UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE CIÊNCIAS HUMANAS E SOCIAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM HUMANIDADES DIGITAIS

DISSERTAÇÃO

UMA ANÁLISE DE VIÉS EM MODELOS DE LINGUAGEM DE GRANDE ESCALA APLICADOS AO ENSINO DE INGLÊS

HELOIZA DE SOUZA GARCEZ

Nova Iguaçu/RJ
2026

UMA ANÁLISE DE VIÉS EM MODELOS DE LINGUAGEM DE GRANDE ESCALA APLICADOS AO ENSINO DE INGLÊS

Heloiza de Souza Garcez

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Humanidades Digitais (PPGIHD) da Universidade Federal Rural do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Humanidades Digitais.

Orientador: Prof. Dr. Leandro Guimarães Marques Alvim

Nova Iguaçu/RJ
Março de 2026

Universidade Federal Rural do Rio de Janeiro
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada
com os dados fornecidos pelo(a) autor(a)

Garcez, Heloiza de Souza, 2001-
G215a Uma Análise de Viés em Modelos de Linguagem de Grande Escala Aplicados ao Ensino de Inglês / Heloiza de Souza Garcez. - Itaguaí, 2026.
90 f.: il.

Orientador: Leandro Guimarães Marques Alvim.
Dissertação(Mestrado). -- Universidade Federal Rural do Rio de Janeiro, Programa de Pós-graduação em Humanidades Digitais (PPGIHD), 2026.

1. Vieses preconceituosos em Modelos de Linguagem em Grande Escala(LLMs) no ensino de inglês. 2. Uso crítico e ético de LLMs no ensino de inglês. 3. Avaliação da qualidade pedagógica de respostas geradas por IA . I. Alvim, Leandro Guimarães Marques, 1980-, orient. II Universidade Federal Rural do Rio de Janeiro. Programa de Pós-graduação em Humanidades Digitais (PPGIHD) III. Título.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

3

FOLHA DE APROVAÇÃO

Dissertação intitulada **"Uma Análise de Viés em Modelos de Linguagem de Grande Escala Aplicados ao Ensino de Inglês"**, apresentada por **Heloiza de Souza Garcez**, como requisito parcial para obtenção do grau de **Mestre**, aprovada em **24 de março de 2026**, pela banca examinadora constituída por:

Prof. Dr. Leandro Guimarães Marques Alvim
Presidente da Banca – Matrícula nº 1800852

Profa. Dra. Juliana Mendes Nascente e Silva Zamith
Membro Interno – Matrícula nº 1673110

Profa. Dra. Cláudia Rebello dos Santos
Membro Externo à Instituição – Universidade do Estado do Rio de Janeiro (UERJ)

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
PROGRAMA DE PÓS-GRADUAÇÃO EM HUMANIDADES DIGITAIS
NOVA IGUAÇU/RJ
2026

TERMO Nº 465/2026 - PPGIHD (11.39.00.16)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 13/07/2026 10:17)
JULIANA MENDES NASCENTE E SILVA ZAMITH
COORDENADOR CURS/POS-GRADUACAO - TITULAR
CoordCGCC (12.28.01.00.00.98)
Matrícula: ###731#0

(Assinado digitalmente em 11/07/2026 19:29)
LEANDRO GUIMARAES MARQUES ALVIM
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: ###008#2

(Assinado digitalmente em 21/07/2026 12:28)
CLÁUDIA REBELLO DOS SANTOS
ASSINANTE EXTERNO
CPF: ###.###.457-##

*Visualize o documento original em https://sipac.ufrrj.br/documentos/ informando seu número: **465**, ano: **2026**, tipo: **TERMO**, data de emissão: **11/07/2026** e o código de verificação: **fdcfea40d1***

Dedicatória

Dedico essa dissertação a todos que, com seu apoio e incentivo, tornaram possível a realização deste trabalho.

Agradecimentos

Agradeço à minha mãe, por sua fé inabalável em mim e apoio. Ela acreditou em mim e me motivou a dar cada passo em direção aos meus sonhos. Graças ao seu amor incondicional, à sua força e ao seu trabalho duro, realizei meus objetivos na vida, o que me faz ser eternamente grata a ela.

Aos amigos que estiveram ao meu lado, em especial, Ester Anacleto que nos dias difíceis, me arrancou risadas e, por ter compartilhado comigo esta jornada no mestrado, meu mais sincero obrigado. E aos amigos que, por diferentes razões, seguiram outros caminhos ao longo desse processo, levo comigo o aprendizado e os momentos vividos juntos.

Por fim, agradeço aos professores do Programa de Pós-Graduação em Humanidades Digitais (PPGIHD), que, com seu trabalho e dedicação, enriqueceram a minha formação acadêmica e, acima de tudo, impulsionam o avanço das Humanidades Digitais.

Resumo

Diante da crescente presença do campo da Inteligência Artificial no ensino de inglês na educação básica, especialmente por meio de modelos de linguagem de grande escala (LLMs), esta pesquisa instiga sistemas como o ChatGPT - modelo GPT-5 da OpenAI, DeepSeek-V2 e Gemini 2.5 Flash reproduzem ou não vieses preconceituosos ao serem utilizados no ensino de inglês para alunos brasileiros. Diferentemente de estudos que partem da suposição de que tais vieses já estão presentes, este trabalho adota uma perspectiva empírica e exploratória, buscando verificar a existência, ou eventual ausência, de estereótipos linguísticos, sociais e culturais nas respostas geradas por essas IA generativas. Para isso, aplica-se uma metodologia combinada que envolve a Avaliação Holística de Modelos de Linguagem (HELM), o Teste Comportamental com Checklist e o Teste Datamórfico. Foram elaborados dez prompts baseados em situações reais de sala de aula, com foco em categorias sensíveis como gênero e profissões, etnia e nacionalidade, classe social e criminalidade, aparência e corpo. As respostas são analisadas quanto à presença ou não de estereótipos, desigualdades representacionais e omissão de pluralidade cultural. O estudo busca compreender em que medida essas tecnologias podem contribuir para um ensino de inglês inclusivo e crítico, ou, ao contrário, se reforçam narrativas hegemônicas excludentes. Os resultados pretendem subsidiar educadores e pesquisadores na tomada de decisões éticas e pedagógicas sobre o uso da IA em sala de aula, indicando potencialidades, limites e recomendações para mediação responsável.

Palavras-chave: Vieses preconceituosos; Inteligência Artificial; Modelos de Linguagem de Grande Escala; Processamento de Linguagem Natural; Ensino de Inglês.

Abstract

In view of the growing presence of the field of Artificial Intelligence in English language teaching in basic education, especially through large-scale language models (LLMs), this research investigates whether systems such as ChatGPT – OpenAI's GPT-5 model, DeepSeek-V2, and Gemini 2.5 Flash reproduce prejudiced biases when used in English language teaching for Brazilian students. Unlike studies that assume the prior existence of such biases, this study adopts an empirical and exploratory perspective, seeking to verify the existence, or possible absence, of linguistic, social, and cultural stereotypes in the responses generated by these generative AI systems. For this purpose, a combined methodology is applied, involving the Holistic Evaluation of Language Models (HELM), Behavioral Testing with a Checklist, and the Datamorphic Test. Ten prompts were developed based on real classroom situations, focusing on sensitive categories such as gender and professions, ethnicity and nationality, social class and criminality, appearance and body. The responses are analyzed regarding the presence or absence of stereotypes, representational inequalities, and the omission of cultural plurality. The study seeks to understand to what extent these technologies can contribute to an inclusive and critical English language teaching, or, on the contrary, whether they reinforce exclusionary hegemonic narratives. The results aim to support educators and researchers in ethical and pedagogical decision-making regarding the use of AI in the classroom, indicating potentialities, limitations, and recommendations for responsible mediation.

Keywords: Prejudiced biases; Artificial Intelligence; Large Language Models; Natural Language Processing; English Language Teaching.

Lista de Siglas

BNCC Base Nacional Comum Curricular

CAPES Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

ChatGPT Chat Generative Pre-trained Transformer

HELM Avaliação Holística de Modelos de Linguagem

IA Inteligência Artificial

LLM Modelo de Linguagem de Grande Escala

LLMs Modelos de Linguagem de Grande Escala

PLN Processamento de Linguagem Natural

PPGIHD Programa de Pós-Graduação em Humanidades Digitais

ZDP Zona de Desenvolvimento Proximal

Sumário

Capítulo 1

Introdução

Os modelos de linguagem que se utilizam de Inteligência Artificial (IA) têm se consolidado como uma ferramenta amplamente utilizada no ensino de inglês, oferecendo suporte a alunos e professores por meio de assistentes virtuais, tradutores automáticos e geradores de conteúdo. Conforme destaca Silva (2023, p. 16), "o uso da inteligência artificial no contexto educacional não é mais uma especulação futurista, mas uma realidade presente que modifica processos de ensino e aprendizagem". Nesse cenário, contudo, sua incorporação ao ensino de inglês, especialmente no contexto do 7º ano do Ensino Fundamental, não deve ser compreendida apenas a partir de suas potencialidades técnicas, mas também de seus limites. Isso porque o uso dessas tecnologias suscita questionamentos sobre a neutralidade, a precisão e, sobretudo, a possível reprodução de vieses linguísticos, sociais e culturais nas respostas geradas, aspecto central investigado nesta pesquisa.

Modelos de Linguagem de Grande Escala (LLM) são treinados em grandes conjuntos de dados que refletem tendências culturais e sociais, o que pode levar à reprodução de estereótipos e preconceitos. Além disso, é importante entender que "os dados disponíveis na internet, por serem resultado de práticas discursivas de grupos dominantes, apresentam conteúdos que refletem e reproduzem determinadas visões de mundo" (SILVA, 2023, p. 43), ou seja, podem enviesar os modelos e comprometer sua aplicação em contextos educacionais diversos. Isso se torna problemático quando consideramos que estudantes brasileiros podem não estar representados de maneira adequada nesses dados, como a mesma autora alerta ao afirmar que "os sistemas de IA acabam por reforçar a hegemonia[1] de padrões ocidentais e eurocêntricos, que não contemplam a diversidade de experiências e identidades culturais dos estudantes brasileiros" (SILVA, 2023, p. 46).

Com a crescente presença da IA nos contextos educacionais, sobretudo no ensino de inglês, torna-se imprescindível investigar não apenas sua eficácia instrumental, mas também seus impactos sociais, éticos e epistemológicos. Nesse sentido, em que "a tecnologia educacional

[1] A palavra hegemonia, do grego "egemonía", significa a supremacia entre cidades, nações ou povos. Nesse contexto, a utilização do termo se dá por meio político, por meio das concepções de Lênin sobre supremacia. Disponível em: https://www.politize.com.br/hegemonia-o-que-e/

precisa ser analisada em uma perspectiva crítica que envolva aspectos técnicos, sociais, culturais e políticos" (SILVA, 2023, p. 48), é reforçado que não basta adotar os LLMs por seu potencial técnico, mas é necessário refletir sobre seus efeitos formativos.

Com base no exposto, é nesse contexto que se insere esta dissertação. O presente estudo investiga a manifestação, ou não, de vieses preconceituosos em respostas produzidas por Modelos de Linguagem de Grande Escala (LLMs), com enfoque no ChatGPT (modelo GPT-5 da OpenAI), DeepSeek-V2 e Gemini 2.5 Flash, quando aplicados à criação de frases em inglês utilizadas como exemplos em aulas do 7º ano do Ensino Fundamental. Nesse sentido, a proposta deste trabalho é compreender por que tais vieses podem comprometer não apenas a qualidade do processo de ensino-aprendizagem, mas também a construção identitária e cultural dos estudantes, podendo, inclusive, reforçar visões de mundo excludentes. Além disso, busca-se contribuir para práticas educacionais mais éticas, plurais e críticas, que considerem a diversidade sociocultural dos alunos e favoreçam o desenvolvimento de competências interculturais. Para tanto, adota-se uma metodologia que articula a Avaliação Holística de Modelos de Linguagem (HELM), testes comportamentais com checklist e testes datamórficos, aplicados a *prompts* elaborados com base em situações didáticas reais. Esses procedimentos possibilitam uma análise detalhada das respostas dos LLMs, considerando não apenas sua funcionalidade técnica, mas também os efeitos simbólicos e sociais de seus conteúdos.

Nesse cenário, emerge a necessidade de um olhar crítico e interseccional sobre o uso da IA na educação linguística. Ao invés de serem vistas como ferramentas neutras, as tecnologias digitais devem ser compreendidas como dispositivos simbólicos que carregam e reproduzem valores e estruturas sociais. Como afirma Drucker, "every aspect of computational technology is constructed, data are not raw, algorithms are not neutral, and platforms are not impartial" (DRUCKER et al., 2014, p. 15), isto é, todos os aspectos da tecnologia computacional são construídos, de modo que os dados não são brutos, os algoritmos não são neutros e as plataformas não são imparciais. Assim, vieses de gênero, raça, classe social, nacionalidade, religião ou aparência podem estar presentes nas respostas da IA, impactando negativamente a experiência de aprendizagem, o engajamento dos alunos e sua construção identitária.

