Uma Abordagem Baseada em LLMs para Análise de Desinformação sobre Vacinas

Athus Cavalini[1,2], Leandro Furlam Turi[2], André G. C. Pacheco[2], Giovanni Comarela[2]

[1]Instituto Federal do Espírito Santo
Vitória – ES – Brasil

[2]Universidade Federal do Espírito Santo
Vitória – ES – Brasil

athus.cavalini@ifes.edu.br, leandro.turi@edu.ufes.br,
apacheco@inf.ufes.br, gc@inf.ufes.br

Abstract. *This article proposes a fully LLM-based approach to extract insights from large-scale data on vaccine misinformation. Using BERT and GPT models, we examined millions of anti-vaccine messages from Telegram. The analysis identified 12 key misinformation narratives, highlighting alarmist concerns and conspiracy theories, while also suggesting mitigation strategies. The results demonstrate the potential of LLMs for large-scale automated misinformation analysis, mapping its dynamics and providing insights for the development of agile and evidence-based strategic responses.*

Resumo. *Este artigo propõe uma abordagem totalmente baseada em Grandes Modelos de Linguagem (LLMs) para obter percepções a partir de dados em massa de desinformação sobre vacinas. Utilizando modelos BERT e GPT, foram analisadas milhões de mensagens antivacina coletadas do Telegram. A análise revelou 12 principais narrativas desinformativas, destacando preocupações alarmistas e teorias conspiratórias, além de indicar estratégias de mitigação. Os resultados evidenciam o potencial dos LLMs na análise automatizada de desinformação em larga escala, mapeando as dinâmicas e fornecendo insights para o desenvolvimento de respostas estratégicas ágeis e baseadas em evidências.*

1. Introdução

A desinformação em saúde tem se mostrado um dos maiores desafios contemporâneos para a saúde pública global. Essa problemática se consolidou durante a pandemia da COVID-19 [Ministério da Saúde 2024], quando narrativas desinformativas amplificadas pelas plataformas digitais contribuíram fortemente para a hesitação vacinal, afetando diretamente as taxas de imunização e prolongando a crise sanitária [Albuquerque 2023, World Health Organization 2022]. Diante desse cenário, entender a dinâmica de disseminação e os padrões argumentativos dessas narrativas tornou-se uma etapa relevante no desenvolvimento de estratégias eficazes de comunicação e intervenção para o combate à desinformação.

Nos últimos anos, os avanços em Grandes Modelos de Linguagem (*Large Language Models*, LLMs) revolucionaram o Processamento de Linguagem Natural (*Natural Language Processing*, NLP), possibilitando análises em larga escala de grandes volumes de dados textuais complexos [Brown et al. 2020, Devlin et al. 2019, Radford et al. 2018]. Esses modelos têm demonstrado eficácia na identificação de padrões semânticos, modelagem de tópicos e classificação de sentimentos, sendo amplamente utilizados para analisar conteúdos gerados em mídias sociais [Chang et al. 2024, Burghardt et al. 2020]. Apesar do crescente interesse acadêmico, poucos estudos exploram o uso integrado dessas tecnologias para compreender a desinformação em saúde pública e seus impactos no comportamento social.

Diante desse contexto, este estudo busca responder à seguinte pergunta de pesquisa: **[RQ] Como os LLMs podem ser utilizados para identificar e categorizar padrões narrativos na desinformação sobre vacinas?**. Para isso, propôs-se uma abordagem totalmente baseada em LLMs que combina modelagem de tópicos utilizando *Bidirectional Encoder Representations from Transformers* (BERT) [Devlin et al. 2019] e análise semântica assistida por *Generative Pre-Trained Transformers* (GPTs) [Radford et al. 2018], empregando o modelo `DeepSeek-R1 32B` [DeepSeek-AI et al. 2025] para investigar padrões narrativos em milhões de mensagens antivacina coletadas do Telegram. Ao categorizar as principais narrativas de desinformação, a pesquisa oferece percepções sobre suas estratégias argumentativas, revelando padrões emergentes e destacando como amplificam o medo e a desconfiança em relação às vacinas.

Além disso, este estudo explora como a aplicação de LLMs pode fornecer subsídios estratégicos para a formulação de ações de mitigação baseadas em comunicação científica. Ao combinar técnicas avançadas de NLP com comunicação estratégica, a pesquisa contribui para o desenvolvimento de abordagens no combate à desinformação em saúde pública, promovendo uma comunicação mais assertiva e orientada por dados.

