

Análise Linguística de Comentários de Humanos e de Modelos de Linguagem a Postagens em Comunidades Brasileiras do Reddit

Fernanda Luiza Tobias
Universidade Federal de Minas Gerais
fernanda.tobias@dcc.ufmg.br

Ana Paula Couto da Silva
Universidade Federal de Minas Gerais
ana.coutosilva@dcc.ufmg.br

ABSTRACT

In this work, we analyze the linguistic differences between comments written by humans and those generated by large language models in social support communities on Reddit. Overall, the models tend to better align with the semantic content and linguistic style of the posts and are generally more informative, but their responses are often repetitive and harder to read. In contrast, human comments are typically more concise, diverse, and include personal experiences. Additionally, differences across communities suggest that the type of support being sought influences writing patterns.

KEYWORDS

Plataformas sociais online, Suporte Emocional, Julgamento Social, Modelos de Inteligência Artificial Generativa.

1 INTRODUÇÃO

Por vários anos, as redes sociais online têm sido utilizadas para diversas finalidades, como manter contato com amigos antigos, fazer novas amizades e para a formação de grupos de apoio virtuais, reduzindo o sentimento de solidão e permitindo que os indivíduos busquem apoio emocional e informativo, conselhos e orientações imparciais sobre diversos assuntos [15].

Mais recentemente, como alternativa ao suporte emocional e social oferecido pelos grupos de apoio virtuais, grandes modelos de linguagem vêm sendo amplamente utilizados como ferramenta de aconselhamento psicológico [11]. Por um lado, o uso desses modelos no suporte ao bem estar apresenta benefícios ligados à disponibilidade e à ausência de julgamentos. Por outro, esses recursos enfrentam desafios relacionados à geração de conteúdo nocivo e a inconsistências no estilo de comunicação, além das questões éticas envolvidas [10, 19].

Assim, a busca cada vez maior por aconselhamento *automatizado* em tópicos pessoais, como questões de saúde mental e julgamento de situações do dia a dia, gera a necessidade de uma análise mais aprofundada da qualidade do suporte oferecido por estes modelos. Em outras palavras, é fundamental compreender o quão similares e alinhados os conselhos gerados por modelos de linguagem estão daqueles oferecidos por humanos.

O desafio é ainda mais relevante quando consideramos a língua portuguesa, pois há uma lacuna de estudos que avaliem a capacidade dos modelos de linguagem de responder questões em idiomas diferentes do inglês. Há também uma limitação de recursos técnicos,

como por exemplo, escassez de modelos pré-treinados específicos para o português [6, 17].

Neste contexto, este trabalho analisa as diferenças linguísticas entre o suporte oferecido por seres humanos e aquele gerado por modelos de linguagem, com base em conteúdos publicados em duas comunidades brasileiras do Reddit: (i) r/desabafos, onde usuários relatam experiências pessoais em busca de apoio emocional; e (ii) r/EuSouOBabaca, comunidade em que usuários compartilham situações do cotidiano esperando receber julgamentos sobre o seu comportamento. Três modelos de linguagem foram utilizados para a geração de texto: ChatGPT, Gemini e Sabiá. As análises apresentadas neste artigo representam um passo inicial para compreender as vantagens, limitações e os desafios associados ao uso de modelos automatizados como substitutos (ou complementares) de interações humanas em contextos sensíveis.

2 TRABALHOS RELACIONADOS

A literatura relacionada ao estudo comparativo entre suporte emocional fornecido por humanos e por grandes modelos de linguagem ainda é limitada.

Estudos focados em chatbots específicos para tarefa de suporte social evidenciam a capacidade dos mesmos de promover melhorias em sintomas de sofrimento mental, ansiedade, depressão, burnout [2, 18]. Considerando o uso de grandes modelos de linguagem para mesma tarefa, os autores de [7] destacaram a divergência entre julgamentos do ChatGPT e de humanos, além de inconsistências na decisão do modelo de linguagem no contexto de aconselhamento sobre relacionamentos, usando dados do Reddit. Já os autores em [19], compararam as respostas fornecidas por modelos de linguagem e por humanos para conteúdos compartilhados em comunidades de saúde mental online do Reddit. Além de destacarem as diferenças linguísticas encontradas entre os conjuntos de textos, os autores também discutiram aspectos relacionados à ética, à privacidade e às particularidades das interações humanas.

