paulo: Estação das Letras e Cores, 2023.

SCHULHOFF, Sander; ILIE, Michael; BALEPUR, Nishant; KAHADZE, Konstantine; LIU, Amanda; SI, Chenglei; LI, Yinheng; GUPTA, Aayush; HAN, HyoJung; SCHULHOFF, Sevien; DULEPET, Pranav Sandeep; VIDYADHARA, Saurav; KI, Dayeon; AGRAWAL, Sweta; PHAM, Chau; KROIZ, Gerson; LI, Feileen; TAO, Hudson; SRIVASTAVA, Ashay; DA COSTA, Hevander; GUPTA, Saloni; ROGERS, Megan L.; GONCEARENCO, Inna; SARLI, Giuseppe; GALYNKER, Igor; PESKOFF, Denis; CARPUAT, Marine; WHITE, Jules; ANADKAT, Shyamal; HOYLE, Alexander; RESNIK, Philip. The prompt report: a systematic survey of prompting techniques. arXiv preprint arXiv:2406.06608, 2024. https://doi.org/10.48550/arXiv.2406.06608

SOUSA, Diego. **62% dos brasileiros não sabem reconhecer fake news, diz pesquisa**. Canaltech. Disponível em: https://canaltech.com.br/seguranca/brasileiros-nao-sabem-reconhecer-fake-news-diz-pesquisa-160415/. Acesso em: 10 mar. 2025.

SOUSA, Natália. Recomeços. **Para Dar Nome às Coisas** [podcast]. Produzido por Natália Sousa. 12 set. 2023. Disponível em: https://open.spotify.com/show/7g6BfZvLNQjrj68MNXyDqf?si=cc3d216ac8124caa. Acesso em: 01 nov. 2024.

SPRENT, Peter; SMEETON, Nigel C. **Applied Nonparametric Statistical Methods**. 4. ed. Boca Raton: Chapman & Hall/CRC, 2007.

TUNHOLI, Murilo. **Brasil é o 4º país que mais usa ChatGPT no mundo; conheça o top 10**. Gizmodo Brasil, 11 mar. 2024. Disponível em: https://gizmodo.uol.com.br/brasil-e-o-4o-pais-que-mais-usa-chatgpt-no-mundo-conheca-o-top-10/. Acesso em: 26 jun. 2024.

VASWANI, Ashish et al. **Attention is all you need**. Advances in Neural Information Processing Systems, v. 30, 2017.

VIANNA, Rodolfo. **Avanço das big techs e crise no modelo de negócio do jornalismo**. Le Monde Diplomatique Brasil, 25 abr. 2024. Disponível em: https://diplomatique.org.br/big-techs-crise-modelo-de-negocio-jornalismo/. Acesso em: 19 abr. 2025.

VOSOUGHI, Soroush; ROY, Deb; ARAL, Sinan. **The spread of true and false news online**. Science, v. 359, n. 6380, p. 1146-1151, 9 mar. 2018. DOI: 10.1126/science.aap9559. Disponível em: https://www.science.org/doi/10.1126/science.aap9559. Acesso em: 10 out. 2024. https://doi.org/10.1126/science.aap9559

WARDLE, Claire; DERAKHSHAN, Hossein. **Desordem informacional: para um quadro interdisciplinar de investigação e elaboração de políticas públicas**. Tradução de Pedro Caetano Filho e Abilio Rodrigues. Campinas: UNICAMP, Centro de Lógica, Epistemologia e História da Ciência, 2023.

WICKHAM, Hadley; ÇETİNKAYA-RUNDEL, Mine; GROLEMUND, Garrett. **R for data science: import, tidy, transform, visualize, and model data**. 2. ed. Sebastopol: O'Reilly Media, 2023. Disponível em: https://r4ds.hadley.nz/. Acesso em: 6 mar. 2025.