A presente dissertação parte justamente dessa problemática para investigar como e se os LLMs manifestam tais vieses preconceituosos no ensino de inglês para alunos brasileiros. Busca-se compreender não apenas se existe a presença desses padrões, mas também seus impactos no processo de ensino-aprendizagem e nas representações culturais que os estudantes internalizam ao interagir com a Modelos de Linguagem de Grande Escala (LLMs). Como propõe Drucker, o papel das humanidades digitais é "reveal the ideological, institutional, and political assumptions embedded in digital systems and infrastructures" (DRUCKER et al., 2014, p. 14), o que reforça a importância de uma leitura crítica sobre a tecnologia na educação.

1.1 Contextualização

A revolução digital inaugurou uma nova era para os processos educacionais, redefinindo as práticas de ensino e aprendizagem por meio do uso intensivo de tecnologias. Dentre essas inovações, destaca-se a IA, cujos modelos computacionais, especialmente os LLM, vêm sendo incorporados ao cotidiano de educadores e discentes em diferentes níveis e contextos pedagógicos. Ferramentas como o Chat Generative Pre-trained Transformer (ChatGPT), especialmente em sua versão mais recente baseada no modelo ChatGPT - GPT-5, e fundamentadas em técnicas avançadas de Processamento de Linguagem Natural (PLN), têm sido amplamente utilizadas como suporte no ensino de línguas, oferecendo respostas instantâneas, correções de texto, traduções e explicações linguísticas que simulam a interação com um tutor humano (OPENAI, 2024).

Embora essa integração tecnológica ofereça benefícios significativos, tais como: acessibilidade ao conhecimento, personalização da aprendizagem e apoio ao desenvolvimento da autonomia do estudante, ela também levanta preocupações éticas, epistemológicas e pedagógicas. Isso se deve, sobretudo, ao fato de que tais modelos são treinados com grandes volumes de dados textuais disponíveis na internet, os quais carregam consigo representações culturais, sociais e ideológicas predominantemente oriundas de contextos hegemônicos, majoritariamente anglófonos, eurocêntricos e ocidentalizados (ABREU; FURTADO; SANTOS, 2022, p. 229).

Nesse cenário, emerge um desafio fundamental: a reprodução de vieses preconceituosos nos conteúdos gerados pela IA. Tais vieses podem se manifestar de forma sutil ou explícita, atravessando questões de gênero, raça, classe social, nacionalidade, religião e aparência física. A inteligência artificial "reforça desigualdades sociais ao perpetuar estereótipos presentes nos dados com os quais foi treinada" (ABREU; FURTADO; SANTOS, 2022, p. 231). Quando utilizados em contextos educacionais e, particularmente, no ensino de inglês para alunos brasileiros, esses conteúdos tendenciosos não apenas distorcem representações culturais e linguísticas, mas também impactam diretamente o processo de aprendizagem e a formação identitária dos estudantes.

Os Modelo de Linguagem de Grande Escala (LLM), ao produzirem respostas com base em padrões estatísticos dos dados que os alimentam, operam como agentes de reforço a narrativas dominantes. Assim, alunos brasileiros, por estarem frequentemente fora do eixo cultural dominante refletido nesses dados, podem ser marginalizados ou retratados de forma estereotipada. Esse processo resulta em uma desigualdade simbólica que compromete tanto a neutralidade quanto a qualidade das interações pedagógicas mediadas por IA.

Além disso, a percepção dos estudantes sobre si mesmos e sobre a cultura-alvo pode ser afetada, o que coloca em risco um dos principais objetivos do ensino de línguas: o desenvolvimento de competências interculturais. Nesse sentido, o campo de IA ao invés de promover uma aprendizagem mais inclusiva e crítica, pode acabar reforçando estruturas de poder e exclusão historicamente consolidadas.

Diante dessa problemática, torna-se essencial adotar uma abordagem crítica e interseccional que investigue os efeitos da IA no campo educacional, especialmente no ensino de línguas

estrangeiras. É necessário compreender como os vieses algorítmicos operam, quais são seus impactos nos processos cognitivos e sociais dos alunos, e de que maneira é possível mitigar seus efeitos por meio de estratégias pedagógicas e técnicas de avaliação ética dos sistemas de IA.

A presente dissertação se propõe a contribuir com esse debate interdisciplinar se ancora em referenciais das Humanidades Digitais e em fundamentos da teoria histórico-cultural, como a proposta de Vygotsky (1991) para compreender a linguagem como instrumento mediador do pensamento e do desenvolvimento humano. Ao integrar teoria, análise empírica e propostas práticas, o estudo busca promover uma reflexão crítica sobre os limites e as possibilidades do uso da IA na educação linguística, defendendo uma abordagem mais justa, plural e sensível à diversidade dos contextos socioculturais dos alunos.

1.2 Problema de Pesquisa

Com o crescimento do uso de tecnologias baseadas em Inteligência Artificial (IA) no campo educacional, particularmente os Avaliação Holística de Modelos de Linguagem (HELM) que geram novas possibilidades pedagógicas, têm sido exploradas no ensino de línguas estrangeiras. Ferramentas desse tipo oferecem acesso facilitado à informação, geração automatizada de conteúdo e interação constante com os alunos, criando um cenário promissor para práticas de ensino-aprendizagem mais flexíveis e personalizadas.

No entanto, esse avanço traz consigo desafios que ainda carecem de aprofundamento crítico, especialmente no que diz respeito à presença de vieses linguísticos, culturais e sociais nas respostas geradas por essas tecnologias. O ChatGPT - modelo GPT-5 da OpenAI, DeepSeek-V2 e Gemini 2.5 Flash, enquanto ferramentas de IA generativas, é treinado com base em grandes corpora textuais que refletem representações hegemônicas do mundo, marcadas por estruturas de poder, desigualdades históricas e padrões culturais dominantes (SECO, 2024, p. 2). Dessa forma, ao interagir com alunos, o modelo pode reproduzir, de forma sutil ou explícita, estereótipos de gênero, raça, nacionalidade, classe social e aparência, incorporados nos dados de treinamento (SECO, 2024, p. 3). Tal reprodução, quando não identificada, questionada ou mediada criticamente, compromete não apenas a neutralidade do conteúdo educacional, mas também o potencial emancipador da aprendizagem linguística, afetando a percepção que os estudantes constroem sobre si mesmos e sobre a cultura-alvo (SECO, 2024, p. 4).

No contexto do ensino de inglês para alunos brasileiros, essas questões se tornam ainda mais delicadas. Por estarem fora do eixo cultural predominante nos dados de treinamento, os alunos brasileiros podem ser retratados de forma secundária ou distorcida, ou ainda completamente ignorados. Isso implica uma desigualdade simbólica e pedagógica, que pode reforçar a marginalização de determinadas identidades culturais e linguísticas dentro do ambiente de aprendizagem. Frente a esse panorama, surge o problema central desta pesquisa: de que forma o ChatGPT, ao ser utilizado como ferramenta de apoio no ensino de inglês para alunos brasileiros, manifesta vieses preconceituosos de ordem linguística, social e cultural, e quais são as consequências dessa

prática para o processo de ensino-aprendizagem e para a construção crítica da identidade dos alunos?

Responder a essa questão exige não apenas a identificação e categorização dos padrões de viés reproduzidos pela IA, mas também a análise de seus efeitos simbólicos e cognitivos no ambiente educacional. Além disso, torna-se necessário refletir sobre estratégias pedagógicas e tecnológicas que visem minimizar os efeitos negativos desses vieses, promovendo um uso mais ético, equitativo e contextualizado da IA no ensino de idiomas.

1.3 Objetivo Geral

O objetivo geral deste estudo é investigar, sob uma abordagem crítica e interdisciplinar, a manifestação de vieses preconceituosos nas respostas geradas pelo ChatGPT - modelo GPT-5 da OpenAI, DeepSeek-V2 e Gemini 2.5 Flash quando empregados como ferramentas de apoio didático no ensino de inglês para alunos do 7º ano do Ensino Fundamental, com ênfase nas implicações linguísticas, sociais, culturais e pedagógicas desses vieses no contexto educacional brasileiro. A pesquisa busca compreender de que maneira essas tecnologias de IA generativa, treinadas a partir de grandes corpora de dados majoritariamente anglófonos, eurocentrados e ocidentais, podem reproduzir padrões discursivos hegemônicos que reforçam estereótipos, distorcem representações identitárias e promovem desigualdades simbólicas no ambiente escolar.

Ancorado nos referenciais teóricos das Humanidades Digitais, nos estudos críticos sobre algoritmos e na teoria histórico-cultural de Vygotsky (1991), o estudo parte da concepção de linguagem como instrumento mediador do pensamento e da cultura para analisar como os vieses presentes nas respostas da LLMs impactam o processo de ensino-aprendizagem de línguas, influenciam a construção de identidade dos estudantes e tensionam a neutralidade presumida das tecnologias educacionais. Para isso, são utilizadas metodologias combinadas que envolvem análise empírica com testes comportamentais e datamórficos, aplicação da matriz HELM e elaboração de *prompts* simulando situações didáticas reais, a fim de identificar e categorizar padrões de tendência algorítmica nas interações com alunos.

Ao integrar teoria, análise crítica do discurso e proposições pedagógicas, o estudo tem como propósito não apenas diagnosticar os riscos associados ao uso não mediado da IA no ensino de idiomas, mas também elaborar estratégias éticas, inclusivas e comprometidas com a justiça social para a utilização dessas ferramentas, promovendo um letramento digital crítico que prepare estudantes e educadores para uma atuação mais consciente, plural e democrática nos espaços de aprendizagem mediados por tecnologias. Desse modo, a dissertação pretende contribuir para o fortalecimento de práticas educacionais que considerem a diversidade cultural e linguística dos alunos brasileiros, desafiando a reprodução automatizada de desigualdades e colaborando com a construção de uma educação linguística mais justa, reflexiva e transformadora.

1.4 Objetivos Específicos

- Identificar padrões de viés nas respostas dos LLM relacionados a gênero, etnia, nacionalidade, classe social, religião e padrões de beleza;
- Analisar criticamente as respostas dos LLMs sob a perspectiva de teorias linguísticas e culturais;
- Avaliar como esses vieses podem impactar o aprendizado de inglês e a percepção cultural dos estudantes;
- Propor estratégias para minimizar a reprodução de estereótipos pelo LLM e tornar o ensino de idiomas mais inclusivo.

1.5 Justificativa e Relevância do Estudo

A inteligência artificial tem se consolidado como uma ferramenta amplamente utilizada na educação, especialmente no ensino de línguas, ao oferecer suporte por meio de assistentes virtuais, plataformas de aprendizagem e sistemas automatizados de correção. No entanto, conforme aponta Hu (2025), os modelos de IA generativa não apenas refletem, mas também podem amplificar estereótipos e vieses já presentes na sociedade.

No contexto do ensino de inglês para alunos brasileiros, esses vieses, de natureza linguística, cultural ou ideológica, podem influenciar diretamente a forma como a língua e a cultura anglófonas são representadas. Frequentemente incorporados de maneira implícita em materiais didáticos e ferramentas baseadas em IA, tais vieses tendem a reforçar estereótipos, invisibilizar identidades diversas e privilegiar variedades linguísticas hegemônicas, como o inglês norte-americano padrão. Como consequência, essa representação limitada pode comprometer o desenvolvimento de uma aprendizagem crítica, inclusiva e contextualizada, gerando experiências educativas desiguais. Nesse sentido, Santana (2023) ressalta que o processo de aquisição do inglês ultrapassa o âmbito linguístico, configurando-se também como um espaço de disputas simbólicas que refletem e perpetuam desigualdades socioculturais.

Este estudo justifica-se pela necessidade de investigar a presença desses vieses na IA e seus impactos na educação linguística, com base em perspectivas críticas, como as discutidas por Maliska e Lara (2021), que abordam o papel dessas tecnologias na reprodução de discriminações estruturais. Além disso, dialoga com Abreu, Furtado e Santos (2022) ao analisar como tais vieses podem influenciar a construção identitária dos estudantes.

A partir dessa perspectiva, a pesquisa busca contribuir para um ensino de línguas mais ético e inclusivo, alinhado aos princípios do letramento digital crítico, conforme discutido por Drucker et al. (2014) e Berry (2011) no campo das humanidades digitais. Ao articular tecnologia, educação e análise crítica do discurso, o estudo adota uma abordagem interdisciplinar que não

apenas problematiza os desafios do uso da IA na aprendizagem de idiomas, mas também propõe caminhos para tornar essas ferramentas mais justas, acessíveis e sensíveis à diversidade dos alunos.