A principal contribuição do nosso trabalho é caracterizar a linguagem usada por humanos e por modelos de linguagem em diferentes contextos de suporte social online, considerando interações em português.

3 METODOLOGIA

3.1 Conjunto de Dados

O Reddit é uma mídia social on-line multilíngue, fundada em 2005, organizada em subcomunidades por áreas de interesse (*subreddits*). Nesta rede, os usuários discutem diferentes assuntos, através de interações do tipo postagem-comentários, chamadas de *threads*.

Como o objetivo do nosso trabalho é quantificar e entender as diferenças linguísticas entre os conselhos dados por humanos e

conselhos gerados pelos grandes modelos de linguagem, coletamos postagens e comentários de dois subreddits: r/desabafos e r/EuSouOBabaca. A descrição dessas comunidades sugere o compartilhamento de situações cotidianas e aconselhamento de modo geral e cada uma delas possui mais de 50 mil membros que utilizam predominantemente a língua portuguesa. A proposta do subreddit r/desabafos é ser um espaço acolhedor e livre de julgamentos, onde os indivíduos possam desabafar e/ou pedir conselhos. Já no subreddit r/EuSouOBabaca, os usuários são convidados a compartilhar conflitos pessoais, buscando apresentar a perspectiva de todos os envolvidos, a fim de entender se agiram de forma adequada. Em outras palavras, os usuários da primeira comunidade procuram por suporte emocional, enquanto os da segunda comunidade buscam entender o possível julgamento social em relação às situações compartilhadas.

Devido às restrições de pagamentos dos modelos de linguagem utilizados, selecionamos de forma aleatória e sem reposição, 200 postagens para cada um dos subreddits, sendo 100 realizadas em 2022 e 100 realizadas em 2023. A escolha de dois anos busca representar duas janelas distintas de tempo de interações nestas comunidades. No processo de seleção realizado, as publicações escolhidas possuem pelo menos um comentário válido. Embora a estrutura do Reddit permita que os usuários interajam com uma postagem respondendo a outros comentários, em uma estrutura de árvore de discussão ou *thread*, consideramos em nossa análise apenas os comentários do primeiro nível, diretamente associadas à postagem inicial. Essa escolha se justifica pelo fato de que comentários em níveis mais profundos, por serem respostas a outros comentários, contam com um contexto adicional que pode influenciar significativamente seu conteúdo e formato.

O nosso conjunto de dados é formado por 200 postagens e 1.229 comentários (respostas de primeiro nível) relacionados ao r/desabafos, sendo 647 vinculados às postagens de 2022 e 582 às postagens de 2023. Adicionalmente, temos 200 postagens e 1.928 comentários do r/EuSouOBabaca, sendo 976 vinculados às postagens de 2022 e 952 às postagens de 2023. Nossos dados contemplam somente o texto publicado, o subreddit e a data de publicação. Adicionalmente, o Reddit é conhecido por ser uma plataforma em que os usuários são anônimos.

3.2 Grandes Modelos de Linguagem

Para cada publicação da amostra, foram geradas respostas com o auxílio de três modelos de linguagem: (1) ChatGPT, da OpenAI¹, (2) Gemini, da Google², e (3) Sabiá, da Maritaca AI³. Os dois primeiros modelos foram selecionados por serem altamente usados pelo público em geral e o terceiro por ser um modelo treinado com documentos em língua portuguesa.

Os resultados foram obtidos via interface web, utilizando a configuração padrão dos seus hiperparâmetros (temperatura e número máximo de tokens). Para garantir consistência, um novo chat era iniciado para cada publicação analisada. No caso do r/desabafos, não há um tipo específico de interação esperada, ora a postagem consiste apenas em um relato, ora inclui questionamentos. Portanto,

optamos por apenas replicar o conteúdo original da postagem no *prompt*, sem fornecer instruções adicionais ao modelo. Essa escolha visa preservar a ambiguidade e espontaneidade característica do subreddit, permitindo que o modelo interprete livremente a intenção do autor da publicação. O objetivo de postagens no r/EuSouOBabaca é obter uma classificação do comportamento do autor com base em sua percepção dos eventos narrados. Assim, foi definido o seguinte *prompt* como entrada para os modelos: *Você é um membro do subreddit r/EuSouOBabaca. Comente e classifique a submissão abaixo. Use uma das seguintes flairs: Babaca, Não é o babaca, Ninguém é babaca, Todo mundo é babaca, Não avaliado, Falta informação, Meta, Trouxa. [conteúdo original da postagem]*

Cada modelo gerou uma resposta vinculada a cada uma das 400 postagens pertencentes ao nosso conjunto de dados.