WUKOSON, George; FORTUNA, Joey. **The predominant use of high-authority commercial web publisher content to train leading LLMs**. SSRN, 2024. Disponível em: https://ssrn.com/abstract=5009668. Acesso em: 19 abr. 2025. https://doi.org/10.2139/ssrn.5009668

APÊNDICES

APÊNDICE A - CÓDIGO PYTHON UTILIZADO PARA TRANSFORMAR OS TÍTULOS DAS CHECAGEM DAS AGÊNCIAS EM *INPUTS*

```
'''
INTERAÇÃO GPT4O
CUSTO: $10.4
'''

import pandas as pd
import openai

# Configurar a chave da API da OpenAI
openai.api_key = "XXX"

# Ler o arquivo CSV
df = pd.read_csv("/content/noticias_1200_dated.csv")

titulos = df["titulo"].tolist()

# Função para reformular títulos
def reformular_titulo(titulo):
    prompt = f"""
     Dado um título de notícia, reescreva-o no formato de uma pergunta objetiva e afirmativa.
     O título original pode conter negações ou ser ambíguo, então reformule-o para uma estrutura clara de pergunta.
    Preserve o sentido original da frase, mas torne-a direta e questionadora.

    **Exemplos**:
     1. **Título original:** Não é verdade que o governo vai taxar os produtores de café.
            **Título modificado:** O governo vai taxar os produtores de café?

     2. **Título original:** Não é recente vídeo que mostra bloqueio de caminhoneiros.
        **Título modificado:** O vídeo que mostra bloqueio de caminhoneiros é recente?

    Agora, reformule este título:

    **Título original:** {titulo}
    **Título modificado:**"""

        # Updated call to the chat completion API using client.chat.completions.create()
    response = client.chat.completions.create(
```

```
        model="gpt-4",
            messages=[{"role": "system", "content": "Você é um
assistente útil."},
                      {"role": "user", "content": prompt}]
    )
    #Extract content from response
    return response.choices[0].message.content.strip()


df_test = df.copy()
df_test["titulo_modificado"]                                    =
df_test["titulo"].apply(reformular_titulo)

# Exibir o resultado
#  Instead  of  using  ace_tools,  display  the  DataFrame  using
pandas' display function
display(df_test)
```

APÊNDICE B - CÓDIGO PYTHON UTILIZADO PARA CARREGAR O DATAFRAME COM O DATASET CRIADO COM OS 1200 PARES DE ENTRADAS[13]

```
from google.colab import drive
drive.mount('/content/drive')

import pandas as pd

file_path                                                    =
'/content/drive/MyDrive/ColabGPT/dataset_completo_27_03_2025.c
sv'
df = pd.read_csv(file_path, sep=';', encoding='utf-8')

print(df.head())

Drive already mounted at /content/drive; to attempt to
forcibly    remount,    call    drive.mount("/content/drive",
force_remount=True).
                                      titulo_checagem  \
0  É falso que Tim deixará de operar em regiões d...
1  Trump ignora vantagem comercial e barreiras am...
2  Bens de Gilmar Mendes em Portugal não foram bl...
3  É falso que parlamentares de oposição apagaram...
4  Vídeo de homem mergulhando em trecho do rio Sã...

                                    primeiro_paragrafo_agencia
data_checagem  \
0    Não é verdade que a operadora de telefonia Tim...
07/03/2025
1     Em discurso no Congresso na última terça-feira...
07/03/2025
2    Não é verdade que o ministro do STF (Supremo T...
07/03/2025
3    Não é verdade que parlamentares de oposição ap...
06/03/2025
4    Não foi gravado durante o governo de Jair Bols...
06/03/2025

                                     titulo_modificado  \
0  Tim deixará de operar em regiões dominadas por...
1  Trump está ignorando a vantagem comercial e as...
2  Os bens de Gilmar Mendes em Portugal foram blo...
3  Os parlamentares de oposição apagaram posts em...
4  O vídeo de um homem mergulhando em trecho do r...

                                                    resposta_gpt
data_resposta  \
```