As principais contribuições deste estudo incluem:

- A demonstração do potencial dos LLMs para análise de desinformação em larga escala, através da identificação de padrões narrativos da comunidade antivacina alinhados aos identificados em pesquisas anteriores;
- O refinamento do uso de LLMs para fornecer diretrizes estratégicas de mitigação, explorando a capacidade dos LLMs de sugerir narrativas contra-argumentativas e intervenções baseadas em evidências; e
- A proposição de um *framework* metodológico estratégico para análise automatizada da desinformação, replicável para pesquisadores, agências de saúde e formuladores de políticas públicas.

Até onde sabemos, este é o primeiro estudo a propor uma abordagem totalmente baseada em LLMs para a análise de narrativas de desinformação em grandes volumes de dados. Além de demonstrar a viabilidade da técnica, os resultados obtidos foram validados por pesquisas que utilizaram metodologias tradicionais, reforçando a robustez dos achados e destacando o potencial dessas ferramentas para ampliar a compreensão das dinâmicas da desinformação e subsidiar estratégias eficazes de mitigação.

2. Trabalhos Relacionados

A análise de narrativas desinformativas e o uso de grandes modelos de linguagem têm sido foco de diversos estudos recentes. O avanço em NLP possibilitou novas abordagens para a modelagem de tópicos e o agrupamento de grandes volumes de dados textuais em saúde pública. Neste contexto, explorou-se trabalhos que se relacionam diretamente com a abordagem proposta, dividindo-os em três principais panoramas: análise de desinformação em mídias sociais, técnicas de modelagem de tópicos utilizando BERT e aplicação de algoritmos de agrupamento para a análise de dados em saúde pública.

A desinformação sobre vacinas é uma preocupação crescente, conforme já destacado por canais do SUS [Sistema Único de Saúde 2023]. [De and Vats 2023] aplicaram técnicas avançadas de NLP baseadas em modelos BERT e GPT-3.5 para classificar sentimentos em postagens sobre vacinas em redes sociais. O trabalho apresentou, a partir da combinação dessas tecnologias, uma categorização de preocupações da população, possibilitando uma melhor compreensão das motivações por trás da hesitação vacinal.

Dentre os estudos que analisaram padrões narrativos da desinformação antivacina, destacam-se os trabalhos de [Massarani et al. 2021], [Malini et al. 2024], [Gehrke and Benetti 2021] e [Hughes et al. 2021]. [Massarani et al. 2021] demonstraram que as narrativas desinformativas sobre vacinação nas redes sociais desempenharam um papel importante na formação de opinião e comportamento popular. No mesmo sentido, [Malini et al. 2024] analisaram mensagens de canais antivacina no Telegram e mapearam cinco principais tipos de desinformação: pseudoprotetiva, imunológica, falsa utilidade, medicalizante e conspiracionista.

Os trabalhos de [Gehrke and Benetti 2021] e [Hughes et al. 2021] complementam essas análises, mapeando temas recorrentes e classificando as retóricas desinformativas, como o uso de jargões pseudocientíficos para conferir credibilidade a informações falsas, a promoção de tratamentos sem comprovação científica, o apelo emocional de testemunhos pessoais, a desconfiança em instituições de saúde e teorias da conspiração. Os estudos fornecem um entendimento detalhado das estratégias discursivas utilizadas para gerar desconfiança em relação às vacinas e influenciar a hesitação vacinal.

Além da análise de narrativas desinformativas, a modelagem de tópicos tem sido amplamente empregada para organizar grandes volumes de dados textuais. [George and Sumathy 2023] propõem um arcabouço que combina BERT, *Latent Dirichlet Allocation* (LDA) e técnicas de agrupamento para aprimorar a modelagem de tópicos. Os resultados indicaram que essa abordagem híbrida permite a extração de tópicos mais coerentes e semanticamente ricos em comparação com métodos tradicionais. [Baratieri et al. 2021] aplicaram o arcabouço a um conjunto de trabalhos relacionados à pandemia da COVID-19, identificando os principais temas abordados na literatura e permitindo uma visão abrangente da evolução do conhecimento sobre o vírus.