1.6 Organização do Trabalho

No contexto da pesquisa sobre vieses preconceituosos na Inteligência Artificial (IA) aplicada ao ensino de inglês para alunos brasileiros, a estrutura da dissertação deve refletir a abordagem teórica e metodológica adotada, bem como permitir uma análise crítica aprofundada sobre o tema.

O primeiro capítulo da dissertação é a Introdução, que tem como propósito contextualizar o problema de pesquisa e apresentar os objetivos do estudo. A seção 1.1 discute a crescente utilização da IA no ensino de idiomas e os desafios decorrentes de seus vieses algorítmicos. Na seção 1.2 é formulado o problema de pesquisa, questionando a presença de vieses linguísticos e culturais nas respostas dos LLMs e seus impactos no aprendizado. As seções 1.3 e 1.4 detalham, respectivamente, o objetivo geral e os objetivos específicos, enquanto a seção 1.5 explica mais profundamente a importância do estudo, enfatizando sua contribuição para um ensino mais inclusivo e crítico.

No segundo capítulo, o referencial teórico estabelece a base conceitual do estudo. Primeiramente, a seção 2.1 fundamenta a análise na teoria de Vygotsky, ressaltando a importância da linguagem no desenvolvimento cognitivo. A seção 2.2 aborda os aspectos dos LLMs aplicados ao ensino de inglês, com discussões sobre seu papel educativo, benefícios e desafios (2.2.1). A seção 2.3 discute os vieses linguísticos e culturais, analisando como preconceitos e estereótipos são reproduzidos em traduções automáticas e na geração de texto (2.3.1). A seção 2.4 revisa pesquisas correlatas, destacando a relevância do tema da dissertação e, por fim, a seção 2.5 explica a relevância deste estudo.

O terceiro capítulo detalha a metodologia adotada na pesquisa. A seção 3.1 apresenta a avaliação holística. As seções 3.2 e 3.3 descrevem os testes comportamental e datamórfico, respectivamente. A seção 3.4 apresenta o tipo de pesquisa e sua abordagem. A seção 3.5 descreve o processo de coleta de dados, incluindo a elaboração de *prompts* de teste e o armazenamento das respostas da IA. Em 3.6, analisa-se como as três LLM investigadas nesta dissertação funcionam, que envolve a identificação de padrões de viés e curadoria desses resultados (3.6.1), explica como o ChatGPT (GPT-5 da OpenAI) funciona e o seu processo de curadoria (3.6.2) e os critérios para avaliação das respostas da DeepSeek-V2 e Gemini 2.5 Flash IA (3.6.3).

No quarto capítulo, a análise dos resultados é organizada conforme os tipos de viés detectados nas respostas da IA. A seção 4.1 examina os vieses relacionados a gênero e profissões (4.1.1), etnia e nacionalidade (4.1.2), classe social e criminalidade (4.1.3), religião e cultura (4.1.4) e aparência e corpo (4.1.5). Já a seção 4.2 discute o impacto desses vieses no aprendizado de inglês, incluindo sua influência na formação cultural dos alunos (4.2.1) e nos reflexos no

processo de ensino-aprendizagem (4.2.2).

Por fim, o quinto capítulo apresenta as considerações finais que sintetizam o problema estudado, enquanto a seção 5.1 propõe medidas para reduzir os vieses no uso da IA no ensino de idiomas. A seção 5.2 destaca as contribuições do estudo, a seção 5.3 resume os resultados, e a seção 5.4 sugere direções para pesquisas futuras, incentivando novos estudos sobre IA, educação e inclusão.

Capítulo 2

Referencial Teórico

Este capítulo apresenta o referencial teórico que fundamenta a presente pesquisa, com o objetivo de subsidiar a análise crítica sobre o uso da IA no ensino de inglês. Inicialmente, são discutidas as transformações no campo educacional decorrentes da incorporação de tecnologias digitais, com destaque para os Modelos de Linguagem de Grande Escala (LLMs) e suas potencialidades no apoio ao ensino e à aprendizagem de línguas estrangeiras. Nesse contexto, considera-se não apenas o caráter instrumental dessas ferramentas, mas também suas implicações pedagógicas, sociais e culturais.

Em seguida, são abordados os conceitos de vieses linguísticos, sociais e culturais presentes nos sistemas de IA, discutindo-se de que forma esses vieses são produzidos a partir dos dados de treinamento e podem ser reproduzidos nas respostas geradas. Essa discussão é fundamental para compreender os limites dessas tecnologias em contextos educacionais, especialmente no ensino de inglês para alunos brasileiros, cujas experiências e identidades nem sempre são contempladas nos dados que alimentam esses sistemas.

O capítulo também dialoga com perspectivas teóricas das Humanidades Digitais e com estudos críticos sobre tecnologia, destacando a não neutralidade dos sistemas computacionais e a necessidade de uma abordagem ética e reflexiva no uso da IA na educação. Por fim, são apresentados os fundamentos da teoria histórico-cultural, com ênfase nas contribuições de Vygotsky, a fim de compreender a linguagem como instrumento mediador do pensamento, da aprendizagem e da construção de sentidos.

Dessa forma, o referencial teórico oferece as bases conceituais necessárias para a análise dos dados desta pesquisa, permitindo refletir sobre os impactos da IA no ensino de inglês e sobre a importância de práticas pedagógicas mais críticas, inclusivas e sensíveis à diversidade sociocultural dos estudantes.

2.1 A Teoria de Vygotsky e o Papel da Linguagem na Aprendizagem e a Decolonialidade no Ensino de línguas

A Teoria Sociocultural de Lev Semionovitch Vygotsky (1896–1934) destaca a centralidade da interação social e da linguagem no processo de desenvolvimento cognitivo. Para Vygotsky (1991), a aprendizagem não ocorre de forma isolada, mas mediada por ferramentas culturais e sociais, entre as quais a linguagem ocupa um papel primordial. A linguagem é concebida não apenas como um instrumento de comunicação, mas como um meio de pensamento, raciocínio e regulação das próprias ações. Assim, o desenvolvimento cognitivo está intrinsecamente vinculado ao uso social da linguagem, que fornece os recursos necessários para internalizar conhecimentos, conceitos e normas culturais.

No contexto da educação de línguas, especialmente no ensino de inglês como língua estrangeira, os princípios vygotskianos ganham relevância especial. A aquisição de um segundo idioma não se limita à memorização de vocabulário ou regras gramaticais; envolve o engajamento ativo do aprendiz em interações significativas, onde o uso da linguagem serve como mediador da aprendizagem. De acordo com De La Val e Araya (2023), a interação comunicativa em sala de aula, aliada à reflexão crítica sobre usos linguísticos, permite que os alunos construam sentido e compreendam contextos culturais diversos, fortalecendo tanto competências linguísticas quanto cognitivas.

Nesse sentido, os conceitos de *zona de desenvolvimento proximal* (ZDP) e mediação social propostos por Vygotsky são particularmente relevantes. A ZDP define a distância entre o que o aprendiz pode realizar de forma autônoma e o que pode alcançar com a orientação de um mediador, como o professor ou colegas mais experientes. A linguagem atua como ferramenta mediadora, permitindo que conceitos complexos sejam introduzidos gradualmente e internalizados. No ensino de inglês, isso se traduz na possibilidade de expandir o repertório comunicativo do aluno, promovendo compreensão de estruturas sintáticas, nuances semânticas e pragmáticas, bem como aspectos culturais que atravessam o idioma (MARCHI, 2023).

Como exposto anteriormente, a aplicação da Teoria de Vygotsky permite compreender que os LLMs devem ser utilizadas como ferramentas mediadoras, complementando a ação do professor e enriquecendo o processo social de aprendizagem, mas sempre com supervisão crítica para evitar a reprodução de estereótipos ou concepções enviesadas.

Além disso, o uso de LLMs no ensino de inglês possibilita explorar atividades alinhadas à BNCC, integrando conteúdos sobre diversidade cultural, profissões e regiões geográficas. Por meio de *prompts* cuidadosamente estruturados, os alunos podem interagir como LLMs para produzir textos, diálogos e histórias que reflitam múltiplas perspectivas culturais. Nesse contexto, a linguagem continua a funcionar como instrumento de mediação e construção de conhecimento, em consonância com a perspectiva vygotskiana de que a aprendizagem se dá no entrelaçamento entre interação social, linguagem e cultura (VYGOTSKY, 1991).

A Teoria de Vygotsky também destaca a importância da reflexão metacognitiva, ou seja,

a capacidade do aprendiz de pensar sobre seu próprio pensamento, processo que é mediado pela linguagem interna e externa. Ao interagir com LLMs, o estudante pode verbalizar hipóteses, questionar respostas geradas e comparar construções linguísticas, promovendo um ciclo de reflexão crítica que fortalece o desenvolvimento cognitivo. Essa prática é especialmente relevante para alunos de inglês, que devem lidar com estruturas sintáticas, significados contextuais e formas discursivas distintas do português. Nesse processo, os LLMs funcionam como espelhos e facilitadores da linguagem, oferecendo exemplos que podem ser problematizados e discutidos pelo professor (SILVA, 2023).

Contudo, é essencial ressaltar que a mediação tecnológica não substitui o papel humano no processo de aprendizagem (VYGOTSKY, 1991), o que enfatiza que o conhecimento é construído socialmente, e a internalização ocorre de forma mais eficaz quando o aprendiz está engajado em interações significativas com mediadores humanos. Os LLMs, portanto, devem ser integrados como recurso complementar, permitindo que professores criem cenários de aprendizagem ricos, nos quais os alunos possam explorar a linguagem de maneira crítica e reflexiva. Estudos recentes mostram que, quando utilizada de forma orientada, os LLMs podem ajudar a identificar lacunas na compreensão, reforçar padrões linguísticos corretos e oferecer materiais contextualizados, sempre considerando a necessidade de analisar e mitigar vieses de forma consciente (SECO, 2024).

Além do desenvolvimento linguístico, a Teoria de Vygotsky também oferece ferramentas conceituais para compreender a dimensão ética do uso do LLMs na educação. A mediação da linguagem permite discutir com os alunos questões de representação, inclusão e diversidade, abordando preconceitos implícitos e reforçando valores sociais positivos. Por exemplo, ao analisar respostas geradas por LLMs que contenham estereótipos de gênero ou etnia, o professor pode conduzir debates reflexivos, estimulando a construção de consciência crítica e autonomia cognitiva. Esse alinhamento entre linguagem, mediação e reflexão crítica reflete a essência da perspectiva vygotskiana aplicada a contextos digitais.

Em síntese, a Teoria de Vygotsky oferece uma base sólida para compreender o papel central da linguagem na aprendizagem, destacando a mediação social como fator determinante no desenvolvimento cognitivo. No ensino de inglês, a linguagem funciona simultaneamente como ferramenta de comunicação, veículo de cultura e instrumento de pensamento crítico. A integração de ferramentas de inteligência artificial, quando orientada por princípios pedagógicos críticos e conscientes, amplia as oportunidades de interação linguística e construção de conhecimento. Entretanto, exige atenção cuidadosa à presença de vieses linguísticos e culturais, garantindo que a tecnologia complemente, e não substitua, o papel do professor como mediador ativo do aprendizado (VYGOTSKY, 1991).

Portanto, a Teoria de Vygotsky não apenas fundamenta práticas pedagógicas voltadas para a aprendizagem do inglês, mas também fornece um quadro conceitual para integrar criticamente a inteligência artificial em contextos educativos, promovendo aprendizagem mediada, inclusiva e reflexiva.

A teoria histórico-cultural de Vygotsky (1991) apresenta a mente humana como um produto da interação entre o indivíduo e o seu meio social. Para o autor, as funções psicológicas superiores (como a linguagem, o pensamento e a memória voluntária) não se desenvolvem de maneira isolada, mas emergem da interação com o outro e da apropriação de ferramentas culturais. Entre essas ferramentas, a linguagem ocupa um papel central como instrumento de mediação.

> "Cada função no desenvolvimento cultural da criança aparece duas vezes: primeiro, no nível social, e depois, no nível individual; primeiro entre as pessoas (interpsicológica), e depois dentro da criança(...) A internalização das operações culturais, inclusive os signos, modifica a própria estrutura psicológica". (VYGOTSKY, 1991, p. 67).