[13] Devido a extensão do dataset, este apêndice retrata apenas o seu cabeçalho. Para ter acesso a versão completa, acesse a pasta do Google Drive “Códigos, Prompts e Dataset” a partir do link a seguir: https://drive.google.com/drive/folders/1HgCIE0VYp8aHTHFX0BuPuJ9JTrV-K_UF?usp=sharing

```
0    Não, a TIM não deixará de operar em regiões d...
08/03/2025
1    Sim, ao criticar o Brasil, Trump ignora que os...
08/03/2025
2    Não, os bens de Gilmar Mendes em Portugal não...
08/03/2025
3    Não há evidências de que parlamentares de opo...
08/03/2025
4    Não, o vídeo que mostra um homem mergulhando ...
08/03/2025

    few_shot_cot_gpt4o   few_shot_cot_deepseek   zero_shot_gpt4o
\
0                 1.00                   1.00              1.00
1                 0.31                   0.25              0.65
2                 1.00                   0.95              1.00
3                 0.86                   0.85              0.85
4                 0.86                   0.90              0.85

   zero_shot_deepseek
0                0.95
1                0.45
2                0.95
3                0.95
4                0.95
```

APÊNDICE C - CÓDIGO PYTHON UTILIZADO PARA MEDIR A SIMILARIDADE SEMÂNTICA GERADA A PARTIR DA TÉCNICA DE PROMPTING ZERO SHOT NO MODELO DEEPSEEK-V3

```
'''
SIMILARIDADE DEEPSEEK ZERO SHOT
EXECUÇÃO: 53:48
CUSTO: $0.35
'''

import nest_asyncio
nest_asyncio.apply()
import asyncio
import httpx
import pandas as pd
import json
from tqdm.asyncio import tqdm_asyncio


DEEPSEEK_API_KEY = "XXX"  # Substitua pela sua chave real
DEEPSEEK_ENDPOINT                                                    =
"https://api.deepseek.com/v1/chat/completions"

file_path = "/content/checagens_gpt_2.csv"
df = pd.read_csv(file_path)

coluna_texto1 = "primeiro_paragrafo_agencia"
coluna_texto2 = "resposta_gpt"

headers = {
    "Authorization": f"Bearer {DEEPSEEK_API_KEY}",
    "Content-Type": "application/json"
}

def build_payload(text1, text2):
    prompt = f"""
      Avalie a similaridade semântica entre os dois textos
abaixo e retorne um número entre 0 e 1.
      O intuito é verificar se ambos os textos, ainda que
escritos de maneiras diferentes, corroboram com a mesma ideia.
    - 1 significa que os textos têm o mesmo significado.
      - 0 significa que os textos não têm nenhuma relação
semântica.
     - Valores intermediários representam diferentes graus de
similaridade.

    Texto 1: "{text1}"
    Texto 2: "{text2}"
```

```
      Responda apenas com um número decimal entre 0 e 1, com
duas casas decimais. Não inclua nenhuma outra informação na
resposta.
    """
    return {
        "model": "deepseek-chat",
        "messages": [
             {"role": "system", "content": "Você é um avaliador
de similaridade semântica."},
            {"role": "user", "content": prompt}
        ],
        "temperature": 0.0
    }

semaphore = asyncio.Semaphore(4)   # Limite de 5 requisições
simultâneas

async def fetch_similarity(client, text1, text2, index):
    async with semaphore:
        try:
            payload = build_payload(text1, text2)
               response = await client.post(DEEPSEEK_ENDPOINT,
headers=headers, data=json.dumps(payload), timeout=30)
            response.raise_for_status()
                                                    content    =
response.json()["choices"][0]["message"]["content"].strip()
             return float(content) if 0 <= float(content) <= 1
else None
        except Exception as e:
            print(f"Erro na linha {index}: {e}")
            return None

async def main():
    async with httpx.AsyncClient() as client:
        tasks = [
              fetch_similarity(client, str(row[coluna_texto1]),
str(row[coluna_texto2]), idx)
            for idx, row in df.iterrows()
        ]
        results = await tqdm_asyncio.gather(*tasks)

        df["similaridade_deepseek"] = results
              df.to_csv("full_deepseek_zero_shot_26-03-25.csv",
index=False)
        print("✅ Processamento finalizado.")
        display(df)

try:
    asyncio.run(main())
except RuntimeError:
    await main()
```