No contexto do agrupamento de dados em saúde pública, estudos demonstram o potencial da técnica na identificação de padrões epidemiológicos e otimização da alocação de recursos. Técnicas de *clustering* foram aplicadas por [YoshimiTanaka et al. 2015] e [Junior et al. 2022] para segmentar unidades de atendimento do SUS e avaliar o desempenho da atenção primária, respectivamente. De forma similar, [Zhong et al. 2024] utilizaram abordagens bayesianas para mapear padrões espaciais e temporais em surtos de

doenças, como COVID-19 e dengue, destacando a importância da análise automatizada para a implementação de estratégias preventivas mais eficazes.

Diante dos estudos mencionados, observa-se que a combinação de técnicas avançadas de Inteligência Artificial e Ciência de Dados tem-se mostrado promissora no apoio à gestão da saúde coletiva, fornecendo suporte na formulação de políticas e na implementação de ações mais eficazes. Este trabalho se insere neste contexto ao propor uma abordagem integrada baseada em BERT e GPT para a identificação e análise de narrativas antivacina, explorando a segmentação semântica e a extração de padrões narrativos em dados coletados de comunidades antivacina do Telegram. Dessa forma, buscou-se contribuir para o avanço das metodologias computacionais aplicadas à análise da desinformação e seus impactos na saúde pública.

3. Metodologia

Esta seção descreve as etapas e procedimentos adotados para analisar as narrativas desinformativas antivacina combinando técnicas de modelagem de tópicos e análise semântica assistida por GPT. A Seção 3.1 descreve a coleta e pré-processamento do conjunto de dados, enquanto as Seções 3.2 e 3.3 apresentam a modelagem de tópicos e a análise semântica assistida por GPT, para identificação e classificação dos padrões argumentativos. A Figura 1 apresenta uma visão geral das etapas aplicadas.

3.1. Coleta e Pré-processamento de Dados

Os dados utilizados foram coletados utilizando o *Telegram Observatory* [Cavalini et al. 2023], uma ferramenta que opera por meio da API oficial do Telegram. A aplicação possibilita a coleta em massa de informações de diversos canais e grupos da plataforma. A seleção dos grupos-alvo foi baseada em registros públicos do Observatório da Saúde nas Redes Sociais [Instituto Capixaba de Ensino, Pesquisa e Inovação 2024], que monitora 779 grupos e canais dedicados à disseminação de narrativas antivacinação. No total, foram recuperadas 9.941.879 mensagens compartilhadas entre março de 2020 e agosto de 2023.

Após a coleta, o conjunto de mensagens passou por uma etapa de pré-processamento que removeu mensagens vazias, URLs, nomes de usuários, *emojis*, caracteres especiais e entradas duplicadas. Também foram excluídas as mensagens com menos de 60 caracteres ou que não estivessem em Língua Portuguesa. Esse procedimento resultou em um *corpus* final de 690.552 mensagens, correspondendo a 6,94% do total inicialmente coletado.

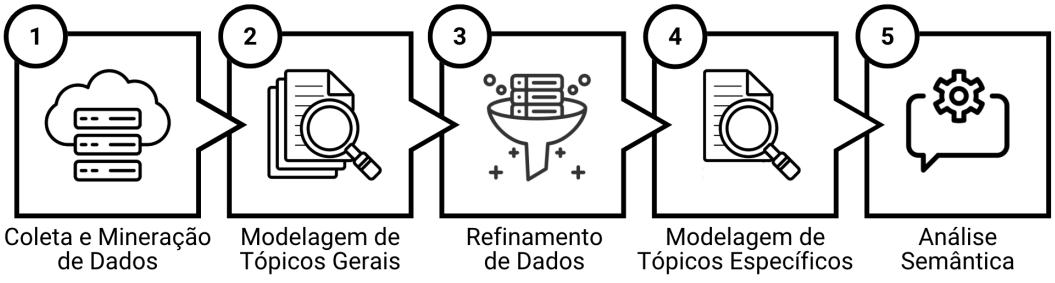

Figura 1. Etapas da metodologia utilizada.

***Prompt* 1**:
Tenho um tópico que contém os seguintes documentos: - *[LISTA DE DOCUMENTOS]* O tópico é representado pelas seguintes palavras-chave: *[PALAVRAS-CHAVE]* Com base nessas informações, crie uma descrição curta que represente o tópico.

Figura 2. *Prompt* utilizado para instruir a rotulagem dos tópicos.

3.2. Modelagem de Tópicos

A modelagem inicial de tópicos foi baseada na ferramenta BERTopic [Grootendorst 2022]. Primeiro, foram geradas representações numéricas das sentenças (*embeddings*) usando o modelo `paraphrase-multilingual-MiniLM-L12-v2`, disponível no *framework Sentence Transformers* [Reimers and Gurevych 2019].