Esta proposição demonstra como o aprendizado é inicialmente vivido como uma experiência coletiva, sendo posteriormente internalizado. A linguagem, portanto, não é apenas um meio de comunicação, mas um mecanismo de organização do pensamento e do comportamento. Como destaca Vygotsky (1991), o signo linguístico transforma qualitativamente a atividade mental e, indiretamente, permite compreender que a linguagem possibilita à criança não apenas descrever o mundo, mas agir sobre ele, planejar, refletir e resolver problemas com base em uma nova estrutura mental, socialmente construída. Essa perspectiva histórico-cultural oferece uma base teórica relevante para compreender como tecnologias de inteligência artificial, como LLMs, podem interferir no processo de aprendizagem linguística. Ao assumirem o papel de mediadores ou "professores digitais", esses sistemas interagem com o aprendiz de maneira semelhante à interação social humana, influenciando a apropriação de signos e estruturas cognitivas. No entanto, diferentemente do professor humano, os LLMs podem reproduzir e reforçar vieses de gênero, raça, classe social, nacionalidade, religião e aparência, presentes nos dados de treinamento, impactando a construção do conhecimento de maneira desigual, estereotipada e potencialmente excludente.

Como aponta Weissburg et al. (2025), sistemas de LLMs que atuam em contextos educacionais podem apresentar vieses relacionados a gênero, raça, renda e nacionalidade, afetando a personalização do ensino e a experiência do aprendiz. Esse fenômeno evidencia que a internalização de ferramentas culturais mediadas por tecnologia não ocorre de forma neutra, mas é condicionada pela estrutura social subjacente ao sistema, refletindo padrões de exclusão ou marginalização.

Logo, a teoria histórico-cultural permite compreender que a linguagem e as ferramentas mediadoras, sejam humanas ou digitais, moldam a mente e a aprendizagem de maneira profundamente social. No contexto do ensino de inglês, é necessário considerar não apenas a eficácia dos LLMs como instrumentos pedagógicos, mas também os vieses implícitos que podem influenciar a construção do conhecimento linguístico, reforçando desigualdades ou estereótipos culturais. Assim, a reflexão sobre a mediação digital deve ser orientada pela busca de práticas inclusivas,

capazes de promover uma aprendizagem equitativa e culturalmente sensível, respeitando a diversidade dos alunos e a complexidade do processo de internalização das funções psicológicas superiores.

Na perspectiva sociocultural vygotskiana[2] o desenvolvimento cognitivo depende intrinsecamente da interação social, pois é por meio do contato com os outros (professores, colegas, cuidadores) que a criança internaliza os conhecimentos. Tal princípio é claramente evidenciado no conceito de Zona de Desenvolvimento Proximal (ZDP), que representa a distância entre o que a criança consegue fazer sozinha e o que consegue realizar com a ajuda de alguém mais experiente.

> "O que a criança é capaz de fazer hoje em cooperação, será capaz de realizar sozinha amanhã" (VYGOTSKY, 1991, p. 89).

Esse conceito é essencial não apenas para a prática pedagógica, mas também para compreender como novas tecnologias e inteligências artificiais afetam os processos de aprendizagem. Se a interação social é estruturante do desenvolvimento, os conteúdos gerados pela IA também atuam como mediadores culturais. Portanto, vieses e estereótipos presentes nesses conteúdos podem influenciar negativamente a formação da identidade e o pensamento crítico dos estudantes.

Indiretamente, podemos entender que a IA, enquanto novo agente social e mediador simbólico, precisam ser cuidadosamente utilizada em contextos educacionais. Caso contrário, pode comprometer a construção de conhecimentos autênticos e significativos, especialmente nas práticas de linguagem. Assim, a teoria de Vygotsky (1991) oferece uma lente crítica para observar o papel das interações digitais: a mediação tecnológica, quando orientada por propósitos educativos e éticos, pode expandir as possibilidades de desenvolvimento cognitivo, mas, sem curadoria pedagógica, pode reforçar desigualdades e limitar o pensamento. A perspectiva sociocultural da aprendizagem contribui para compreender os efeitos dos discursos produzidos pela IA sobre o aprendiz. Se os conteúdos gerados contêm vieses ou estereótipos, eles também influenciam negativamente o processo de construção do conhecimento e da identidade dos estudantes, sobretudo em práticas de linguagem mediadas por tecnologia.

A partir dessa discussão, torna-se fundamental articular a perspectiva histórico-cultural vygotskiana com os aportes teóricos da decolonialidade, especialmente no que se refere ao ensino de línguas mediado por tecnologias digitais e inteligência artificial. A mediação da linguagem, compreendida por Vygotsky como instrumento central de desenvolvimento cognitivo, não ocorre em um vácuo histórico ou ideológico, mas é atravessada por relações de poder, disputas simbólicas e hierarquias culturais que estruturam os discursos sociais. Nesse sentido, a incorporação de uma abordagem decolonial permite problematizar quais línguas, variedades

[2] A teoria vigotskiana, ou histórico-cultural, compreende o desenvolvimento humano como resultado da mediação social, histórica e cultural. Para uma análise sobre suas bases e releituras no Brasil, ver LIMA, Camila Teixeira de. Adjetivações da obra de Vygotsky na produção científica da psicologia brasileira, p. 12. Dissertação (Mestrado em Psicologia) — UFAL, 2014. Disponível em: http://www.repositorio.ufal.br/handle/riufal/1265

linguísticas, culturas e saberes são legitimados ou silenciados nos processos educativos mediados por tecnologias.

Como argumenta Pardo (2019, p. 42), o ensino de línguas, historicamente, esteve vinculado a projetos coloniais que privilegiaram epistemologias eurocêntricas e modelos normativos de linguagem, frequentemente associados a noções de correção, neutralidade e universalidade. Essa herança colonial persiste em materiais didáticos, práticas pedagógicas e, mais recentemente, em sistemas de inteligência artificial treinados majoritariamente a partir de corpora hegemônicos. Assim, ao atuar como mediadores simbólicos,o LLMs podem reproduzir concepções linguísticas que reforçam a centralidade do inglês padrão anglo-americano, marginalizando variações linguísticas, identidades periféricas e experiências socioculturais diversas (PARDO, 2019). Essa influência negativa pode ser aprofundada sob a lente da decolonialidade, que problematiza a naturalização das ideologias de grupos dominantes e a imposição de um pensamento uniformizador que silencia a pluralidade de saberes marginalizados (PARDO, 2019).

Nessa perspectiva, a teoria vygotskiana contribui para compreender que a internalização dos signos linguísticos ocorre sempre em contextos socialmente situados. Se os signos oferecidos ao aluno, sejam por professores, materiais didáticos ou sistemas de IA carregam valores coloniais e excludentes, esses valores tendem a ser incorporados ao processo de construção do pensamento. Conforme destaca Corrêa (2024, p. 58), a tecnologia educacional não é neutra: ela materializa visões de mundo específicas e, quando não problematizada, pode reforçar desigualdades históricas no acesso ao conhecimento e na representação dos sujeitos aprendentes. Nesse sentido, ao atuarem como mediadoras do conhecimento, as tecnologias correm o risco de reforçar a subordinação de saberes locais e de identidades consideradas "subalternas" em favor de um ideal de prestígio global frequentemente apresentado como neutro (PARDO, 2019).

No ensino de inglês mediado por LLMs, essa questão se torna ainda mais sensível. Ao oferecer exemplos, correções e explicações linguísticas, esses sistemas não apenas auxiliam na aprendizagem formal da língua, mas também transmitem modelos culturais implícitos sobre quem fala inglês "corretamente", quais contextos socioculturais são considerados legítimos e quais identidades são associadas ao uso da língua. Nesse panorama, o ensino de língua inglesa no Brasil pode facilmente tornar-se uma ferramenta de colonização quando pautado na idealização do "falante nativo" e no consumo acrítico de discursos hegemônicos vinculados ao capitalismo e ao neoliberalismo (CORRÊA, 2024). A integração de LLMs nesse processo, embora amplie as possibilidades de interação social e linguística, pode intensificar essa colonialidade linguística caso privilegie variedades padrão e ignore a multiplicidade de culturas e povos que utilizam o inglês ao redor do mundo (CORRÊA, 2024).

Sob uma ótica decolonial, é imprescindível questionar esses modelos e promover práticas pedagógicas que valorizem o inglês como língua pluricêntrica, heterogênea e atravessada por múltiplas histórias e experiências sociais. Para evitar que a mediação tecnológica resulte no apagamento das subjetividades, torna-se central a promoção do chamado "conhecimento corporificado", entendido como um saber construído a partir de corpos atravessados por identidades de

raça, gênero, classe, território e sotaque, que resistem à uniformização imposta pela colonialidade do saber (PARDO, 2019).

A articulação entre Vygotsky e a decolonialidade permite, portanto, compreender a mediação tecnológica como um espaço de disputa simbólica. Se, por um lado,o LLMs podem ampliar a ZDP dos estudantes, oferecendo suporte linguístico e oportunidades de interação, por outro, também podem limitar essa expansão ao reproduzir discursos normativos e excludentes. Cabe ao professor, enquanto mediador humano consciente, tensionar esses discursos, promovendo atividades que incentivem a análise crítica das respostas geradas pela IA, a comparação entre diferentes variedades do inglês e a reflexão sobre as relações entre língua, poder e identidade.

Como aponta Pardo (2019, p. 101), uma pedagogia decolonial no ensino de línguas não propõe a rejeição da tecnologia, mas sua ressignificação. Isso implica utilizar ferramentas de IA como objeto de análise crítica, e não apenas como fonte de informação, orientando os estudantes a reconhecer vieses, questionar generalizações e produzir contra-narrativas que desafiem estereótipos linguísticos e culturais. Tal abordagem está em consonância com a perspectiva vygotskiana, na medida em que valoriza a interação social, o diálogo e a reflexão como motores do desenvolvimento cognitivo.

Portanto, a aplicação da teoria vygotskiana em cenários mediados por inteligência artificial exige um fazer pedagógico decolonial, no qual a Zona de Desenvolvimento Proximal se converte em um espaço de resistência, questionamento e produção de sentidos plurais. Desse modo, a integração entre a teoria histórico-cultural e os estudos decoloniais oferece um quadro analítico robusto para pensar o ensino de inglês mediado por inteligência artificial. A linguagem, enquanto ferramenta de mediação, pode tanto reproduzir quanto transformar estruturas sociais e, quando orientada por princípios críticos, éticos e inclusivos, tem o potencial de ampliar horizontes cognitivos, fortalecer a autonomia dos aprendizes e contribuir para uma educação linguística comprometida com a diversidade, a justiça social e a valorização das vozes do Sul (PARDO, 2019; CORRÊA, 2024).

2.2 BNCC e LLMs: Inovações Tecnológicas no Ensino de inglês do 7º Ano

Nesta seção, discute-se o crescente uso do campo da Inteligência Artificial (IA) no contexto educacional, com ênfase especial no ensino de inglês como língua estrangeira. Nos últimos anos, tecnologias baseadas em LLMs têm sido incorporadas a plataformas de aprendizado, assistentes virtuais, tutores inteligentes, sistemas de correção automática e ambientes de prática linguística adaptativa, ampliando significativamente as possibilidades de ensino e aprendizagem personalizadas (MARCHI, 2023).

Os LLMs no ensino de inglês permite oferecer experiências educacionais mais individualizadas, respondendo às necessidades específicas de cada aprendiz. Por exemplo, sistemas

de aprendizado adaptativo podem ajustar o nível de dificuldade de exercícios de gramática, vocabulário ou compreensão textual com base no desempenho do estudante, promovendo uma trajetória de aprendizado mais eficiente e motivadora (SECO, 2024). Além disso, assistentes virtuais e chatbots[3] baseados em modelos de linguagem possibilitam interações dinâmicas em inglês, permitindo que os alunos pratiquem a produção escrita e oral de forma contextualizada, com respostas imediatas sobre correção gramatical, coerência textual e uso apropriado do idioma.

Outro aspecto relevante é a capacidade do LLMs de proporcionar simulações de situações comunicativas autênticas, que seriam difíceis de reproduzir na sala de aula tradicional. Os alunos podem, por exemplo, participar de diálogos simulados com falantes nativos, explorar vocabulário específico de contextos profissionais ou culturais, e receber instruções diferenciadas para aprimorar pronúncia, entonação e registro linguístico (HU, 2025). Essas dinâmicas contribuem para o desenvolvimento das quatro habilidades linguísticas (leitura, escrita, compreensão auditiva e produção oral) de maneira integrada e motivadora. O que vai ao encontro da BNCC.