APÊNDICE D - CÓDIGO PYTHON UTILIZADO PARA MEDIR A SIMILARIDADE SEMÂNTICA GERADA A PARTIR DA TÉCNICA DE PROMPTING FEW SHOT COT NO MODELO DEEPSEEK-V3

```
'''
SIMILARIDADE DEEPSEEK FEW SHOT COT
EXECUÇÃO: 1:06:55
CUSTO: $0.09

'''

import nest_asyncio
nest_asyncio.apply()
import asyncio
import httpx
import pandas as pd
import json
from tqdm.asyncio import tqdm_asyncio


DEEPSEEK_API_KEY = "XXX"  # Substitua pela sua chave real
DEEPSEEK_ENDPOINT                                                    =
"https://api.deepseek.com/v1/chat/completions"

file_path = "/content/checagens_gpt_2.csv"
df = pd.read_csv(file_path)

coluna_texto1 = "primeiro_paragrafo_agencia"
coluna_texto2 = "resposta_gpt"

headers = {
    "Authorization": f"Bearer {DEEPSEEK_API_KEY}",
    "Content-Type": "application/json"
}

def build_payload(text1, text2):
    prompt = f"""
       Avalie a similaridade semântica entre os dois textos
abaixo e retorne um número entre 0 e 1.

    Seu objetivo é identificar se os textos, mesmo escritos de
maneiras diferentes, transmitem a mesma ideia.

    Escala de Similaridade:
    - 0.0 – 0.3: Fracos ou inexistentes paralelos semânticos
    - 0.31 – 0.6: Similaridade parcial
    - 0.61 – 0.85: Boa correspondência de sentido
     - 0.86 – 1.0: Correspondência quase total de conteúdo e
intenção

    ### Exemplos:
```

Texto 1: "Não é verdade que a operadora de telefonia Tim anunciou que deixará de fornecer seus serviços
em comunidades do Rio de Janeiro devido à interferência do tráfico de drogas e das milícias.
Em nota, a empresa desmentiu as alegações que circulam nas redes e afirmou que "segue operando normalmente
em todas as regiões que já atende"."

Texto 2: "Não, a TIM não deixará de operar em regiões dominadas por tráfico e milícia no Rio de Janeiro.
A empresa desmentiu um comunicado falso que circula nas redes sociais e afirmou que 'segue operando
normalmente em todas as regiões que já atende'. "

Passo a passo:
- Foco: Ambos os textos tratam da mesma alegação ou temática
- Veracidade: Ambos apresentam a mesma conclusão sobre a veracidade do conteúdo
- Fundamentação: O raciocínio e as evidências utilizadas são semelhantes
Similaridade: **1.00**

Texto 1: "Não é recente o vídeo em que Donald Trump, eleito presidente dos Estados Unidos na
última terça-feira (5), pede a libertação de israelenses feitos reféns pela grupo extremista
palestino Hamas, como afirmam publicações nas redes. A declaração ocorreu em julho deste ano
durante a Convenção Nacional do Partido Republicano."

Texto 2: "O vídeo em que Donald Trump pede a liberação dos reféns do Hamas é recente,
datado de 5 de março de 2025. Nessa ocasião, Trump emitiu um "último aviso" ao Hamas para que
libertasse todos os reféns ainda mantidos em Gaza. Portanto, o vídeo é pós-eleição."