3.3 Análise Linguística

A análise linguística realizada neste trabalho se baseou em [19]. Para cada tipo de comentário vinculado à uma postagem, isto é, comentários feitos por: (i) humanos; (ii) GPT; (iii) Gemini e; (iv) Sabiá, calculamos a distribuição de valores das métricas e realizamos comparações estatística entre todos os pares (humano-modelo). Foram aplicados testes estatísticos para diferença de média (teste t) e distribuição de probabilidade (Kolmogorov-Smirnov).

As métricas consideradas foram organizadas em cinco dimensões: psicolinguística, estrutura linguística, estilo linguístico, adaptabilidade linguística e detecção de posicionamento.

3.3.1 Análise Psicolinguística. Analisamos as propriedades psicolinguísticas dos tipos de comentários a partir da utilização da ferramenta LIWC (*Linguistic Inquiry and Word Count*) para português (*Brazilian Portuguese LIWC 2007 Dictionary*⁴). Para cada comentário, calculamos o número de palavras em cada uma das 64 categorias e normalizamos pelo número total de palavras.

3.3.2 Estrutura Linguística. Para compreender melhor a estrutura linguística dos comentários analisados, foram utilizadas três métricas: *verbosidade*, *legibilidade* e *repetibilidade*. A primeira mede o total de palavras em um texto. Textos com maior verbosidade tendem a ser mais elaborados. Pesquisas anteriores indicam associação positiva entre verbosidade e suporte social informativo [3], assim como engajamento de outros usuários [12]. A segunda métrica avalia o grau de dificuldade associado à leitura de um texto e é calculada a partir do Índice de Coleman-Liau [4]. Este índice estima o total de anos de escolaridade necessário para compreender um texto na primeira leitura. Quanto maior o valor do índice, maior a estimativa de anos de estudo necessários para entender o texto.

Para medir o grau de *repetibilidade* em um texto, analisamos a proporção de palavras não únicas em um comentário que é dada pela razão entre a frequência total de palavras repetidas e o número total de palavras. Um comentário com repetibilidade igual a zero não apresenta repetições, enquanto um valor igual a 1 indica que todas as palavras foram utilizadas mais de uma vez. De modo geral, a repetição pode desempenhar funções distintas em um texto, contribuindo para compreensão, reforçando conceitos chaves, ou levando à fadiga do leitor [21].

¹<https://chat.openai.com>

²<https://gemini.google.com>

³<https://chat.maritaca.ai>

⁴<http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>

3.3.3 Estilo Linguístico. O estilo linguístico é representado pelas escolhas linguísticas recorrentes que caracterizam a forma como uma pessoa se expressa verbalmente ou por escrito. Neste trabalho, usamos como métricas para esta análise o *Índice Categórico-Dinâmico (CDI)* [14] e a *formalidade*. O CDI classifica o estilo de linguagem em uma escala entre categórico e dinâmico, a partir da estrutura gramatical. A linguagem categórica é mais sofisticada, combinando pensamento abstrato e capacidade cognitiva. Já a linguagem dinâmica ou narrativa é marcada pela fluidez e pessoalidade. Valores mais altos de CDI estão vinculados ao estilo categórico. A frequência das classes gramaticais necessárias para o cálculo do CDI foram calculadas a partir do dicionário LIWC para português.

Já a *formalidade* textual reflete o nível de polidez e o uso (ou não) de expressões coloquiais em um texto [20]. Um maior nível de formalidade na comunicação pode aumentar a credibilidade e influência exercida pela mensagem [9]. Em ambientes online, usuários tendem a seguir o nível de formalidade estabelecido na conversa e quanto maior a frequência de iteração de um indivíduo, menor a formalidade esperada [13]. Usamos o modelo pré treinado *formality-classifier-mdeberta-v3-base* proposto em [8] para classificar os comentários como formais, informais ou neutros.

3.3.4 Adaptabilidade Linguística. Nesta dimensão, o principal objetivo é analisar a capacidade de um indivíduo em ajustar seu uso de linguagem conforme o contexto no qual a comunicação está sendo estabelecida. Assim, verificamos o nível de *similaridade semântica* e *acomodação de estilo linguístico* entre o comentário e a postagem relacionada, bem como medimos a *diversidade/criatividade* existente entre os comentários.

