

15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

ANÁLISE DOS LIMITES E POTENCIALIDADES DA INTELIGÊNCIA ARTIFICIAL NO RECONHECIMENTO DA VARIAÇÃO LINGUÍSTICA: UM OLHAR SOCIOLINGUÍSTICO

JULIANA DA SILVA ALVES DOS SANTOS¹, MARIA BEATRIZ GAMEIRO CORDEIRO²

¹ Graduanda em Licenciatura em Letras, bolsista do PIBIFSP, IFSP, Campus Sertãozinho, alves.juliana1@aluno.ifsp.edu.br.

² Orientadora, professora e coordenadora na graduação em Licenciatura em Letras, IFSP, Campus Sertãozinho, mbg@ifsp.edu.br.

Área do conhecimento (tabela CNQP): 8.02.00.00-1 Letras

RESUMO:

O objetivo desse trabalho é apresentar os resultados de uma pesquisa que analisa se diferentes sites de Inteligência Artificial (doravante IA), como o “ChatGPT” e o “Gemini” – que geram respostas por meio do modelo *chatbot* guiado pela IA Generativa – reconhecem variações linguísticas regionais, especificamente, os significados e usos de expressões semânticas utilizados em diferentes contextos linguísticos por falantes de diversas localidades do país. O estudo também busca analisar se as IAs possuem a potencialidade de apresentar reflexões sobre o nível de formalidade, preconceitos, “correções”, dentre outros aspectos, em consonância à teoria Sociolinguística. Apesar de estudos que relatam inconsistências nos sistemas, parte-se da hipótese de que haja alto grau de “acerto”, tanto no reconhecimento das variações, bem como na explicação da utilização dos usos e alternância de utilização, visto que o banco de dados da IA se caracteriza por ser complexo e controlado por algoritmos poderosos. Para tanto, essa pesquisa parte de testes e questionamentos qualitativos feitos às IAs, já que o objetivo não é quantificar os resultados, e sim realizar um trabalho interpretativo de análises minuciosas das respostas fornecidas pela plataforma. Os resultados iniciais indicam que o ChatGPT demanda uma maior contextualização do uso e sentido de determinada expressão em uma pergunta a ser testada para fornecer os resultados esperados, ao passo que o Gemini, além de reconhecer as expressões com mais facilidade, também fornece considerações mais complexas as relações sociais que envolvem o uso da expressão. Diante desses resultados iniciais, depreende-se que o grau de precisão das respostas do Gemini é maior do que o ChatGPT no que concerne às variações regionais brasileiras, pois demonstra maior precisão e contextualização do que o ChatGPT, que, além de se limitar à pergunta, no geral, fornece uma resposta que não detalha a complexa relação entre certos usos linguísticos e fatores sociais.

PALAVRAS-CHAVE: Variações linguísticas; Inteligência Artificial; Sociolinguística; banco de dados; ChatGPT; Gemini.

ANALYSIS OF THE LIMITS AND POTENTIALITIES OF ARTIFICIAL INTELLIGENCE IN RECOGNIZING LINGUISTIC VARIATION: A SOCIOLINGUISTIC PERSPECTIVE

ABSTRACT: The objective of this work is to present the results of a study that analyzes whether different Artificial Intelligence (AI) platforms, such as "ChatGPT" and "Gemini"—which generate responses through a chatbot model guided by Generative AI—recognize regional linguistic variations, specifically the meanings and uses of semantic expressions employed in different linguistic contexts by speakers from various parts of the country. The study also seeks to examine whether these AIs have the potential to reflect on aspects such as the level of formality, biases, "corrections," among other factors, in accordance with Sociolinguistic theory. Despite studies reporting inconsistencies in these systems, the hypothesis is that there is a high degree of "accuracy," both in recognizing variations and in explaining the use and alternation of usage, given that the AI databases are characterized by being complex and

controlled by powerful algorithms. To this end, this research is based on qualitative tests and questions posed to the AIs, as the goal is not to quantify the results but rather to conduct an interpretive work of thorough analysis of the responses provided by the platforms. Initial results indicate that ChatGPT requires greater contextualization of the use and meaning of a given expression in a test question to deliver the expected results, whereas Gemini, in addition to more easily recognizing the expressions, also provides more complex considerations of the social relations involved in the use of the expression. Based on these initial findings, it can be inferred that Gemini's degree of accuracy is higher than that of ChatGPT regarding Brazilian regional variations, as it demonstrates greater precision and

15° CONICT 2024

contextualization than ChatGPT, which, besides being limited to the question, generally provides a response that does not detail the complex relationship between certain linguistic uses and social factors.