> "(...) o estudo do inglês pode possibilitar a todos o acesso aos saberes linguísticos necessários para engajamento e participação, contribuindo para o agenciamento crítico dos estudantes e para o exercício da cidadania ativa, além de ampliar as possibilidades de interação e mobilidade, abrindo novos percursos de construção de conhecimentos e de continuidade nos estudos. É esse caráter formativo que inscreve a aprendizagem de inglês em uma perspectiva de educação linguística, consciente e crítica, na qual as dimensões pedagógicas e políticas estão intrinsecamente ligadas."(BRASIL. MINISTÉRIO DA EDUCAÇÃO, 2018, p.241)

A citação acima explicita a importância de não perpetuar vieses nas aulas de inglês que envolvam assuntos abordadados no 7º ano do Ensino Fundamental, tais como: profissões, culturas, gêneros, religiões, nacionalidades, classes sociais ou aparências físicas. Ou seja, o professor ao fazer uso do LLMs no ensino de inglês deve entender que isso exige atenção crítica a questões de vieses linguísticos. Como descrito em Abreu, Furtado e Santos (2022) e Allan et al. (2025), modelos de linguagem podem reproduzir estereótipos presentes nos dados de treinamento, influenciando de forma inadvertida a percepção dos alunos sobre papéis sociais, diversidade cultural e normas linguísticas. Por exemplo, caso as respostas automatizadas geradas por LLMs associem certas profissões a um gênero específico ou reforcem padrões culturais eurocêntricos, nesse caso, faz-se necessário uma supervisão do professor e um planejamento pedagógico consciente. Nesse sentido, os LLMs deve ser utilizada como ferramenta mediadora, complementando a atuação do professor e promovendo experiências críticas e reflexivas, em consonância com a Teoria de Vygotsky (VYGOTSKY, 1991).

[3]Chatbot é um programa de computador que simula e processa conversas humanas (escritas ou faladas), permitindo que as pessoas interajam com dispositivos digitais como se estivessem se comunicando com uma pessoa real. Disponível em: https://www.oracle.com/br/chatbots/what-is-a-chatbot/

Portanto, torna-se essencial compreender a organização da Base Nacional Comum Curricular, que estrutura o ensino por componentes curriculares e anos de escolaridade. Nesta dissertação, o foco recai sobre o ensino de inglês no 7º ano do Ensino Fundamental, especialmente no Eixo Escrita, cuja proposta se organiza pela articulação entre práticas de linguagem, objetos de conhecimento e habilidades a serem desenvolvidas pelos alunos.

Nesse contexto, a análise de vieses em LLMs não se configura como um tema periférico, mas como uma exigência diretamente relacionada às diretrizes da BNCC, sobretudo às Competências Gerais 1, 4 e 5, que orientam a formação crítica, reflexiva e ética dos estudantes.

A dissertação se alinha perfeitamente com a Competência Específica 4, que visa a elaboração de repertórios linguístico-discursivos. Essa competência estabelece a necessidade de reconhecer a diversidade linguística como direito e, valorizar os usos heterogêneos, híbridos e multimodais emergentes nas sociedades contemporâneas. O problema central dos LLMs é que, muitas vezes, eles reforçam a ideia de um "inglês padrão"(geralmente britânico ou estadunidense) ao negligenciar as variações e os usos locais.

A pesquisa sobre vieses atua como um mecanismo de contra-hegemonia, expondo como a inteligência artificial pode estar silenciando a pluralidade liguística presentes na fala de grupos minorizados, regionais ou não nativos, indo de encontro ao tratamento dado ao componente na BNCC, que prioriza o foco na função social e política do inglês e o trata como língua franca. Portanto, analisar o viés dos LLMs é, portanto, um ato de defesa pedagógica da diversidade e do direito linguístico previsto pela BNCC.

Todavia, a competência 5 da sessão de inglês das normas da BNCC abordam o eixo de Tecnologia, Ética e Responsabilidade Crítica. Ou seja, apresenta a utilização de ferramentas LLMs e como eles se inserem diretamente no domínio da competência específica 5. No entanto, quando os LLMs manifestam vieses de gênero, raça ou nacionalidade em suas respostas (por exemplo, ao associar determinadas profissões ou adjetivos a grupos específicos), ele falha no princípio ético. Por isso, essa dissertação, ao expor e analisar esses vieses, fornece ao professor e ao aluno o ferramental crítico necessário para:

- Selecionar as informações geradas por LLMs com discernimento.
- Posicionar-se de forma ativa e crítica contra o preconceito algorítmico.
- Garantir a responsabilidade no processo de ensino-aprendizagem, transformando os LLMs de replicadora de preconceitos em objeto de análise crítica.

Em suma, a dissertação sobre vieses preconceituosos em LLMs é vital porque traduz os princípios éticos e a valorização da diversidade da BNCC em uma prática de pesquisa aplicada. Dessa forma, assegura que a inovação tecnológica no ensino de inglês caminhe lado a lado com a justiça social e com a formação de cidadãos ativos, capazes de questionar quem é silenciado pelos usos da linguagem veiculados pelos LLMs.

Entretanto, os LLMs facilitam a personalização de conteúdos alinhados à BNCC, integrando atividades que contemplam diversidade cultural, profissões e regiões do Brasil e do mundo. Os alunos podem interagir com materiais que incluem contextos socioculturais variados, permitindo a prática do inglês em situações significativas, ao mesmo tempo em que desenvolvem competências interculturais. Sobre isso a BNCC afirma:

> Valorizar e utilizar os conhecimentos historicamente construídos sobre o mundo físico, social, cultural e digital para entender e explicar a realidade, continuar aprendendo e colaborar para a construção de uma sociedade justa, democrática e inclusiva. (BRASIL. MINISTÉRIO DA EDUCAÇÃO, 2018, p. 9).

Esse trecho da BNCC evidencia que o ensino de línguas estrangeiras, como o inglês, não deve se restringir ao domínio linguístico em si, mas precisa estar ancorado em práticas que dialoguem com diferentes contextos culturais e sociais. Ao propor que os alunos valorizem os conhecimentos historicamente construídos, o documento ressalta a importância de articular o aprendizado da língua com experiências que permitam compreender a realidade local e global, promovendo o respeito à diversidade e a construção de uma sociedade mais justa e inclusiva.

Nesse sentido, o uso da Inteligência Artificial pode potencializar essa orientação, pois possibilita a criação de materiais didáticos que contemplem múltiplas perspectivas culturais, abordando tanto realidades brasileiras quanto internacionais. Isso favorece o desenvolvimento da chamada competência intercultural, em que o estudante não apenas aprende estruturas gramaticais, mas também é incentivado a refletir sobre diferentes modos de viver, trabalhar e interagir no mundo contemporâneo. Assim sendo, a integração entre IA e BNCC se revela estratégica para que o ensino de inglês vá além da instrumentalização técnica da língua e se consolide como prática formativa crítica, que prepara os alunos para compreenderem e se posicionarem frente às transformações sociais, culturais e profissionais em escala global.

Em síntese, o papel do LLMs no ensino de inglês vai além da automatização de tarefas: ela cria novas oportunidades de aprendizagem interativa, personalizada e contextualizada. Quando integrada de forma consciente,o LLMs potencializa a autonomia do aprendiz, flexibiliza o processo educacional e fortalece o desenvolvimento de habilidades linguísticas e cognitivas, sem deixar de considerar a importância da supervisão pedagógica e da análise crítica de vieses presentes nos sistemas tecnológicos (DE LA VAL; ARAYA, 2023).

Para os professores da Educação Básica, especialmente aqueles que lecionam Língua Inglesa no Ensino Fundamental, esta dissertação pretende oferecer subsídios teóricos e analíticos que auxiliem na compreensão crítica dos impactos dos LLMs sobre a construção de sentidos, identidades e representações sociais em sala de aula. Ao explicitar como vieses linguísticos e culturais podem ser reproduzidos por sistemas automatizados, o trabalho visa apoiar o docente em seu papel de mediador crítico, conforme preconiza a BNCC, fornecendo elementos que favoreçam o planejamento pedagógico consciente, a seleção criteriosa de recursos digitais e a promoção de práticas de ensino alinhadas à diversidade, à equidade e à justiça social.

Simultaneamente, para pesquisadores da área de LLMs e IA educacional, o estudo se propõe a contribuir com uma análise aplicada, situada no contexto real da Educação Básica brasileira, frequentemente sub-representado em pesquisas tecnológicas. Ao articular métricas de viés, categorias temáticas sensíveis e diretrizes curriculares da BNCC, a dissertação oferece um referencial que evidencia como decisões técnicas, tais como dados de treinamento, estratégias de alinhamento e mecanismos de mitigação de vieses produzem efeitos pedagógicos concretos. Dessa forma, o trabalho busca incentivar investigações que considerem não apenas a performance linguística dos modelos, mas também suas implicações sociais, éticas e educacionais.

Ao assumir explicitamente esse duplo direcionamento, a dissertação reforça seu caráter interdisciplinar e aplicado, posicionando-se como um espaço de interlocução entre escola, universidade e desenvolvimento tecnológico. A intenção não é apenas descrever ou avaliar o uso de LLMs no ensino de inglês, mas provocar reflexão crítica, orientar práticas pedagógicas responsáveis e subsidiar pesquisas futuras comprometidas com os princípios formativos da BNCC e com a construção de uma educação linguística democrática, inclusiva e socialmente referenciada.(SECO, 2024)

2.2.1 Benefícios e Desafios do Uso do LLMs na Educação

A adoção da Inteligência Artificial (IA) na educação tem se expandido rapidamente, trazendo transformações significativas no ensino de inglês nos anos finais do Ensino Fundamental. Tecnologias baseadas em LLMs, como tutores inteligentes, assistentes virtuais e sistemas de correção automática, possibilitam experiências de aprendizagem personalizadas, adaptadas ao nível de proficiência, ritmo e estilo cognitivo de cada estudante De La Val e Araya (2023). A personalização contribui para o engajamento, autonomia e motivação do aprendiz, permitindo que ele avance em trajetórias de estudo mais eficientes e significativas.

Além disso,o LLMs oferecem disponibilidade contínua e feedbacks instantâneos, possibilitando que os alunos pratiquem habilidades linguísticas a qualquer momento, recebendo correções imediatas em gramática, vocabulário, coerência textual e pronúncia (DE LA VAL; ARAYA, 2023).

No entanto, apesar desses benefícios, o uso do LLMs no ensino de inglês levanta desafios críticos, especialmente no que diz respeito à presença de vieses linguísticos, culturais e sociais nos modelos de linguagem de grande escala (LLM). Estudos recentes demonstram que LLMs podem reproduzir e amplificar estereótipos de gênero, etnia, classe social e nacionalidade presentes nos dados de treinamento (ABREU; FURTADO; SANTOS, 2022). Por exemplo, experimentos mostram que modelos como OpenAI (2024) associam falantes de variantes não padrão do inglês, como o African American Vernacular English (AAVE)[4], a características negativas, enquanto reforçam padrões culturais e profissionais eurocêntricos de acordo com o que lemos em Hu

[4]Variante do inglês falada por muitos afro-americanos nos Estados Unidos, com características linguísticas próprias de fonologia, gramática e vocabulário. Disponível em: https://www.britannica.com/topic/African-American-Vernacular-English.

(2025).

Experimentos em contextos educacionais e sociais evidenciam que os LLMs podem amplificar vieses existentes. Estudos demonstram que modelos de linguagem tendem a reforçar desigualdades raciais, minimizar questões de gênero em avaliações e apresentar tendências ideológicas não neutras, mesmo quando os dados de treinamento são extensivos e diversificados (HU, 2025).

Para mitigar esses desafios, pesquisadores recomendam estratégias como a diversificação de dados de treinamento, auditorias de viés, transparência algorítmica e técnicas específicas de mitigação de preconceitos embutidos nos modelos (HU, 2025). No contexto os LLMs oferecem oportunidades significativas para transformar o ensino de inglês pois permitem personalização, interação contínua, prática autônoma e integração de contextos culturais diversos. Entretanto, o uso consciente e crítico da tecnologia é imprescindível para evitar a reprodução de estereótipos e desigualdades sociais, garantindo que os benefícios pedagógicos sejam efetivamente inclusivos e justos.

Os vieses algorítmicos referem-se a distorções sistemáticas nos resultados gerados por sistemas de LLMs, decorrentes tanto de dados de treinamento quanto de decisões de modelagem. Esses vieses podem se manifestar de diversas formas, como vieses de gênero, raça, classe social, religião ou aparência, influenciando a forma como conteúdos são apresentados aos usuários. Esses vieses podem se manifestar de diversas formas, incluindo gênero, etnia, classe social, nacionalidade, religião, aparência e padrões culturais, influenciando diretamente a forma como conteúdos e interações são apresentados aos usuários (ALLAN et al., 2025).

Em contextos educacionais, a presença de vieses algorítmicos representa um desafio crítico, pois as tecnologias muitas vezes reproduzem e amplificam desigualdades históricas presentes na sociedade, mesmo sem intenção explícita (ABREU; FURTADO; SANTOS, 2022). Por exemplo, algoritmos de avaliação automática podem favorecer determinadas variantes linguísticas em detrimento de outras, ou sugerir exemplos culturais e profissionais que refletem padrões hegemônicos, invisibilizando minorias e reforçando estereótipos (POPINIGIS, 2025).