Em seguida, foi aplicado o método UMAP [McInnes et al. 2020] para redução de dimensionalidade dos dados, a fim de facilitar a identificação de padrões. Por fim, o algoritmo HDBSCAN [McInnes et al. 2017] foi empregado para agrupar automaticamente os textos, sem a necessidade de definir previamente um número fixo de tópicos.

Após a primeira execução do BERTopic, os tópicos associados à desinformação sobre vacinas foram revisados manualmente pelos autores a fim de descartar conteúdos irrelevantes ou fora do escopo. Esse refinamento manual possibilitou a seleção de um subconjunto de mensagens estritamente relacionadas à desinformação antivacina.

Com o *corpus* refinado, o modelo foi reexecutado utilizando a mesma abordagem inicial. Essa modelagem de tópicos em duas etapas permitiu (i) isolar apenas o conteúdo relevante para o objetivo da pesquisa, (ii) identificar com maior precisão os tópicos abordados no conjunto de mensagens desinformativas e (iii) refinar as representações e permitir uma interpretação mais detalhada dos padrões narrativos e estratégias discursivas usadas na disseminação da desinformação sobre vacinas.

3.3. Análise Semântica

Nesta etapa, os tópicos identificados na modelagem foram rotulados utilizando o modelo GPT de código aberto `DeepSeek-R1 32B`[1] [DeepSeek-AI et al. 2025]. Cada tópico recebeu um rótulo representativo a partir da análise de seus documentos, orientado pelo *prompt* apresentado na Figura 2.

Essa abordagem explorou a capacidade dos modelos GPT de capturar nuances semânticas complexas e gerar rótulos mais descritivos e representativos dos tópicos identificados, possibilitando uma análise mais contextualizada das narrativas e conferindo maior interpretabilidade aos resultados.

Além da rotulagem, as mensagens pertencentes aos tópicos refinados foram submetidas a uma análise semântica detalhada pelo modelo DeepSeek. Esse processamento envolveu a identificação de padrões recorrentes nas narrativas desinformativas a partir do processamento massivo das mensagens de cada tópico identificado, de acordo com

[1]A escolha do `DeepSeek-R1 32B` se justifica por ser um modelo gratuito e acessível, permitindo sua execução local. O tamanho do modelo foi definido a partir dos recursos computacionais disponíveis.

***Prompt* 2**:
Analise as mensagens a seguir, veiculadas em grupos antivacina. Liste as principais narrativas utilizadas para espalhar desinformação sobre o tema. Para cada uma das narrativas, sugira uma ação de mitigação com foco na comunicação estratégica das agências públicas de saúde.

Figura 3. *Prompt* utilizado para identificar as narrativas de desinformação.

o *prompt* apresentado na Figura 3. Conforme destacado por [Malini et al. 2024], esta análise proporciona uma compreensão abrangente das estratégias de comunicação utilizadas para disseminar desinformação sobre vacinas, além de oferecer percepções para o desenvolvimento de estratégias de comunicação eficazes e baseadas em evidências.

4. Resultados e Discussão

4.1. Modelagem de Tópicos

A execução inicial do BERTopic resultou na identificação de 250 tópicos, que foram agrupados em 50 categorias principais. Dentre esses, dois tópicos mostraram-se particularmente relevantes para a análise de desinformação sobre vacinas. O primeiro, com 9.143 mensagens, foi representado pelas palavras-chave *vacina*, *vacinas*, *dose*, *vacinação*, *passaporte*, *vacinados*, *hepatite*, *vírus*, *coronavac*. Esse conjunto de mensagens evidenciou uma narrativa centrada em preocupações sobre a segurança das vacinas, eficácia e exigências de passaporte vacinal. O segundo tópico, composto por 132 mensagens, foi caracterizado pelas palavras-chave *certificadodevacina*, *conectesus*, *certificado*, *vacina*, *passaporte*, *vacinação*, *doença*, indicando um foco em questões relacionadas ao passaporte sanitário e ao sistema ConecteSUS [Sistema Único de Saúde 2025].

Tabela 1. Resultado da segunda modelagem de tópicos.