Mais especificamente, a *similaridade semântica* é medida a partir da similaridade do cosseno entre as representações vetoriais do comentário e da postagem. Para obter as representações vetoriais dos textos, utilizamos o modelo pré treinado *paraphrase-multilingual-MiniLM-L12-v2* [16]. Valores próximos de 1 indicam que há alta semelhança no conteúdo semântico dos dois textos. A *acomodação de estilo linguístico* mede a similaridade no uso de palavras funcionais, ou seja, palavras que possuem pouco significado semântico e que têm papel estrutural e gramatical na construção das frases. Para avaliar esta métrica, usamos a similaridade do cosseno entre os vetores de frequência normalizada de palavras funcionais relativo a cada comentário e a publicação. Valores próximos de 1 apontam que o comentário está alinhado ao estilo da publicação.

Por fim, a partir da representação vetorial dos comentários obtidas usando o mesmo modelo anterior, calculamos o vetor médio destas representações (centróide) para cada conjunto (humanos, *GPT*, *Gemini* e *Sabiá*). Para cada par comentário-centróide do conjunto calculamos a distância do cosseno. Quanto maior a distância, maior *diversidade/criatividade* dos textos de um dado conjunto

3.3.5 Detecção de Posicionamento. Esta dimensão avalia se um indivíduo se posiciona a favor ou contra um determinado tópico. Esse posicionamento é influenciado por diferentes fatores sociais, culturais e sociais [1]. Para definir o posicionamento de um comentário em relação à sua postagem, aplicamos o modelo de regressão baseado no BERTimbau e proposto em [5]. O *stance score* varia entre -1 e 1, onde valores mais positivos indicam que a resposta está em maior concordância com a publicação.

4 RESULTADOS

A seguir apresentamos os resultados preliminares do nosso trabalho. As análises serão descritas para cada subreddit e cada dimensão de análise linguística em estudo.

4.1 Suporte Emocional: r/desabafos

4.1.1 Análise Psicolinguística. Das 64 categorias analisadas, há diferenças significativas nas comparações humano-modelos em 58 (91%). Para simplicidade e generalização, exemplificamos algumas das categorias que divergem na comparação com os três modelos. Nos comentários escritos por humanos, é mais comum o uso da primeira pessoa do singular e tempo verbal do passado, que podem ser associados a narrativas pessoais. Paralelamente, há mais termos relacionados ao assentimento e à negação, indicando que humanos são mais propícios a demonstrar algum tipo de posicionamento parcial do que os modelos. Nos comentários escritos pelos modelos, há maior presença de palavras funcionais, artigos, conjunções, preposições e quantificadores, que estão vinculados a uma escrita mais articulada. Também é recorrente a manifestação de emoções positivas.

4.1.2 Estrutura Linguística. A *verbosidade* dos comentários é maior para os textos gerados por modelos de linguagem. O *Gemini* gerou comentários mais elaborados, com média de 441 palavras. Esse valor equivale a cerca de dez vezes o comprimento médio dos comentários escritos por humanos (43 palavras). Em média, os comentários do *GPT* e de *Sabiá* possuem 329 e 343 palavras, respectivamente. Adicionalmente, em geral, é mais fácil compreender um comentário escrito por humanos do que um comentário gerado por um modelo. A *legibilidade* média estimada é de 7 anos de estudo para entender os comentários feitos por humanos, 10, 11 e 12 anos para entender os comentários gerados pelos modelos *GPT*, *Gemini* e *Sabiá*, respectivamente.

A *repetibilidade* é consideravelmente menor nos comentários feitos por humanos, onde, em média, 80% das palavras são usadas uma única vez. Para os comentários gerados por modelos, mais da metade do texto consiste em palavras repetidas.

4.1.3 Estilo Linguístico. Os valores médios de *CDI* indicam que os comentários gerados pelo *GPT* possuem estilo mais dinâmico do que os comentários feitos por humanos. Por outro lado, a linguagem categórica é mais presente nos textos do *Gemini* e *Sabiá* do que dos humanos.

Proporcionalmente, há uma maior tendência de *informalidade* nos comentários elaborados por humanos, sendo 88% dos comentários classificados como informais. Esse percentual corresponde a 64% para o *GPT*, 55% para o *Gemini* e apenas 7% para o *Sabiá* que tem os comentários predominantemente classificados como neutros (73%).