KEYWORDS: Linguistic variations; Artificial Intelligence; Sociolinguistics; database; ChatGPT; Gemini.

INTRODUÇÃO

Com o avanço da utilização de plataformas como o ChatGPT, ferramenta que fornece respostas para os usuários por meio de um banco de dados guiado pela Inteligência Artificial (doravante IA), pesquisar sua real eficiência se tornou um objetivo relevante para a sociedade. De acordo com Mitsunaga (2023), o avanço da tecnologia na sociedade fez com que a IA se tornasse um aspecto comum na vida de muitas pessoas, portanto, um projeto que analise as respostas concernentes sobre variação linguística torna-se relevante. Timpone e Guidi (2023) explicam que plataformas como o ChatGPT funcionam com base na IA Generativa, que possui a capacidade de gerar respostas em modelo *chat* por meio de algoritmos que complementam e estruturam seu banco de dados. Portanto, o objetivo deste trabalho é avaliar se plataformas como o ChatGPT e Gemini (plataformas de IA com a mesma funcionalidade) são capazes de demonstrar uma análise coerente, com fundamentos linguísticos, sociolinguísticos e variacionistas a partir de testes estruturados, com o intuito de verificar se há qualquer tipo de preconceito linguístico ou social. Bagno (2007) contribui com essa discussão quando explica sobre o mito existente na sociedade de que o único conceito de língua é aquele referente ao padrão (uma norma de pronúncias, palavras e regras eleitas como corretas), estabelecido ao longo da tradição escrita, literária e escolar. Essa definição ignora as variedades dialetais e aponta que há apenas uma forma “certa” de falar, que corresponderia à escrita, contribuindo assim para o aumento de preconceitos e desconhecimento das variantes linguísticas. Assim, considerando que na sociedade já existem mitos e preconceitos a respeito da língua em uso, testar os limites e potencialidades da IA sobre a temática mostra-se fundamental para a sociedade contemporânea, em que seus usos é crescente.

MATERIAL E MÉTODOS

Para verificar, testar, confirmar ou refutar a hipótese de que as plataformas de IA (em específico, ChatGPT e Gemini) são capazes de averiguar ou não variações linguísticas presentes na língua em uso da sociedade brasileira, elaboramos as categorias de análise listadas a seguir, que corroboram para sintetizar pontos importantes sobre as conclusões que esperamos obter. Além disso, possibilitam que haja uma comparação da funcionalidade das plataformas ChatGPT e Gemini de forma mais específica. Quanto à metodologia, a pesquisa segue a abordagem qualitativa, visto que seu objetivo não é produzir uma quantidade numérica grande de resultados, mas sim uma análise significativa e minuciosa das respostas geradas pela IA, observando os possíveis valores sociais expostos, a concepção linguística que fundamenta os textos-respostas e a partir dessas relações, construir sentidos, como define Knechel (2014). Portanto, trata-se de um trabalho interpretativo que parte da pesquisa bibliográfica como recurso metodológico primordial, constituindo a base teórica para as reflexões desenvolvidas. Juntamente com a pesquisa bibliográfica, dados são levantados para a análise por meio de perguntas feitas às IAs.

Tabela 1. Categorias de análise

Maior precisão: MP	Menor precisão: RP
--------------------	--------------------

Apresenta exemplos de variação: EX	Não apresenta exemplos de variação: NE
Especifica as regiões: ER	Não especifica as regiões: NR
Tece considerações sobre níveis de formalidade e uso: TC	Não tece considerações sobre níveis de formalidade e uso: NT

Fonte: Elaboração própria

RESULTADOS E DISCUSSÃO

Após a sondagem inicial nas plataformas e elaboração das categorias de análise, reunimos expressões retiradas de falas jornalísticas (selecionadas de forma aleatória) e de pesquisas bibliográficas que continham expressões linguísticas correlatas a fatores sociais. Até o momento, foram realizados 25 testes, porém, devido às limitações de espaço, somente será apresentada uma ilustração de como esses testes e análises têm sido implementados. A seguir, apresentam-se duas tabelas com um mesmo questionamento feitos às diferentes ferramentas: "Quais são os significados de "negrinho" na linguagem popular nas diferentes regiões do Brasil?", as quais evidenciam a forma como as respostas são criadas pelas diferentes plataformas de inteligência artificial generativa:

Tabela 2: Respostas dada pela plataforma " ChatGPT"

Categorias de análise:	MP	RP	EX	NE	ER	NR	TC	NT
ChatGPT			x		x		x	
Gemini	x		x	x			x	

O termo "negrinho" na linguagem popular brasileira pode ter diferentes significados e conotações dependendo da região do país. A seguir, apresento alguns desses significados regionais:

1.Rio Grande do Sul:

Negrinho do Pastoreio: Na cultura gaúcha, "Negrinho do Pastoreio" é uma figura do folclore. É um mito de um escravo negro que, após ser injustamente castigado e morrer, torna-se um protetor dos perdidos nas pampas. Sua lenda é amplamente conhecida e tem um tom de respeito e reverência.