Passo a passo:
- Foco: Ambos os textos tratam da mesma temática, mas não da mesma alegação, pois se contradizem quanto à data do ocorrido.
- Veracidade: Ambos não apresentam a mesma conclusão sobre a veracidade do conteúdo
- Fundamentação: O raciocínio e as evidências utilizadas são diferentes.
Similaridade: **0.00**

```
    ---

    Agora avalie o seguinte par de textos:

    Texto 1: "{text1}"
    Texto 2: "{text2}"

     Pense passo a passo, identifique a similaridade e retorne
apenas um número entre 0 e 1 com até duas casas decimais.
    **Não inclua nenhuma explicação ou justificativa.**
    """
    return {
        "model": "deepseek-chat",
        "messages": [
             {"role": "system", "content": "Você é um avaliador
de similaridade semântica."},
            {"role": "user", "content": prompt}
        ],
        "temperature": 0.0
    }

semaphore = asyncio.Semaphore(5)   # Limite de 5 requisições
simultâneas

async def fetch_similarity(client, text1, text2, index):
    async with semaphore:
        try:
            payload = build_payload(text1, text2)
               response = await client.post(DEEPSEEK_ENDPOINT,
headers=headers, data=json.dumps(payload), timeout=30)
            response.raise_for_status()
                                                    content    =
response.json()["choices"][0]["message"]["content"].strip()
             return float(content) if 0 <= float(content) <= 1
else None
        except Exception as e:
            print(f"Erro na linha {index}: {e}")
            return None

async def main():
    async with httpx.AsyncClient() as client:
        tasks = [
              fetch_similarity(client, str(row[coluna_texto1]),
str(row[coluna_texto2]), idx)
            for idx, row in df.iterrows()
        ]

        results = await tqdm_asyncio.gather(*tasks)

        df["similaridade_deepseek"] = results
```

```
            df.to_csv("full_deepseek_few_shot_cot_26-03-25.csv",
index=False)
        print("✅ Processamento finalizado.")
        display(df)

try:
    asyncio.run(main())
except RuntimeError:
    await main()
```

APÊNDICE E - CÓDIGO PYTHON UTILIZADO PARA MEDIR A SIMILARIDADE SEMÂNTICA GERADA A PARTIR DA TÉCNICA DE PROMPTING ZERO SHOT NO MODELO GPT-4O

```
'''
SIMILARIDADE GPT4O ZERO SHOT
EXECUÇÃO: 12:16
CUSTO: $2.63

'''

import nest_asyncio
nest_asyncio.apply()
import asyncio
import httpx
import pandas as pd
import json
from tqdm.asyncio import tqdm_asyncio


# === CONFIGURAÇÕES ===
file_path = "/content/checagens_gpt_2.csv"
df = pd.read_csv(file_path)

coluna_texto1 = "primeiro_paragrafo_agencia"
coluna_texto2 = "resposta_gpt"

# === CHAVE E ENDPOINT DO GPT-4o ===
OPENAI_API_KEY = "XXX"  # Substitua com sua chave real
OPENAI_ENDPOINT = "https://api.openai.com/v1/chat/completions"

headers_openai = {
    "Authorization": f"Bearer {OPENAI_API_KEY}",
    "Content-Type": "application/json"
}

# === CONTROLE DE CONCORRÊNCIA ===
semaphore = asyncio.Semaphore(5)   # Limita para 3 requisições
simultâneas

# === PROMPT PADRÃO ===
def build_prompt(text1, text2):
    return f"""
      Avalie a similaridade semântica entre os dois textos
abaixo e retorne um número entre 0 e 1.
      O intuito é verificar se ambos os textos, ainda que
escritos de maneiras diferentes, corroboram com a mesma ideia.
    - 1 significa que os textos têm o mesmo significado.
      - 0 significa que os textos não têm nenhuma relação
semântica.
```

```
     - Valores intermediários representam diferentes graus de
similaridade.

    Texto 1: "{text1}"
    Texto 2: "{text2}"

     Responda apenas com um número decimal entre 0 e 1, com
duas casas decimais. Não inclua nenhuma outra informação na
resposta.
    """

# === MONTAR PAYLOAD ===
def build_payload(text1, text2):
    return {
        "model": "gpt-4o",
        "messages": [
             {"role": "system", "content": "Você é um avaliador
de similaridade semântica."},
                {"role": "user", "content": build_prompt(text1,
text2)}
        ],
        "temperature": 0.0
    }

# === FUNÇÃO ASSÍNCRONA COM BACKOFF ===
async def fetch_similarity(client, text1, text2, retries=3,
delay=5):
    async with semaphore:
        for attempt in range(retries):
            try:
                payload = build_payload(text1, text2)
                response = await client.post(
                    OPENAI_ENDPOINT,
                    headers=headers_openai,
                    data=json.dumps(payload),
                    timeout=60
                )
                response.raise_for_status()
                                                    result   =
response.json()["choices"][0]["message"]["content"].strip()
                  return float(result) if 0 <= float(result) <=
1 else None
            except httpx.HTTPStatusError as e:
                if e.response.status_code == 429:
                    wait_time = delay * (2 ** attempt)
                        print(f"[GPT-4o] Limite de requisições
atingido. Aguardando {wait_time}s...")
                    await asyncio.sleep(wait_time)
                else:
                    print(f"[GPT-4o] Erro HTTP: {e}")
                    break
```

```
            except Exception as e:
                print(f"[GPT-4o] Erro geral: {e}")
                break
        return None

# === LOOP PRINCIPAL ===
async def main():
    async with httpx.AsyncClient() as client:
        tasks = []
        for _, row in df.iterrows():
            text1 = str(row[coluna_texto1])
            text2 = str(row[coluna_texto2])
                  tasks.append(fetch_similarity(client,  text1,
text2))

        results = await tqdm_asyncio.gather(*tasks)

        df["similaridade_gpt4o"] = results
                df.to_csv("full_gpt_4o_zero_shot_26-03-25.csv",
index=False)
          print("✅ Processamento concluído. Resultados salvos
em 'full_gpt_4o_zero_shot_26-03-25.csv'.")
        display(df)

# === EXECUTAR ===
asyncio.run(main())
```