Tópico	No. de Msgs	Descrição
1	135	Registro de doses no ConectSUS
2	130	Gripe Aviária H5N1 Detectada no Brasil
3	133	Passaporte Sanitário e Registro de Doses no ConectSUS
4	7892	Preocupações sobre a Segurança das Vacinas contra a COVID-19
5	255	Níveis de Dímero-D após a Vacinação
6	120	Doenças Transmitidas por Mosquitos
7	177	Alegações de Adição de HIV nas Vacinas COVID-19
8	112	Remédios Antivirais Naturais
9	131	Emissão de Documentos de Vacina
10	220	Serviços de Passaporte Vacinal

Ao aplicar um refinamento sobre esses dois tópicos (terceira e quarta etapas da metodologia), o BERTopic foi reexecutado, resultando na identificação de 10 novos tópicos, detalhados na Tabela 1. O tópico mais expressivo foi “Preocupações sobre a Segurança das Vacinas contra a COVID-19” (Tópico 4), que concentrou 7.892 mensagens, evidenciando um alto volume de narrativas voltadas à desconfiança em relação aos possíveis

efeitos adversos das vacinas. Ainda, a análise revelou que essa preocupação foi amplificada por discursos alarmistas que correlacionavam a vacinação com eventos de saúde graves, incluindo discussões sobre "Níveis de Dímero-D após a Vacinação" (Tópico 5), "Alegações de Adição de HIV nas Vacinas" (Tópico 7) e a promoção de "Remédios Antivirais Naturais" (Tópico 8).

4.2. Análise Semântica das Narrativas

A análise semântica dos tópicos refinados realizada pelo modelo GPT revelou 12 padrões narrativos recorrentes na desinformação antivacina. Os principais temas identificados incluem: teorias da conspiração, alarmismo sobre efeitos adversos, desconfiança em instituições de saúde e promoção de tratamentos alternativos sem comprovação científica. Essas narrativas, em grande parte, corroboram os achados de estudos anteriores baseados em métodos tradicionais, indicando a eficácia dos LLMs na identificação de padrões semânticos complexos.

Para evidenciar o alinhamento entre os resultados do modelo e as pesquisas anteriores, a Tabela 2 apresenta um comparativo entre as narrativas identificadas pelo modelo e as equivalentes descritas na literatura. O resultado completo na análise semântica, incluindo a descrição das narrativas, palavras-chave e propostas de ações de mitigação, pode ser acessado no repositório do projeto[2].

A análise semântica do Tópico 4, que se refere à preocupação com os efeitos adversos das vacinas, destacou que os padrões narrativos focam em *Tratamentos Alternativos, Efeitos Adversos, Mortes e Perigos à Saúde, e Desconfiança em Instituições e Empresas Farmacêuticas*. Esses padrões foram amplamente documentados por estudos anteriores, que demonstraram como a desinformação frequentemente utiliza relatos de eventos adversos para gerar medo e hesitação vacinal [Malini et al. 2024].

A promoção de tratamentos alternativos também se destaca, citando especialmente o uso da ivermectina, medicamento sem comprovação de eficácia, como substituto para a vacina. Esse achado se alinha aos estudos de [Gehrke and Benetti 2021] e [Malini et al. 2024], que demonstraram como essas narrativas foram difundidas.

Analisando os Tópicos 4 e 7, o modelo identificou padrões que sugerem riscos à saúde associados às vacinas. No Tópico 5, o modelo identificou os seguintes padrões narrativos: *Alarme sobre os Níveis de Dímero-D Pós-Vacinação, Desconfiança em Profissionais de Saúde, Conspiração Médica e Confusão Científica usando Jargões Médico*, enquanto no Tópico 7: *Conspiração sobre HIV e Vacinas, e Narrativas Históricas e Comparativas com a AIDS e Outras Doenças Virais*. Já na análise semântica do Tópico 8, o modelo destacou narrativas conspiracionistas, especificamente *Conspiração sobre Remédios Naturais e "Fraudemia" e Negacionismo da Pandemia*.

Comparando os resultados com estudos anteriores, as narrativas conspiratórias e de desconfiança em instituições de saúde e farmacêuticas, destacados nos trabalhos de [Massarani et al. 2021] e [Malini et al. 2024], sugerem intenções maliciosas por parte de empresas farmacêuticas e outros integrantes de uma "elite global",A narrativa baseia-se na suposição de que a pandemia foi uma situação criada e/ou mantida para gerar lucro, poder e até mesmo promover a "despopulação" mundial.