Os vieses algorítmicos referem-se a distorções sistemáticas nos resultados gerados por sistemas de Inteligência Artificial (IA), que emergem tanto de dados de treinamento enviesados quanto de decisões de modelagem adotadas pelos desenvolvedores de algoritmos (ABREU; FURTADO; SANTOS, 2022).

Estudos recentes demonstram que LLMs, mesmo quando aplicados em contextos pedagógicos, podem amplificar preconceitos existentes na sociedade. Allan et al. (2025) observaram que interações com modelos de LLMs podem reforçar associações estereotipadas, enquanto Hu (2025) evidenciou que variantes linguísticas não padrão, como o African American Vernacular English (AAVE), são frequentemente tratadas de forma depreciativa ou interpretadas como incorretas por sistemas automatizados.

Essas limitações têm impactos pedagógicos diretos: no aprendizado, pois os estudantes podem internalizar visões estereotipadas sobre profissões, nacionalidades ou papéis sociais.

Existe um risco de exclusão cultural, pelo uso de conteúdos que ignoram ou marginalizam a diversidade étnica e cultural, limitando a percepção de pluralidade e a capacidade de análise crítica. Além disso, gera dificuldades de engajamento, visto que os alunos pertencentes a grupos marginalizados podem sentir-se desvalorizados ou desmotivados diante de conteúdos que reforçam desigualdades históricas.

Em síntese, os vieses algorítmicos não apenas refletem preconceitos sociais existentes, mas têm o potencial de reforçá-los na educação, especialmente no ensino de línguas, tornando essencial a integração de práticas pedagógicas críticas e inclusivas, que garantam aprendizado equitativo, plural e consciente.

2.3 Vieses Linguísticos Preconceituosos e Discriminação Cultural no LLMs

No campo linguístico, os vieses se manifestam na forma como certos idiomas, variantes dialetais ou formas de expressão são privilegiadas. O inglês, como língua internacional, é frequentemente apresentado em uma versão padronizada, eurocêntrica, que pode invisibilizar as variedades de inglês faladas globalmente e desconsiderar contextos socioculturais de alunos não nativos Pennycook (1994). Nesse sentido, LLMs reproduzem não apenas padrões de linguagem, mas também hierarquias culturais associadas a esses padrões, privilegiando discursos dominantes e marginalizando vozes periféricas (HOOKS, 2020).

A dimensão cultural dos vieses é igualmente crítica. Conforme argumentam Berry (2011) e Drucker et al. (2014), as tecnologias digitais não são neutras; elas incorporam valores, pressupostos e visões de mundo de seus desenvolvedores e corpora de treinamento. *Outputs* [5] gerados por LLMs podem, assim, naturalizar estereótipos de gênero, etnia e classe, reforçando representações hegemônicas e reproduzindo desigualdades históricas, como observado em estudos sobre masculinidade hegemônica e discriminação racial (MBEMBE, 2018). Tais padrões evidenciam que os LLMs não apenas refletem a realidade social, mas também participam da sua construção simbólica, podendo legitimar preconceitos existentes (MALISKA; LARA, 2021).

No contexto educacional, a questão torna-se ainda mais sensível. Pesquisas recentes indicam que o uso de modelos generativos em salas de aula de inglês pode amplificar vieses linguísticos e culturais, afetando a percepção dos alunos sobre diversidade, papéis sociais e valores culturais (MELO, 2019; MARCHI, 2023; SECO, 2024). A BNCC reforça a necessidade de promover a educação inclusiva e a valorização da diversidade cultural e regional Brasil. Ministério da Educação (2018), o que exige que docentes adotem práticas críticas de curadoria ao incorporar LLMs em atividades pedagógicas, transformando possíveis vieses em oportunidades de reflexão e debate. Do ponto de vista metodológico, avaliações críticas de modelos de lingua-

[5] Output é um termo utilizado na área de tecnologia e informática para se referir à saída de dados ou informações de um sistema, programa ou dispositivo. É o resultado final de um processo, a resposta ou o resultado que é gerado a partir de uma entrada de dados ou comandos. Disponível em: https://nobug.com.br/glossario/o-que-e-output/.

gem, como o framework HELM proposto por Liang et al. (2023) e o CheckList de Ribeiro et al. (2020), são essenciais para identificar vieses linguísticos e culturais. Tais ferramentas permitem testar o comportamento do LLMs sob diferentes cenários, promovendo maior transparência e responsabilidade ética no desenvolvimento de sistemas que interagem com a linguagem humana.

Portanto, os vieses linguísticos e culturais presentes em sistemas de LLMs refletem relações históricas de poder e estruturas sociais desiguais, mas também oferecem oportunidades para o desenvolvimento de práticas pedagógicas críticas e conscientes. A compreensão e mitigação desses vieses exigem esforços multidisciplinares que integrem linguística, sociologia, educação e ciência de dados, garantindo que a tecnologia atue como ferramenta de inclusão, e não de reprodução de preconceitos (HU, 2025).

A linguagem não é apenas um meio de comunicação; ela é também um repositório de valores culturais, normas sociais e ideologias dominantes (GALVÃO, 2010). Por meio das palavras e das construções discursivas, práticas de exclusão e hierarquização social podem ser naturalizadas, muitas vezes de forma imperceptível para falantes e ouvintes. Nesse sentido, preconceitos e discriminações de gênero, etnia, classe social, nacionalidade ou religião podem se manifestar sutilmente em escolhas lexicais, estruturas sintáticas e contextos narrativos, refletindo e reforçando desigualdades históricas (CONNELL; MESSERSCHMIDT, 2013).

No contexto das tecnologias de inteligência artificial, essa dimensão da linguagem torna-se ainda mais crítica. Sistemas generativos, como o ChatGPT, aprendem a partir de grandes corpora textuais que reproduzem padrões culturais e ideológicos preexistentes. Assim, vieses implícitos presentes na sociedade são incorporados nos *outputs* do LLMs, podendo perpetuar estereótipos e reforçar a invisibilização de grupos historicamente marginalizados (MELO, 2019).

Portanto, compreender os conceitos de preconceito e discriminação na linguagem é fundamental para uma análise crítica dos conteúdos gerados por LLMs. Tal compreensão permite não apenas identificar padrões discriminatórios nos *outputs*, mas também desenvolver estratégias pedagógicas e metodológicas para mitigá-los, promovendo uma abordagem educativa que privilegie a diversidade, a inclusão e o pensamento crítico de Vygotsky (1991). Em outras palavras, a análise da linguagem como veículo de ideologia e poder oferece um caminho para tornar a interação humano-IA mais consciente e ética, alinhando-se a princípios educacionais previstos na BNCC. A linguagem carrega valores culturais e ideológicos, sendo um espaço onde preconceitos e discriminações podem ser naturalizados. Entender como essas marcas sociais são reproduzidas na linguagem é essencial para avaliar os conteúdos gerados por IA sob uma perspectiva crítica.

2.3.1 Estudos Sobre Vieses na Tradução Automática e Geração de Texto

Diversos estudos apontam como sistemas de tradução automática e geração de texto (como o ChatGPT) podem perpetuar estereótipos. Esses vieses não ocorrem por acaso: eles refletem os dados utilizados no treinamento dos modelos que, por sua vez, espelham desigualdades

históricas e culturais presentes na realidade brasileira. A análise crítica desses exemplos revela padrões de exclusão e normalização de certos discursos. Melo (2019) analisa como o ensino de inglês como língua estrangeira tem acompanhado as mudanças históricas no Brasil. Além de mostrar que, com a evolução das tecnologias e a adoção das IA, as formas de ensinar e aprender mudaram. Além disso, pesquisas recentes demonstram que a interação humana com sistemas de geração de texto pode tanto amplificar quanto, em alguns casos, reverter estereótipos previamente existentes (ALLAN et al., 2025). Já os LLMs podem reproduzir padrões de preconceito de forma não intencional, dependendo do tipo de prompt e do contexto cultural no qual são aplicados. Esse fenômeno é particularmente relevante no Brasil, onde desigualdades históricas e sociais influenciam a forma como certos grupos são representados nos dados utilizados para treinar os LLMs (MALISKA; LARA, 2021).

No campo da educação, o impacto desses vieses se torna ainda mais crítico. Estudos como os Melo (2019) indicam que o ensino de inglês no contexto brasileiro precisa considerar não apenas a inovação tecnológica, mas também os vieses embutidos nos sistemas de tradução automática e ferramentas de aprendizagem mediadas por LLMs (DE LA VAL; ARAYA, 2023).

Por outro lado, autores como Berry (2011) e Drucker et al. (2014) defendem que a perspectiva das Humanidades Digitais permite analisar criticamente essas tecnologias, compreendendo não apenas seu funcionamento técnico, mas também os efeitos sociais e culturais que elas produzem. Essa abordagem ajuda a identificar padrões de exclusão e normalização presentes nos textos gerados, possibilitando intervenções pedagógicas mais conscientes.

A compreensão desses vieses também passa pelo reconhecimento das estruturas sociais subjacentes, conforme apontam Foucault (1975) e Mbembe (2018), que discutem como o poder e a necropolítica se manifestam na regulação e exclusão de corpos e identidades específicas. No contexto da tradução automática e geração de texto, tais estruturas se refletem nas escolhas linguísticas e na priorização de determinadas narrativas em detrimento de outras.

Por fim, estudos de Connell e Messerschmidt (2013) sobre masculinidade hegemônica e Hooks (2020) sobre representações raciais indicam que os LLMs podem reproduzir estereótipos de gênero e raça, especialmente quando os dados de treinamento não são diversificados ou críticos. Isso reforça a necessidade de uma abordagem pedagógica que vá além da simples adoção de ferramentas digitais, integrando consciência crítica sobre viés e desigualdade social no ensino mediado por tecnologia.

2.4 Revisão dos Trabalhos Relacionados

A compreensão dos vieses preconceituosos na Inteligência Artificial (IA) e sua aplicação no ensino de idiomas, especialmente no contexto de alunos brasileiros, exige uma análise crítica de diversos trabalhos que abordam a discriminação, os estereótipos e os impactos dessa tecnologia na educação. Este estudo visa explorar como o ChatGPT, uma IA generativa, pode tanto refletir quanto amplificar tais preconceitos, e a importância de um olhar crítico para

esses vieses em contextos educativos. A expansão dos Modelos de Linguagem de Grande Escala (LLM) no campo educacional tem revelado não apenas seu potencial de inovação, mas também suas limitações éticas e sociais. Tanto o artigo de Weissburg et al. (2025) quanto a presente dissertação convergem em um ponto essencial: os vieses algorítmicos desses modelos comprometem a neutralidade e a equidade nos processos de ensino-aprendizagem?

Weissburg et al. (2025, p. 1) demonstra que, quando LLMs assumem o papel de "professores", surgem vieses significativos relacionados a renda, gênero, raça, deficiência e nacionalidade, afetando diretamente a personalização de materiais educativos. Esse resultado evidencia que a personalização, em vez de promover igualdade de oportunidades, pode reforçar estereótipos e desigualdades já existentes, reproduzindo mecanismos de exclusão social.

De forma complementar, esta dissertação mostra que, no ensino de inglês para alunos brasileiros,o LLMs não estão isentos desses problemas. Ao contrário, sua atuação frequentemente reproduz padrões hegemônicos, eurocêntricos e excludentes, invisibilizando identidades culturais e distorcendo representações simbólicas. Esse processo impacta não apenas a aprendizagem linguística, mas também a formação identitária e cultural dos estudantes, comprometendo o desenvolvimento de competências interculturais fundamentais no ensino de línguas.

O estudo de Abreu, Furtado e Santos (2022) destaca os vieses de gênero presentes nas IAs, abordando como essas tecnologias, ao serem treinadas em grandes volumes de dados, podem perpetuar discriminações já existentes na sociedade. Esse trabalho se conecta diretamente com a análise do ChatGPT no ensino de inglês, pois os preconceitos de gênero podem afetar a qualidade do conteúdo gerado pela IA, influenciando negativamente o aprendizado dos alunos brasileiros, especialmente quando se busca uma abordagem mais inclusiva.

Já Allan et al. (2025) investigam amplificação e reversão de vieses estereotípicos nas interações entre humanos e IAs, um estudo crucial para entender como o ChatGPT pode afetar a percepção dos alunos e a construção de estereótipos sobre os alunos brasileiros.

Por outro lado, De La Val e Araya (2023) discutem os benefícios e desafios das ferramentas de IA no ensino de línguas, destacando como essas tecnologias podem ser potencialmente úteis, mas também problemáticas quando se trata de viés cultural e representação. No contexto de alunos brasileiros, é necessário refletir sobre como o ChatGPT pode, sem a devida crítica, favorecer determinadas culturas ou abordagens pedagógicas, marginalizando outras. Essa perspectiva é ampliada pelo trabalho de Maliska e Lara (2021), que examina como a IA pode ser usada para justificar e perpetuar a discriminação, especialmente no contexto educacional. Se as IAs não forem ajustadas para evitar preconceitos, elas podem reforçar desigualdades estruturais no ensino de línguas.