4.1.4 Adaptabilidade Linguística. Os comentários feitos por humanos possuem *similaridade semântica* inferior aos dos modelos de linguagem. A similaridade média para esse primeiro conjunto de textos é de 0,34. Dentre os modelos, o *Gemini* produziu, na média, os textos mais similares ao conteúdo publicado (0,56), seguido por *GPT* (0,53) e *Sabiá* (0,51).

Os modelos de linguagem geraram comentários mais *acomodados ao estilo linguístico* das publicações que os humanos. O valor médio

dessa métrica é de 0,91 para textos escritos por humanos e superior a 0,96 para comentários gerados pelos modelos.

Os comentários dos humanos são significativamente mais *diversos e criativos* que os comentários dos modelos de linguagem. A medida de criatividade média equivale a 0,50 para textos escritos por humanos, 0,32 para *GPT* e 0,35 para *Gemini* e *Sabiá*.

4.1.5 Detecção de Posicionamento. Humanos e modelos tendem a concordar com o autor da postagem. O valor médio do *Stance Score* é 0,44 para comentários escritos por humanos, 0,46 para o *Gemini*, 0,48 para o *GPT* e 0,52 para o *Sabiá*. As diferenças não são estatisticamente significativas.

4.2 Julgamento Social: r/EuSouOBababaca

4.2.1 Análise Psicolinguística. Nas comparações humano-modelos, há diferenças significativas na frequência de palavras em 61 (95%) das 64 categorias existentes. Exemplificamos algumas das divergências que ocorrem em relação aos três modelos. Os comentários dos humanos usam mais a primeira pessoa do singular, o que está ligado a uma escrita com narrativa pessoal. Além disso, citam termos relacionados à assentimento e audição com maior frequência. Esses atributos podem indicar posicionamento favorável e demonstração de acolhimento através da escuta, por exemplo. Já nos comentários escritos pelos modelos, há uma maior presença de palavras categorizadas como tristeza, raiva e ansiedade, que sugerem uma tentativa de conexão emocional com o autor da postagem. Adicionalmente, o uso frequente de palavras funcionais, artigos, preposições, advérbios, conjunções, indica um estilo de escrita mais estruturada e coesa nesse conjunto de comentários.

4.2.2 Estrutura Linguística. A *verbosidade* é menor nos comentários escritos por humanos. Em média, esses comentários contêm 51 palavras. Em contraste, os comentários gerados pelos modelos de linguagem apresentam maior extensão: aproximadamente 178 palavras no caso do *GPT*, 196 no *Sabiá* e 254 no *Gemini* (cerca de 5 vezes o comprimento médio dos comentários feitos por humanos).

Os valores médios de *legibilidade* estimam que apenas 2 anos de escolaridade são suficientes para compreender os comentários escritos por humanos. Por outro lado, *GPT* e *Gemini* requerem 10 anos e *Sabiá* 12 anos de estudo.

A *repetibilidade* nos comentários gerados por modelos equivale, em média, ao dobro do valor dessa métrica nos textos escritos por humanos. No segundo conjunto, cerca de 22% das palavras foram repetidas ao menos um vez. Para cada um dos modelos de linguagem, a frequência de palavras repetidas é superior a 50%.

4.2.3 Estilo Linguístico. Os valores do *CDI* sugerem que os modelos de linguagem geraram comentários com uma narrativa mais fluida que os humanos. Por outro lado, há mais *informalidade* nos comentários escritos por humanos. 84% dos textos de humanos são classificadas como informais. Esse percentual é estatisticamente inferior para textos gerados pelos modelos e equivale a 79% dos comentários do *Gemini*, 71% do *GPT* e apenas 12% do *Sabiá*.

4.2.4 Adaptabilidade Linguística. A *similaridade semântica* é significativamente inferior nos comentários escritos por humanos. Em média, o *Gemini* possui maior alinhamento semântico (0,52), seguido do *GPT* (0,50), do *Sabiá* (0,48) e dos humanos (0,27).

Em geral, a *acomodação ao estilo linguístico* é alta. Os valores são ainda maiores para os comentários gerados pelos modelos. O valor médio dessa métrica é superior a 0,97 para os modelos e igual a 0,94 para humanos.