[2] `github.com/dsl-ufes/llm-health`

Tabela 2. Comparação entre Narrativas Identificadas e Pesquisas Anteriores

#	Narrativa Identificada	Correspondência em Pesquisas Anteriores
1	Tratamentos Alternativos	**Cura:** trata da eficácia de certos remédios, especialmente hidroxicloroquina, ivermectina, nitazoxanida e azitromicina, além de receitas caseiras e terapêuticas alternativas [Gehrke and Benetti 2021]
11	Conspiração sobre Remédios Naturais	
2	Depoimentos Pessoais	**Personalização:** destaca histórias pessoais que envolvem a vacina abordada [Massarani et al. 2021]; **Pseudoprotetiva:** testemunhos usados como evidência clínica para espalhar medo e desconfiança [Malini et al. 2024]; **Pessoas estão dizendo:** a evidência existe simplesmente porque outras pessoas supostamente a estão afirmando [Gehrke and Benetti 2021]
3	Efeitos Adversos e Perigos à Saúde	**Incertezas científicas:** riscos à saúde, efeitos adversos e limites da ciência [Massarani et al. 2021]; **Lesão por vacina:** todos os danos que a vacina pode causar a você [Gehrke and Benetti 2021]
5	Mortes e Eventos Adversos	
6	Dímero-D e Exames Pós-Vacinação	
4	Desconfiança em Instituições e Empresas Farmacêuticas	**Elites Corruptas:** Elite forçando os *lockdowns* e medidas sanitárias por lucro e/ou poder (ex: farmacêuticas) [Hughes et al. 2021]; **Conspiracionismo:** plano da "elite global", incluindo países e farmacêuticas, para reduzir a população mundial [Malini et al. 2024]
12	"Fraudemia" e Negacionismo da Pandemia	
7	Desconfiança em Profissionais de Saúde	**Controvérsias científicas:** foca nas controvérsias científicas relacionadas à vacina e vacinação [Massarani et al. 2021]; **Dizendo a verdade por poder:** Médicos, enfermeiros e outros especialistas que se manifestam contra o alarmismo da COVID são corajosos, trazendo a verdade para o povo. [Hughes et al. 2021]
8	Confusão Científica com Jargão Médico	**Medicalizante:** uso de argumento de autoridade médica para validar desinformação [Malini et al. 2024]; **Aparência de Autoridade:** utiliza símbolos de autoridade e expertise para dar maior peso a um argumento [Gehrke and Benetti 2021]
9	Conspiração sobre HIV e Vacinas	**Imunológica:** centrada na ideia de que existe perda de imunidade ao se vacinar [Malini et al. 2024]
10	Narrativa Histórica e Comparativa com a AIDS	

Além disso, o modelo detectou o uso recorrente de depoimentos pessoais como estratégia de legitimação emocional, uma tática também já documentada por [Massarani et al. 2021]. Essas narrativas foram amplificadas pelo uso de linguagem pseudocientífica para conferir credibilidade às alegações, uma estratégia também destacada por [Malini et al. 2024].

Além das convergências nas narrativas, a análise orientada por GPT trouxe novas perspectivas ao revelar temas anteriormente não identificados, como a comercialização de certificados de vacinação e a inclusão de doses no sistema público de controle, o ConectSUS (Tópicos 1, 3, 9 e 10). Esses achados demonstram que a aplicação de LLMs na análise de desinformação possibilita a identificação de padrões argumentativos emergentes, contribuindo para a formulação de estratégias de comunicação baseadas em evidências.

4.3. Sugestão de Ações Estratégicas de Mitigação

Diante das análises obtidas, destaca-se a importância de uma comunicação estratégica, guiada por dados, para mitigar o impacto da desinformação em saúde [Sistema Único de Saúde 2023]. Nesse sentido, as sugestões do modelo GPT para ações de mitigação concentram-se em três abordagens estratégicas principais: (i) transparência e clareza na comunicação científica, promovendo informações acessíveis e baseadas em evidências sobre a eficácia e segurança das vacinas, contrapondo-se a narrativas pseudocientíficas; (ii) engajamento com influenciadores confiáveis, incluindo profissionais de saúde respeitados, para desmistificar desinformações e fortalecer a credibilidade das campanhas de vacinação; e (iii) desconstrução de narrativas alarmistas, utilizando análises críticas de vídeos e postagens desinformativas, esclarecendo o contexto científico correto de dados médicos descontextualizados. A saída exata do modelo contendo as propostas de ações de mitigação pode ser acessada através do repositório do projeto.