APÊNDICE F - CÓDIGO PYTHON UTILIZADO PARA MEDIR A SIMILARIDADE SEMÂNTICA GERADA A PARTIR DA TÉCNICA DE PROMPTING FEW SHOT COT NO MODELO GPT-4O

```
'''
SIMILARIDADE GPT4O FEW SHOT COT
EXECUÇÃO: 32:05
CUSTO: $0.32

'''


import nest_asyncio
nest_asyncio.apply()
import asyncio
import httpx
import pandas as pd
import json
from tqdm.asyncio import tqdm_asyncio


# === CONFIGURAÇÕES ===
file_path = "/content/checagens_gpt_2.csv"
df = pd.read_csv(file_path)

coluna_texto1 = "primeiro_paragrafo_agencia"
coluna_texto2 = "resposta_gpt"

# === CHAVE E ENDPOINT DO GPT-4o ===
OPENAI_API_KEY = "XXX"  # Substitua com sua chave real
OPENAI_ENDPOINT = "https://api.openai.com/v1/chat/completions"

headers_openai = {
    "Authorization": f"Bearer {OPENAI_API_KEY}",
    "Content-Type": "application/json"
}

# === CONTROLE DE CONCORRÊNCIA ===
semaphore = asyncio.Semaphore(3)  # Limita para 3 requisições
simultâneas

# === PROMPT PADRÃO ===
def build_prompt(text1, text2):
    return f"""
      Avalie a similaridade semântica entre os dois textos
abaixo e retorne um número entre 0 e 1.

    Escala de Similaridade:
    - 0.0 – 0.3: Fracos ou inexistentes paralelos semânticos
    - 0.31 – 0.6: Similaridade parcial
    - 0.61 – 0.85: Boa correspondência de sentido
```

- 0.86 – 1.0: Correspondência quase total de conteúdo e intenção

Exemplos:

Texto 1: "Não é verdade que a operadora de telefonia Tim anunciou que deixará de fornecer seus serviços
em comunidades do Rio de Janeiro devido à interferência do tráfico de drogas e das milícias.
Em nota, a empresa desmentiu as alegações que circulam nas redes e afirmou que "segue operando normalmente
em todas as regiões que já atende"."