Os comentários de humanos apresentam maior nível de *diversidade* entre si que as respostas geradas pelos modelos. Em média, a medida de diversidade para comentários feitos por humanos é 0,47, para *GPT*, *Gemini* e *Sabiá* é próximo de 0,30.

4.2.5 Detecção de Posicionamento. Os valores do *stance score* são mais baixos para os comentários dos humanos. Embora o valor médio seja positivo (0,25), a distribuição também assume valores negativos (primeiro quartil = -0,28), sinalizando que humanos adotam, com mais frequência, um posicionamento contrário ou menos favorável ao conteúdo inicial. Para os três modelos, o valor médio é cerca de 0,51 e o primeiro quartil é superior a 0,25.

5 DISCUSSÃO FINAL E TRABALHOS FUTUROS

De modo geral, os modelos de linguagem demonstram vantagens em alguns dos aspectos avaliados e os humanos em outros.

Os comentários gerados por modelos conseguem se adaptar melhor ao conteúdo semântico e ao estilo linguístico, aumentando a conexão com o autor da postagem. Além disso, por serem textos mais longos, são mais informativos. Porém, são repetitivos, e requerem maior legibilidade, podendo dificultar a compreensão por parte de quem está buscando o aconselhamento.

Os comentários feitos por humanos são mais objetivos e menos redundantes. Embora de forma mais sutil que os modelos, os humanos também alinham o estilo linguístico ao da postagem. Uma das principais vantagens dos humanos é a capacidade de produzir textos diversos, que fogem de um padrão genérico. Isso pode estar relacionado, inclusive, com a presença de termos ligados a relatos pessoais e posicionamento.

Há também algumas diferenças nos resultados observados nas duas comunidades que podem estar relacionados ao tipo de suporte buscado. Primeiro, os comentários dos modelos são menos elaborados e de mais fácil compreensão no r/EuSouOBababaca que no r/desabafos. A hipótese é que as orientações mais diretas sobre o tipo de interação esperado nessa comunidade contribuam para geração de respostas mais simples e objetivas. Adicionalmente, humanos assumem um posicionamento contrário ou menos favorável com mais frequência no r/EuSouOBababaca. O que é razoável tendo em vista que as postagens envolvem conflitos morais.

Como continuação deste estudo devemos ampliar a quantidade de postagens e aprimorar a geração de respostas usando os modelos de linguagem. Os principais fatores de melhoria estão relacionados a uma investigação mais aprofundada sobre a construção dos *prompts* e geração de mais de uma resposta por postagem. Para a escalabilidade desse processo generativo, será importante utilizar a API dos modelos. Além disso, pretendemos realizar uma análise qualitativa de um conjunto de postagens e comentários, tanto de humanos quanto de modelos de linguagem, a fim de enriquecer a interpretação dos resultados observados.

Agradecimentos: Este trabalho foi parcialmente financiado pela FAPEMIG, CAPES e CNPq.