A proposta de comunicação estratégica ainda inclui o uso de narrativas contra-argumentativas, capazes de desconstruir desinformações através de *storytelling* positivo e empático, como a promoção de histórias reais e positivas de pessoas que foram vacinadas com segurança, enfatizando experiências positivas e desmistificando efeitos adversos menores por meio de explicações baseadas em ciência [Malini et al. 2024]. Além disso, o engajamento direto com a audiência nas redes sociais, oferecendo respostas claras e empáticas para dúvidas e preocupações, é indicado para reduzir a disseminação de informações falsas e aumentar a confiança na vacinação [Ministério da Saúde 2024].

Os achados deste estudo demonstram o potencial dos LLMs na formulação de estratégias de mitigação, permitindo uma análise rápida e abrangente das narrativas de desinformação. No entanto, é fundamental destacar que essas sugestões são preliminares e devem ser validadas por especialistas em comunicação científica, epidemiologia e políticas públicas, a fim de garantir sua aplicabilidade e eficácia. Além disso, a implementação dessas estratégias deve ser continuamente monitorada e ajustada conforme novas dinâmicas de desinformação emergem. Estudos futuros podem avaliar empiricamente a efetividade das ações sugeridas, bem como explorar novas abordagens no combate à desinformação em saúde pública.

5. Considerações Finais

Este estudo indicou o potencial dos modelos de linguagem como ferramentas para a interpretação de dados em massa, especialmente na análise da desinformação em saúde. Ao utilizar uma abordagem combinada de modelagem de tópicos com BERT e análise semântica assistida por modelo GPT, foi possível identificar padrões narrativos complexos em milhões de mensagens antivacina coletadas do Telegram. Essas narrativas foram classificadas em 12 temas principais, destacando preocupações alarmistas e teorias conspiratórias como estratégias persuasivas centrais.

Os resultados encontrados validam a eficácia dos LLMs na detecção de narrativas desinformativas, corroborando achados de estudos anteriores que utilizaram metodologias tradicionais – frequentemente mais onerosas em termos de tempo e esforço. Ao reproduzir achados consistentes com pesquisas baseadas em análises qualitativas manuais e técnicas de NLP convencionais, este estudo destaca o sucesso e a robustez dos LLMs na análise de desinformação em larga escala, reforçando o valor dessas ferramentas como uma alternativa eficiente e escalável para a interpretação de dados complexos, reduzindo significativamente os custos operacionais e o tempo de análise.

Além disso, a aplicação desses modelos permitiu não apenas mapear padrões discursivos, mas também sugerir diretrizes estratégicas para mitigar a disseminação da desinformação, tornando a comunicação científica mais acessível e eficaz. Ao identificar os principais eixos argumentativos utilizados na propagação desinformativa os modelos trabalharam na formulação de estratégias específicas para desarticulá-las, incluindo o fortalecimento da comunicação baseada em evidências, o engajamento de influenciadores confiáveis e o desenvolvimento de contranarrativas.

Embora este trabalho tenha se concentrado na desinformação sobre vacinas, a metodologia proposta pode ser aplicada a outros contextos de saúde pública, como a análise de desinformação sobre medicamentos, terapias não comprovadas e debates sobre políticas sanitárias. A capacidade dos LLMs de identificar padrões narrativos emergentes e gerar relatórios analíticos estratégicos pode contribuir para a formulação de políticas públicas baseadas em evidências e aprimorar a resposta a desafios informacionais em diferentes áreas da saúde.

Como continuidade deste estudo, propõe-se o desenvolvimento de um agente inteligente atualizado em tempo real com informações provenientes de redes sociais e também de fontes confiáveis, capaz de gerar respostas contextualizadas para combater as narrativas desinformativas nas plataformas digitais e fornecer contranarrativas embasadas em evidências científicas, permitindo a disseminação proativa de informações confiáveis.

Agradecimentos

Os autores agradecem o apoio da Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES) e do Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi).

Referências

Albuquerque, F. (2023). Brasil atingiu em 2021 menor cobertura vacinal em 20 anos. https://agenciabrasil.ebc.com.br/saude/noticia/2023-08/brasil-atingiu-em-2021-menor-cobertura-vacinal-em-20-anos. Acesso em: 15 de mar. de 2025.

Baratieri, T., Lentsck, M. H., and Peres, K. C. et al (2021). Modelagem de tópicos de pesquisa sobre o novo coronavírus: aplicação do latent dirichlet allocation. *Ciência, Cuidado E Saúde*, 20(1):e56403.