2.Nordeste:

Em algumas regiões do Nordeste, "negrinho" pode ser usado de maneira carinhosa ou afetuosa para se referir a crianças negras. No entanto, é importante notar que o contexto é essencial, pois o termo pode ser mal interpretado dependendo do tom e da intenção.

3.Sudeste:

Em áreas urbanas do Sudeste, como São Paulo e Rio de Janeiro, o termo pode ser utilizado de forma semelhante ao uso no Nordeste, mas também pode ser visto como pejorativo dependendo do contexto e da relação entre as pessoas envolvidas. O uso do termo sem a devida sensibilidade pode ser considerado ofensivo.

4.Gíria e Expressões:

Em algumas gírias urbanas, "negrinho" pode ser usado de maneira afetuosa entre amigos, similar ao uso de "brother" ou "mano", independentemente da cor da pele. Esse uso, no entanto, é restrito a certos grupos e contextos informais.

5.Pejorativo:

Infelizmente, em diversas regiões, "negrinho" pode ser usado de maneira depreciativa para se referir a pessoas negras, carregando conotações racistas. O uso pejorativo é uma manifestação de preconceito e é amplamente condenado.

6.Outros Contextos Regionais:

Em algumas regiões específicas ou em contextos culturais particulares, "negrinho" pode ter significados adicionais, geralmente ligados às tradições locais e ao uso coloquial da linguagem. É importante sempre considerar o contexto, o tom e a relação entre as pessoas ao utilizar termos como "negrinho", dado que a linguagem é sensível e carregada de conotações culturais e históricas. O uso respeitoso e consciente é fundamental para evitar mal-entendidos e ofensas.

Fonte: Elaboração própria

Tabela 3. Tabela 3. Respostas dada pela plataforma "Gemini"

O termo "negrinho" na linguagem popular brasileira possui diversos significados e conotações, que variam de acordo com a região, contexto social e até mesmo intenção do falante. É importante analisar cada caso em específico para compreender o sentido pretendido, pois o termo pode ter caráter discriminatório e ofensivo, dependendo do uso.

1. Apelido carinhoso:

Em algumas regiões, principalmente no sul do Brasil, "negrinho" pode ser utilizado como um apelido carinhoso entre familiares e amigos próximos, especialmente para pessoas negras ou com pele mais escura. É crucial considerar a relação entre as pessoas envolvidas e o contexto em que o termo é usado para determinar se há ou não intenção de ofender.

Exemplo: Um pai carinhosamente chamando seu filho negro de "negrinho".

2. Descrição da cor da pele:

Em um contexto neutro e sem intenção de ofender, "negrinho" pode ser usado para descrever a cor da pele de uma pessoa negra, geralmente de forma informal. É importante evitar o uso excessivo ou repetitivo do termo, pois isso pode soar como estereótipo ou até mesmo como uma microagressão.

Exemplo: "A menina com cabelo preto e pele negrinha estava brincando no parque." **3. Termo discriminatório:**

Em muitos casos, principalmente em contextos formais ou quando usado por pessoas que não têm relação próxima com o indivíduo, "negrinho" pode ser considerado um termo discriminatório e ofensivo. Sua utilização pode reforçar estereótipos negativos sobre pessoas negras e perpetuar o racismo estrutural.

Exemplo: Uma pessoa fazendo um comentário racista sobre um jogador de futebol negro, chamando de "negrinho".

4. Conotações regionais:

Em algumas regiões do Brasil, "negrinho" pode ter conotações específicas que não se aplicam em outras áreas do país. Por exemplo, no Nordeste, o termo pode ser usado para se referir a um tipo específico de doce.