Texto 2: "Não, a TIM não deixará de operar em regiões dominadas por tráfico e milícia no Rio de Janeiro.
A empresa desmentiu um comunicado falso que circula nas redes sociais e afirmou que 'segue operando
normalmente em todas as regiões que já atende'. "

Passo a passo:
- Foco: Ambos os textos tratam da mesma alegação ou temática
- Veracidade: Ambos apresentam a mesma conclusão sobre a veracidade do conteúdo
- Fundamentação: O raciocínio e as evidências utilizadas são semelhantes
Similaridade: **1.00**

Texto 1: "Não é recente o vídeo em que Donald Trump, eleito presidente dos Estados Unidos na
última terça-feira (5), pede a libertação de israelenses feitos reféns pela grupo extremista
palestino Hamas, como afirmam publicações nas redes. A declaração ocorreu em julho deste ano
durante a Convenção Nacional do Partido Republicano."

Texto 2: "O vídeo em que Donald Trump pede a liberação dos reféns do Hamas é recente,
datado de 5 de março de 2025. Nessa ocasião, Trump emitiu um "último aviso" ao Hamas para que
libertasse todos os reféns ainda mantidos em Gaza. Portanto, o vídeo é pós-eleição."

Passo a passo:
- Foco: Ambos os textos tratam da mesma temática, mas não da mesma alegação, pois se contradizem quanto à data do ocorrido.
- Veracidade: Ambos não apresentam a mesma conclusão sobre a veracidade do conteúdo

```
     - Fundamentação: O raciocínio e as evidências utilizadas
são diferentes.
    Similaridade: **0.00**

    ---

    Agora avalie o seguinte par de textos:

    Texto 1: "{text1}"
    Texto 2: "{text2}"

     Pense passo a passo, identifique a similaridade e retorne
apenas um número entre 0 e 1 com até duas casas decimais.
    **Não inclua nenhuma explicação ou justificativa.**
    """

# === MONTAR PAYLOAD ===
def build_payload(text1, text2):
    return {
        "model": "gpt-4o",
        "messages": [
             {"role": "system", "content": "Você é um avaliador
de similaridade semântica."},
               {"role": "user", "content": build_prompt(text1,
text2)}
        ],
        "temperature": 0.0
    }

# === FUNÇÃO ASSÍNCRONA COM BACKOFF ===
async  def  fetch_similarity(client,  text1,  text2,  retries=3,
delay=5):
    async with semaphore:
        for attempt in range(retries):
            try:
                payload = build_payload(text1, text2)
                response = await client.post(
                    OPENAI_ENDPOINT,
                    headers=headers_openai,
                    data=json.dumps(payload),
                    timeout=60
                )
                response.raise_for_status()
                                                        result   =
response.json()["choices"][0]["message"]["content"].strip()
                  return float(result) if 0 <= float(result) <=
1 else None
            except httpx.HTTPStatusError as e:
                if e.response.status_code == 429:
                    wait_time = delay * (2 ** attempt)
```

```
                        print(f"[GPT-4o] Limite de requisições
atingido. Aguardando {wait_time}s...")
                    await asyncio.sleep(wait_time)
                else:
                    print(f"[GPT-4o] Erro HTTP: {e}")
                    break
            except Exception as e:
                print(f"[GPT-4o] Erro geral: {e}")
                break
        return None

# === LOOP PRINCIPAL ===
async def main():
    async with httpx.AsyncClient() as client:
        tasks = []
        for _, row in df.iterrows():
            text1 = str(row[coluna_texto1])
            text2 = str(row[coluna_texto2])
                  tasks.append(fetch_similarity(client,  text1,
text2))

        results = await tqdm_asyncio.gather(*tasks)

        df["similaridade_gpt4o"] = results
             df.to_csv("1200_gpt_4o_few_shot_cot_26-03-25.csv",
index=False)
          print("✅ Processamento concluído. Resultados salvos
em '1200_gpt_4o_few_shot_cot_26-03-25.csv'.")

# === EXECUTAR ===
asyncio.run(main())
```