Brown, T., Mann, B., and Ryder, N. et al. (2020). Language models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Inf. Processing Systems*, volume 33, pages 1877–1901.

Burghardt, K., Chen, K., and Lerman, K. (2020). Large language models reveal information operation goals, tactics, and narrative frames. *arXiv preprint arXiv.2405.03688*, pages 1–15.

Cavalini, A., Malini, F., and Gouveia, F. et al. (2023). Politics and disinformation: Analyzing the use of telegram's information disorder network in brazil for political mobilization. *First Monday*, 28(5):12901.

Chang, Y., Wang, X., and Wang, J. el al. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3):2157–6904.

De, S. and Vats, S. (2023). Decoding concerns: Multi-label classification of vaccine sentiments in social media. In Ghosh, K., Mandl, T., Majumder, P., and Mitra, M., editors, *Working Notes of FIRE 2023 – Forum for Information Retrieval Evaluation (FIRE-WN 2023)*, pages 99–111, India. CEUR-WS.org.

DeepSeek-AI, Guo, D., and Yang, D. et al. (2025). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv.2501.12948*, pages 1–22.

Devlin, J., Chang, M.-W., and Lee, K. et al. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 4171–4186, EUA. Association for Computational Linguistics.

Gehrke, M. and Benetti, M. (2021). A desinformação no brasil durante a pandemia de covid-19:: temas, plataformas e atores. *Fronteiras - estudos midiáticos*, 23(2).

George, L. and Sumathy, P. (2023). An integrated clustering and bert framework for improved topic modelin. *International Journal of Information Technology*, 15:2187–2195.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*, pages 1–10.

Hughes, B., Miller-Idriss, C., Piltch-Loeb, R., Goldberg, B., White, K., Criezis, M., and Savoia, E. (2021). Development of a codebook of online anti-vaccination rhetoric to manage covid-19 vaccine misinformation. *International Journal of Environmental Research and Public Health*, 18(14).

Instituto Capixaba de Ensino, Pesquisa e Inovação (2024). Projeto Observa ICEPi. Acesso em: 10 de fev. de 2025.

Junior, R. V. B. P., Junior, N. C., and Sala, A. et al. (2022). Desempenho da atenção primária à saúde, segundo clusters de municípios convergentes no estado de são paulo. *Revista Brasileira de Epidemiologia*, 25:E220017.

Malini, F., Sodré, F., and Cavalini, A. et al. (2024). Five patterns of vaccine misinformation on telegram. *Lecture Notes in Computer Science*, 15213:181–196.

Massarani, L., Waltz, I., and Leal, T. et al. (2021). Narrativas sobre vacinação em tempos de fake news: uma análise de conteúdo em redes sociais. *Saúde e Sociedade*, 30:e200317.

McInnes, L., Healy, J., and Astels, S. (2017). HDBCSAN: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11):205.

McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv.1802.03426*, pages 1–63.

Ministério da Saúde (2024). Ministra da saúde defende aliança internacional contra a desinformação em saúde. https://www.gov.br/saude/pt-br/assuntos/noticias/2024/outubro/ministra-da-saude-defende-alianca-internacional-contra-a-desinformacao-em-saude. Acesso em: 15 de mar. de 2025.

Radford, A., Narasimhan, K., and Salimans, T. et al. (2018). Improving language understanding by generative pre-training. *OpenAI preprint Technical Report*, pages 1–12.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-Networks. In Padó, S. and Huang, R., editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pages 3982–3992, China. Association for Computational Linguistics.

Sistema Único de Saúde (2023). Saúde sem boato. https://susconecta.org.br/saude-sem-boato/. Acesso em: 10 de fev. de 2025.

Sistema Único de Saúde (2025). ConecteSUS. https://meususdigital.saude.gov.br/. Acesso em: 20 de fev. de 2025.

World Health Organization (2022). Health topics: Infodemic. https://www.who.int/health-topics/infodemic. Acesso em: 15 de mar. de 2025.

YoshimiTanaka, O., Júnior, M. D., and Cristo, E. B. et al. (2015). Uso da análise de clusters como ferramenta de apoio à gestão no sus. *Saúde e Sociedade*, 24(1):34–45.

Zhong, R., Chacón-Montalván, E. A., and Moraga, P. (2024). Bayesian spatial functional data clustering: applications in disease surveillance. *arXiv preprint arXiv:2407.12633*, pages 1–19.