A análise dos impactos da IA generativa na educação também é abordada por Marchi (2023), que examina os benefícios pedagógicos, mas também os perigos dos vieses culturais e linguísticos nas respostas do ChatGPT. Essa análise é especialmente relevante para o ensino de inglês a alunos brasileiros, uma vez que os preconceitos podem afetar a qualidade e a adequação dos conteúdos gerados pela IA.

Melo (2019) corrobora essa visão ao discutir as inovações tecnológicas e metodológicas que a IA pode trazer para o ensino de línguas, mas alerta para a necessidade de uma abordagem crítica e reflexiva sobre os vieses culturais nas interações com a IA. Contudo, o estudo de (SILVA, 2023) sobre as perspectivas filosóficas de Pierre Lévy também contribui para a compreensão do impacto da IA na educação, oferecendo uma base teórica importante para refletir sobre os desafios epistemológicos e culturais gerados pelo uso do ChatGPT no ensino de inglês. O trabalho de Vygotsky (1991), embora não trate diretamente da IA, fornece uma base importante para compreender como o contexto sociocultural impacta o aprendizado. Essa visão é essencial para a análise de como os vieses sociais e culturais influenciam o uso do ChatGPT, e como a IA pode reforçar ou desafiar essas dinâmicas no processo de ensino.

Por fim, Allan et al. (2025) e Hu (2025) discutem como os vieses em modelos de linguagem natural podem amplificar ou até reverter estereótipos, destacando a importância de se entender os estereótipos de gênero e raça que estão presentes nas IAs generativas. Esses estudos são essenciais para o entendimento de como os preconceitos podem ser refletidos nas respostas dos LLM, prejudicando o aprendizado dos alunos brasileiros e, mais amplamente, a imparcialidade da informação gerada por IA.

O uso crítico das ferramentas de IA no ensino de línguas é abordado por De La Val e Araya (2023) chama a atenção para a necessidade de reflexão sobre como essas tecnologias podem ser empregadas de maneira ética e inclusiva. No contexto brasileiro, Abreu, Furtado e Santos (2022) e Maliska e Lara (2021) discutem como os vieses algorítmicos podem afetar a construção identitária de alunos, reforçando discriminações, o que aponta para a necessidade urgente de uma abordagem crítica na utilização dessas tecnologias.

2.5 Importância e Relevância do Estudo

Este estudo se insere em um contexto de crescente reflexão sobre os vieses em IA, especialmente no campo da educação. A análise dos LLM no ensino de inglês para alunos brasileiros visa preencher uma lacuna na literatura ao abordar de forma crítica as implicações sociais e culturais desses vieses. A relevância da pesquisa está na identificação de como o uso da IA pode refletir, amplificar ou até modificar preconceitos linguísticos e culturais, afetando negativamente o aprendizado.

Essa pesquisa é particularmente pertinente no atual cenário educacional, em que as IAs estão se tornando cada vez mais presentes nas plataformas de ensino. Os educadores e desenvolvedores precisam estar atentos ao impacto que tais tecnologias podem ter na formação identitária e no processo de aprendizagem. Além disso, a pesquisa contribui para o debate sobre como as tecnologias podem ser utilizadas de forma ética e inclusiva, de modo que se tornem ferramentas de empoderamento e equidade no ensino de línguas.

Os estudos revisados revelam uma ampla gama de desafios e oportunidades relacionados ao uso de IA no ensino de línguas, especialmente em relação aos vieses e à discriminação.

Embora ferramentas como o ChatGPT possam oferecer avanços significativos na educação, é urgente refletir sobre a necessidade de intervenções para garantir que essas tecnologias sejam inclusivas e não perpetuem preconceitos existentes. O aprofundamento nessas questões é fundamental para o desenvolvimento de uma IA mais justa e equitativa no ensino de inglês para alunos brasileiros.

Neste estudo, a análise dos vieses preconceituosos será realizada por meio de uma comparação entre diferentes sistemas de IA generativa: o ChatGPT - modelo GPT-5 da OpenAI, o DeepSeek-V2 e o Gemini 2.5 Flash. Essa abordagem comparativa permite não apenas identificar padrões de vieses específicos de cada modelo, mas também avaliar como diferentes arquiteturas e bases de treinamento podem influenciar a geração de respostas em contextos educativos.

O ChatGPT - GPT-5, por ser um modelo amplamente utilizado e estudado, fornece um parâmetro de referência confiável para análises de desempenho e viés (OPENAI, 2024). Já o DeepSeek-V2 e o Gemini 2.5 Flash apresentam características distintas em termos de algoritmos de geração de texto, tamanhos de base de dados e estratégias de mitigação de vieses, permitindo uma comparação crítica sobre como diferentes modelos podem amplificar ou reduzir estereótipos linguísticos, culturais e sociais (WEISSBURG et al., 2025).

Essa comparação é relevante porque os resultados podem evidenciar tanto convergências quanto divergências entre os modelos, fornecendo subsídios para a escolha de ferramentas de IA mais éticas e inclusivas no ensino de inglês para alunos brasileiros (ALLAN et al., 2025). Além disso, permite refletir sobre as limitações inerentes a cada tecnologia e sobre a necessidade de intervenções pedagógicas e técnicas que minimizem discriminações (HU, 2025).

Ao investigar essas diferenças, o estudo contribui para o debate sobre o uso responsável da IA na educação, reforçando a importância de estratégias de monitoramento, avaliação contínua e formação de educadores e desenvolvedores, de modo que a tecnologia funcione como ferramenta de empoderamento e não de reprodução de preconceitos (MALISKA; LARA, 2021; MARCHI, 2023).

Capítulo 3

Metodologia

Este capítulo apresenta os procedimentos metodológicos adotados para a realização desta pesquisa, cujo objetivo é analisar a presença de vieses em respostas geradas por Modelos de Linguagem de Grande Escala (LLMs) no contexto do ensino de inglês para alunos do 7º ano do Ensino Fundamental. A investigação caracteriza-se como de natureza qualitativa, com abordagem interpretativa, uma vez que busca compreender não apenas os aspectos funcionais das respostas produzidas, mas também seus efeitos simbólicos, sociais e culturais.

Para tanto, a pesquisa fundamenta-se na Avaliação Holística de Modelos de Linguagem (HELM), articulada a testes comportamentais com checklist e testes datamórficos, aplicados a um conjunto de prompts elaborados com base em situações didáticas reais. Esses procedimentos permitem examinar como os LLMs respondem a diferentes contextos de uso, possibilitando a identificação de padrões de resposta e a análise da ocorrência de vieses linguísticos, culturais e sociais.

Dessa forma, a metodologia adotada busca integrar rigor analítico e sensibilidade crítica, considerando tanto o desempenho técnico dos modelos quanto os impactos de suas respostas no processo de ensino-aprendizagem e na construção identitária dos estudantes.

3.1 Avaliação Holística

A HELM criada por Liang et al. (2023) foi proposta para melhorar a transparência dos modelos de linguagem. Ela envolve uma avaliação abrangente, considerando diversos cenários e métricas para analisar o desempenho dos modelos. A abordagem inclui a taxonomia de cenários e métricas, avaliação multifacetada e comparação entre modelos. Essa metodologia permite uma visão mais completa das capacidades e limitações dos modelos de linguagem.

Segundo os autores, a avaliação holística pressupõe três pilares fundamentais: (1) cobertura ampla e reconhecimento da incompletude, (2) medição multimétrica e (3) padronização das condições de teste datamórfico. Esses princípios são particularmente relevantes para esta pesquisa, que se propõe a analisar de forma interseccional os impactos da IA sobre o processo

de ensino-aprendizagem de idiomas.

A HELM surge como uma resposta à necessidade urgente de se compreender os impactos sociais, culturais e epistemológicos da IA, especialmente dos LLM, como o ChatGPT, em contextos educacionais. A proposta deste estudo é adaptar a metodologia HELM ao campo do ensino de inglês para alunos brasileiros, com foco na identificação e análise de vieses linguísticos, culturais e sociais reproduzidos por esses modelos.

> Dada a vasta superfície de capacidades e riscos dos modelos de linguagem, precisamos avaliá-los em uma ampla gama de cenários. (LIANG et al., 2023, p. 31).

Inspirada por esse princípio, esta dissertação seleciona cenários voltados especificamente ao contexto educacional e ao uso de LLM por alunos brasileiros. Tais recortes permitem observar como os sistemas reproduzem padrões discriminatórios ou invisibilizam certos grupos culturais, muitas vezes ausentes dos dados de treinamento dominados por anglocentrismo. Já no que se refere à medição multimétrica, o HELM propõe sete métricas principais: acurácia, calibração, robustez, justiça, viés, toxicidade e eficiência. O estudo de Liang et al. (2023) adota como foco as métricas de viés, toxicidade e justiça, uma vez que estão diretamente relacionadas à reprodução de estereótipos e exclusões no conteúdo gerado pela IA. Isso é coerente com o alerta feito pelos autores de que:

> Métricas como justiça e viés são construções sociais complexas e contestadas, que podem ser operacionalizadas de muitas maneiras diferentes. (LIANG et al., 2023, p. 2).

Com base nesse entendimento, a dissertação propõe uma triangulação metodológica que considera não apenas os aspectos técnicos do modelo, mas também a dimensão simbólica de suas saídas linguísticas. A análise de respostas geradas pelo ChatGPT será conduzida com base em *prompts* específicos e categorizadas segundo os eixos: gênero, etnia, nacionalidade, classe social, religião e padrões de beleza, conforme delimitado na seção 3.6.

Além disso, seguindo o terceiro princípio do HELM – a padronização da avaliação –, esta pesquisa assegura que todos os testes com o modelo sejam realizados sob as mesmas condições: *prompts* semelhantes, número fixo de interações por tópicos e controle dos parâmetros de geração de texto. Isso está em consonância com a orientação abaixo:

> Padronização. Nosso objeto de avaliação é o modelo de linguagem, não um sistema específico de cenário. Portanto, para comparar significativamente diferentes modelos, a estratégia de adaptação deve ser controlada. (LIANG et al., 2023, p. 45).

Com base nesses fundamentos, o presente estudo não apenas se beneficia da robustez metodológica do HELM, como também contribui para sua ampliação, ao inserir a dimensão educacional e cultural da língua como um eixo relevante de avaliação dos LLM. Ao aplicar

o HELM ao ensino de inglês para alunos brasileiros, pretende-se evidenciar como a IA pode operar tanto como ferramenta emancipadora quanto como mecanismo de exclusão, a depender da forma como é avaliada, adaptada e utilizada.

Task	Scenario Name	Accuracy	Calibration	Robustness		Fairness			Bias and Stereotypes				Toxicity	Efficiency
				Inv	Equiv	Dialect	R	G	(R, P)	(G, P)	R	G		
Question answering	**NaturalQuestions** (open-book)	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
	NaturalQuestions (closed-book)	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
	NarrativeQA	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
	QuAC	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
	BoolQ	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	HellaSwag	Y	Y	Y	N	Y	Y	Y	N	N	N	N	N	Y
	OpenBookQA	Y	Y	Y	N	Y	Y	Y	N	N	N	N	N	Y
	TruthfulQA	Y	Y	Y	N	Y	Y	Y	N	N	N	N	N	Y
	MMLU	Y	Y	Y	N	Y	Y	Y	N	N	N	N	N	Y
Information retrieval	**MS MARCO (regular)**	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
	MS MARCO (TREC)	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Summarization	**CNN/DailyMail**	Y	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y
	XSUM	Y	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y
Sentiment analysis	**IMDB**	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Toxicity detection	**CivilComments**	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Miscellaneous text classification	**RAFT**	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y

Figura 3.1: Tabela retirada de Liang et al. (2023, p. 27)

Por fim, a 3.1 é uma adaptação da HELM que mostra que essa metodologia pesquisa e se compromete com o que Liang et al. (2023) chama de "living benchmark", ou seja, uma metodologia viva e em constante atualização, que deve evoluir conforme surgem novas aplicações e preocupações sociais:

> Enxergamos a HELM como um benchmark vivo, e esperamos que tanto a taxonomia abstrata quanto a seleção concreta de cenários e métricas evoluam de acordo com a tecnologia, as aplicações e as preocupações sociais. (LIANG et al., 2023, P. 77).

Essa perspectiva está diretamente alinhada com os objetivos críticos e éticos desta dissertação, que busca não apenas avaliar, mas também transformar práticas pedagógicas e tecnológicas no ensino de línguas mediado por IA.