REFERÊNCIAS

- [1] Abeer ALDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing Management* 58, 4 (2021), 102597. <https://doi.org/10.1016/j.ipm.2021.102597>
- [2] Gumhee Baek, Chiyoungh Cha, and Jin-Hui Han. 2025. AI chatbots for psychological health for health professionals: Scoping review. *JMIR Human Factors* 12, 1 (2025), e67682. <https://doi.org/10.2196/67682>
- [3] Langtao Chen, Aaron Baird, and Detmar Straub. 2020. A linguistic signaling model of social support exchange in online health communities. *Decision Support Systems* 130 (2020), 113233. <https://doi.org/10.1016/j.dss.2019.113233>
- [4] Meri Coleman and Ta Lin Liao. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology* 60, 2 (1975), 283. <https://doi.org/10.1037/h0076540>
- [5] Gustavo F. Cunha and Ana Paula C. da Silva. 2025. *Caracterizando Polarização nas Eleições Brasileiras de 2018 e 2022: Uma Análise das Discussões no Reddit com um Modelo de Regressão para Stance Detection*. MSI. Universidade Federal de Minas Gerais. <https://monografias.dcc.ufmg.br/monografia/caracterizando-polarizacao-nas-eleicoes-brasileiras-de-2018-e-2022-uma-analise-das-discussoes-no-reddit-com-um-modelo-de-regressao-para-stance-detection/>
- [6] Luiz Fernando de Lima and Renata Mendes Araujo. 2023. A call for a research agenda on fair NLP for Portuguese. In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*. SBC, 187–192. <https://sol.sbc.org.br/index.php/stil/article/view/25450>
- [7] Haonan Hou, Kevin Leach, and Yu Huang. 2024. ChatGPT Giving Relationship Advice—How Reliable Is It?. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 610–623. <https://doi.org/10.1609/icwsm.v18i1.31338>
- [8] LenDigLearn. [n. d.]. Formality Classifier MDeBERTa v3 Base. Retrieved 13 de Julho de 2025 from <https://huggingface.com/LenDigLearn/formality-classifier-mdeberta-v3-base>
- [9] Elizabeth Linos, Jessica Lasky-Fink, Chris Larkin, Lindsay Moore, and Elspeth Kirkman. 2024. The formality effect. *Nature Human Behaviour* 8, 2 (2024), 300–310. <https://doi.org/10.1038/s41562-023-01761-z>
- [10] Zilin Ma, Yiyang Mei, and Zhaoyuan Su. 2024. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. In *AMIA Annual Symposium Proceedings*, Vol. 2023. 1105. <https://doi.org/10.48550/arXiv.2307.15810>
- [11] Birger Moëll. 2024. Comparing the efficacy of GPT-4 and chat-gpt in mental health care: A blind assessment of large language models for psychological support. *arXiv preprint arXiv:2405.09300* (2024). <https://doi.org/10.48550/arXiv.2405.09300>
- [12] Fahad Mansoor Pasha, Fatima Habib, Komal Kamran, Akbar Azam, Zeeshan Ali, and Dildar Hussain. 2025. Let the Customers Speak Their Hearts Out: The Role of Verbosity and Emotions in Online Viewer-to-Viewer Engagement. *Human Behavior and Emerging Technologies* 2025, 1 (2025), 6282833. <https://doi.org/10.1155/hbe2/6282833>
- [13] Ellie Pavlick and Joel Tetreault. 2016. An empirical analysis of formality in online communication. *Transactions of the association for computational linguistics* 4 (2016), 61–74. https://doi.org/10.1162/tacl_a_00083
- [14] James W Pennebaker, Cindy K Chung, Joey Frazee, Gary M Lavergne, and David I Beaver. 2014. When small words foretell academic success: The case of college admissions essays. *PLoS one* 9, 12 (2014), e115844. <https://doi.org/10.1371/journal.pone.0115844>
- [15] Julie Prescott, Amy Leigh Rathbone, and Terry Hanley. 2020. Online mental health communities, self-efficacy and transition to further support. *Mental Health Review Journal* 25, 4 (2020), 329–344. <https://doi.org/10.1108/MHRJ-12-2019-0048>
- [16] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. <http://arxiv.org/abs/1908.10084>
- [17] Luiz Rodrigues, Cleon Xavier, Newarney Costa, Hyan Batista, Luiz Felipe Bag-nhuk Silva, Wesley Chaleghi de Melo, Dragan Gasevic, and Rafael Ferreira Mello. 2025. LLMs Performance in Answering Educational Questions in Brazilian Portuguese: A Preliminary Analysis on LLMs Potential to Support Diverse Educational Needs. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*. Association for Computing Machinery, 865–871. <https://doi.org/10.1145/3706468.3706515>
- [18] Sahand Sabour, Wen Zhang, Xiyao Xiao, Yuwei Zhang, Yinhe Zheng, Jiaxin Wen, Jialu Zhao, and Minlie Huang. 2023. A chatbot for mental health support: exploring the impact of Emohaa on reducing mental distress in China. *Frontiers in digital health* 5 (2023), 1133987. <https://doi.org/10.48550/arXiv.2209.10183>
- [19] Koustuv Saha, Yoshee Jain, and Munmun De Choudhury. 2025. Linguistic Comparison of AI- and Human-Written Responses to Online Mental Health Queries. <https://doi.org/10.48550/arXiv.2504.09271>
- [20] Multimodal strategies for balancing formality and informality: The role of kaomoji in online comment-reply interactions. 2022. Kaneyasu, Michiko. *Internet Pragmatics* 5, 1 (2022), 143–164. <https://doi.org/10.1075/ip.00071.kan>
- [21] Elena Tikhonova, Daria Mezentseva, and Peter Kasatkin. 2024. Text Redundancy in Academic Writing: A Systematic Scoping Review. *Journal of Language and Education* 10, 3 (39) (2024), 128–160. <https://doi.org/10.17323/jle.2024.23747>