Exemplo: "Comprei um negrinho para comer depois do almoço." (Nordeste) **5.**

Evitar o uso:

Devido ao seu potencial discriminatório e às diversas interpretações possíveis, é recomendável evitar o uso do termo "negrinho" em qualquer contexto. Opções mais adequadas e respeitosas para se referir a pessoas negras incluem:

- Negro/Negra: Termos neutros e diretos para se referir à raça de uma pessoa.
- Pessoa negra: Uma forma mais completa de se referir a uma pessoa negra, reconhecendo sua individualidade.
- Afro-brasileiro/Afro-brasileira: Termos que reconhecem a herança africana da pessoa e sua identidade cultural.

É importante lembrar que a linguagem é um instrumento poderoso que pode construir ou destruir pontes. Usar termos discriminatórios, como "negrinho", contribui para a perpetuação do racismo e da exclusão social. Optar por uma linguagem mais inclusiva e respeitosa é fundamental para construir uma sociedade mais justa e igualitária.

Fonte: Elaboração própria

Uma rápida comparação entre a explicação dada pelas duas plataformas já indica o maior grau de detalhamento e precisão nas respostas dadas pelo Gemini do que as fornecidas pelo ChatGPT. A própria diferença no número de linhas usadas para tais explicações ratifica essa concisão e maior objetividade, que, na prática, poderiam configurar “um reducionismo teórico”. Logo no parágrafo introdutório que antecede a lista de significados dados, o ChatGPT apenas menciona que há “diferentes significados e conotações dependendo da região do país”, já o Gemini especifica, de início, “o caráter discriminatório e ofensivo” que a expressão pode gerar. Tais diferenças na precisão e na profundidade da resposta nos levam a deduzir que, no que tange à variação linguística, o Gemini é mais preciso. No que tange aos preceitos Sociolinguísticos, (Tarallo, 1985) que preconizam que um uso linguístico é condicionado por fatores linguísticos e sociais, o ChatGPT lista os usos do termo por região, evidenciando que a expressão apresenta conotações distintas dependendo da localidade do falante, já o Gemini faz um comentário genérico sobre os usos regionais, citando o significado de “doce” no Nordeste. Nesse sentido, é possível concluir que as bases de dados consideram os preceitos sociolinguísticos.

CONCLUSÕES

Os testes preliminares realizados até momento indicam que ambas as plataformas são capazes de tecer considerações a respeito do comando testado. Porém, há um direcionamento distinto em relação às respostas dadas; o ChatGPT exemplifica as regiões do Brasil, conforme solicitado no comando, mas suas considerações em relação aos significados das expressões são mais limitadas, pois, retringe-se às especificações de cada região, sem tecer considerações sobre os usos, conforme demonstrado pelo termo “negrinho”. Já o Gemini, do Google, apesar de não organizar sua resposta conforme o comando solicitado, explicitando os significados de acordo com cada região como faz o ChatGPT, realiza ponderações a respeito dos usos, se mostra mais preciso e completo, visto que em cada uso, a plataforma é capaz de dizer sobre os sentidos e fazer indicações a respeito de questões sociais. Com testes dessa natureza, este trabalho pretendeu ilustrar como as IAs reconhecem usos e variações em consonância à teoria Sociolinguística (Calvet, 2002), que considera a variação sistemática, universal, passível de sistematização e ligada a fatores sociais, como região, classe social, dentre outros fatores.

CONTRIBUIÇÕES DOS AUTORES

Santos, Juliana da Silva Alves dos: Conceitualização, curadoria de dados, análise de dados, recebimento de financiamento, pesquisa, metodologia, design de apresentação de dados, redação do manuscrito original.

Cordeiro, Maria Beatriz Gameiro: Conceitualização, curadoria de dados, análise de dados, pesquisa, metodologia, administração do projeto, disponibilização de ferramentas, supervisão, validação de dados e experimentos, redação- revisão e edição.

Todos os autores contribuíram com a revisão do trabalho e aprovaram a versão submetida.

AGRADECIMENTOS

Os agradecimentos são direcionados para todos aqueles que de forma direta ou indireta contribuíram para a construção deste trabalho. Também agradeço ao PIBIFSP pela oportunidade de realizar esse projeto como bolsista.

REFERÊNCIAS

BAGNO, Marcos. **Nada na língua é por acaso**: por uma pedagogia da variação linguística. 1ª ed. São Paulo: Parábola Editorial, 2007

CALVET, Louis-Jean. **Sociolinguística**: uma introdução crítica. São Paulo: Parábola, 2002

TARALLO, F. **A pesquisa sociolinguística**. São Paulo: Ática, 1985.

TIMPONE, Rich; GUIDI, Michel. **Explorando a mudança de cenário da IA. Da IA Analítica à IA Generativa**. Ipsos, 2023.