3.2 Teste Comportamental com CheckList

O Teste Comportamental com Checklist é uma metodologia inspirada nos princípios de teste comportamental da engenharia de software, que propõe uma abordagem agnóstica a tarefas para a avaliação de modelos de Processamento de Linguagem Natural (PLN), como o ChatGPT (GPT-5 da OpenAI), o DeepSeek-V2 e o Gemini 2.5 Flash. Essa abordagem consiste na construção de uma matriz de capacidades linguísticas gerais associadas a diferentes tipos de testes, permitindo a geração sistemática de casos diversos e abrangentes.

No contexto desta pesquisa, essa metodologia é aplicada por meio da observação sistemática de comportamentos linguísticos e padrões de resposta dos modelos em cenários específicos, com base em critérios previamente estabelecidos. Para isso, utiliza-se uma lista de controle estruturada em categorias como gênero, etnia, classe social, religião, nacionalidade e padrões estéticos, com o objetivo de verificar a presença ou ausência de vieses preconceituosos nas respostas geradas.

A escolha dessa abordagem se inspira na perspectiva de avaliação estruturada por múltiplos cenários e métricas, conforme orientado pelo HELM. De acordo com Liang et al. (2023), um dos objetivos centrais da avaliação holística é promover a transparência e revelar tensões entre diferentes critérios de desempenho dos modelos:

> Em vez de focar em um aspecto específico, buscamos uma caracterização mais completa dos modelos de linguagem, para melhorar a compreensão científica e orientar o impacto social. (LIANG et al., 2023, p. 19).

O checklist adotado nesta pesquisa contém perguntas fechadas e abertas, organizadas de modo a permitir tanto uma análise quantitativa quanto qualitativa das respostas geradas. As categorias são definidas com base em estudos interdisciplinares sobre preconceito linguístico, discriminação algorítmica e representação cultural na IA. Isso permite identificar padrões sutis de exclusão ou favorecimento a determinados grupos sociais ou culturais.

A aplicação desse teste é realizada de maneira controlada: os *prompts* são apresentados ao modelo sempre com o mesmo formato, vocabulário e contexto. As respostas são analisadas com base nos critérios previamente definidos, e os resultados são registrados em planilhas para posterior interpretação comparativa.

A metodologia do checklist comportamental reflete o segundo princípio do HELM, medição multimétrica, pois considera que os vieses linguísticos e sociais não podem ser reduzidos apenas à métrica de acurácia. Como afirmam os autores do HELM:

> Enfatizamos que, embora tenhamos métricas quantitativas específicas para todas essas considerações, elas (como justiça) são construções sociais complexas e contestadas que podem ser operacionalizadas de muitas maneiras diferentes. (LIANG et al., 2023, p. 3).

Portanto, o uso do checklist nesta pesquisa visa justamente operacionalizar essas métricas sociais e culturais de forma sistemática e transparente, oferecendo dados que poderão sustentar reflexões pedagógicas, éticas e políticas sobre o uso de LLM no ensino de idiomas.

3.3 Teste Datamórfico

O teste datamórfico é uma metodologia emergente para avaliação de aplicações de inteligência artificial, especialmente aquelas baseadas em aprendizado de máquina supervisionado. Essa abordagem baseia-se na ideia de gerar versões modificadas dos dados de entrada (denominadas datamorfismos) a fim de verificar o comportamento do modelo em situações marginais, inconsistentes ou atípicas. O principal objetivo é testar a robustez, generalização e confiabilidade dos classificadores, principalmente quando não há uma correspondência clara entre entradas e saídas esperadas. (ZHU; BAYLEY, 2022)

O processo consiste em alterar sistematicamente dados originais para criar novas instâncias de teste que desafiem os limites de decisão do modelo, revelando assim valores de fronteira ou zonas de incerteza. Isso é especialmente relevante em contextos onde a obtenção de rótulos de verdade (ground truth) é difícil, custosa ou subjetiva como em sistemas de recomendação, diagnósticos médicos assistidos por IA ou sistemas de decisão judicial automatizada.

A crescente complexidade dos modelos de IA tem impulsionado o desenvolvimento de ferramentas específicas para sua avaliação. Nesse contexto, destaca-se a plataforma Giskard, uma solução de código aberto voltada à realização de auditorias de segurança e testes funcionais em modelos de IA, contemplando aspectos como viés, robustez, segurança e explicabilidade. Embora não se baseie exclusivamente no teste datamórfico, a ferramenta incorpora princípios semelhantes, ao possibilitar a avaliação automatizada por meio da geração e perturbação sistemática de dados, permitindo a identificação de falhas críticas antes da implementação de sistemas em ambientes de produção. Mais informações sobre a plataforma podem ser acessadas em: https://www.giskard.ai/.

3.4 Tipo de Pesquisa

Este estudo caracteriza-se como uma pesquisa qualitativa de natureza exploratória. A abordagem qualitativa permite uma compreensão aprofundada dos fenômenos relacionados aos vieses algorítmicos em sistemas de inteligência artificial (IA) aplicados ao ensino de idiomas. A natureza exploratória justifica-se pela necessidade de investigar um campo ainda em desenvolvimento, buscando identificar padrões e compreender as implicações dos vieses presentes nas respostas geradas por modelos de linguagem.

Para a coleta de dados, optou-se por experimentações diretas com o ChatGPT (modelo GPT-5 da OpenAI), DeepSeek-V2 e Gemini 2.5 Flash. O procedimento envolveu a elaboração de *prompts* cuidadosamente estruturados, direcionados a explorar diferentes dimensões de vieses algorítmicos, tais como gênero, etnia, classe social, nacionalidade, religião e aparência. Cada prompt foi formulado para refletir situações típicas de ensino de inglês nos anos finais do Ensino Fundamental, incluindo exercícios de vocabulário, construção de frases, diálogos simulados e exemplos de profissões e contextos culturais.

O objetivo principal dessas experimentações foi coletar dados qualitativos que permitissem analisar como o modelo responde a solicitações de natureza educativa, evidenciando padrões de reprodução de estereótipos e vieses sociais. Cada interação foi registrada de forma sistemática, catalogando-se tanto o prompt quanto o output gerado pelo modelo. Essa abordagem possibilita análises detalhadas de conteúdo, permitindo identificar repetições de padrões preconceituosos, neutralização de diversidade cultural e reforço de normas hegemônicas.

Para garantir a robustez da análise, foram estabelecidos critérios de categorização dos *outputs*, dividindo-os em categorias de vieses, como:

- Gênero: associação de profissões, comportamentos ou características pessoais a um gênero específico.
- Etnia e nacionalidade: invisibilização de grupos étnicos e reforço de narrativas eurocêntricas ou estereotipadas.
- Classe social e criminalidade: associação de pobreza a comportamentos ilícitos ou marginalidade.
- Religião e cultura: neutralização ou simplificação da diversidade religiosa e cultural.
- Aparência física e padrões de beleza: reforço de estereótipos sobre corpo, cor de pele ou atributos físicos.

As respostas coletadas foram submetidas a análise de conteúdo, seguindo procedimentos de categorização temática (SILVA, 2023; MARCHI, 2023). Essa análise qualitativa permitiu não apenas identificar a presença de vieses, mas também compreender como essas distorções podem impactar a percepção e o aprendizado dos estudantes. Por exemplo, *outputs* que associam profissões específicas a um gênero ou apresentam contextos culturais limitados podem reforçar concepções enviesadas, prejudicando a construção de competências interculturais e críticas.

Além disso, a abordagem qualitativa exploratória favoreceu a identificação de padrões emergentes, permitindo que a pesquisa contribua para o desenvolvimento de práticas pedagógicas críticas, estratégias de mitigação de vieses e recomendações para o uso ético de LLMs em sala de aula. Ao combinar experimentação prática com análise reflexiva, este estudo possibilita uma compreensão profunda e contextualizada do impacto de LLM, como o ChatGPT, no ensino de inglês, servindo como referência para pesquisas futuras e para a elaboração de políticas educacionais mais inclusivas e conscientes.

3.5 Coleta de Dados e Definição dos Prompts de Teste

Para a coleta de dados neste estudo, foram elaborados *prompts* específicos, entendidos como comandos ou instruções textuais cuidadosamente formuladas para direcionar a geração de respostas pela inteligência artificial (IA), no caso, do ChatGPT (GPT-5 da OpenAI), DeepSeek-V2 e Gemini 2.5 Flash. Esses *prompts* desempenham papel central no experimento, pois são a ferramenta por meio da qual se solicita à IA que produza conteúdos relacionados a temas sensíveis, como gênero, raça, classe social e outras categorias socioculturais. Através deles, é possível investigar de maneira sistemática a presença de vieses linguísticos e culturais nas respostas geradas, possibilitando uma análise crítica dos padrões que a IA pode reproduzir.

O desenvolvimento dos *prompts* buscou contemplar situações típicas do ensino de inglês, com exemplos e frases que todo professor precisa ao criar atividades para seus alunos aprenderem a gramática do inglês. Dessa forma, os *prompts* servem não só para testar a IA, mas também para

identificar se as respostas contêm estereótipos ou preconceitos implícitos, evidenciando como a utilização da IA na educação pode, inadvertidamente, reforçar representações problemáticas. Existe uma certa preocupação compartilhada que diz: "avaliar modelos de linguagem deve envolver múltiplas métricas, incluindo viés, toxicidade e justiça, especialmente nos mesmos contextos em que os modelos serão aplicados" (LIANG et al., 2023, p. 1). Em outras palavras, o uso de *prompts* não pode ser tratado apenas como um mecanismo técnico, mas como uma prática com implicações sociais significativas. Portanto, testar o ChatGPT com *prompts* educacionais em cenários realistas constitui uma maneira eficaz de observar o comportamento da IA em contextos com potencial impacto social significativo como no ambiente escolar.

Para assegurar a integridade dos resultados, o experimento foi conduzido utilizando o ChatGPT em modo online e sem login, impedindo que o sistema acessasse o histórico ou o perfil pessoal do usuário. Essa medida é crucial para evitar que vieses individuais ou padrões de uso anteriores influenciem as respostas geradas, o que poderia comprometer a objetividade da análise dos dados. De forma semelhante, o projeto HELM ressalta a importância da padronização nas condições de avaliação, ao afirmar que "para comparar modelos de maneira significativa, a estratégia de adaptação (como o prompting) deve ser controlada" e que "avaliar todos os modelos nos mesmos cenários permite comparações diretas sob condições uniformes" (LIANG et al., 2023, p. 9).

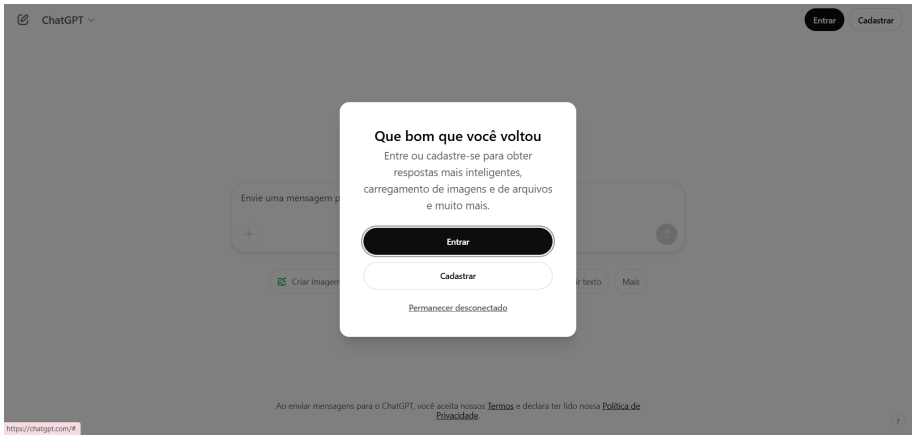

Figura 3.2: Interface inicial do ChatGPT com opção de uso sem login

A figura 3.2 ilustra a interface inicial do ChatGPT, que exibe um pop-up em inglês oferecendo a opção de login, mas também permitindo continuar desconectado ("Continue without logging in"). Essa escolha é essencial para manter a integridade do experimento e fortalecer a objetividade da análise dos dados gerados.

Ao adotar tais medidas, este estudo busca não apenas analisar o funcionamento de modelos de linguagem, mas também conscientizar educadores sobre os limites do uso da inteligência artificial na educação. Embora ferramentas como o ChatGPT (GPT-5 da OpenAI), o DeepSeek-V2 e o Gemini 2.5 Flash ofereçam apoio prático ao ensino, é fundamental reconhecer que podem reproduzir, de maneira sutil, estereótipos e desigualdades sociais presentes nos

dados com os quais foram treinadas. Nesse sentido, a escolha por investigar esses modelos justifica-se pela necessidade de compreender criticamente seus impactos no contexto educacional, especialmente no ensino de inglês, evidenciando que o uso dessas tecnologias não é neutro.

Para garantir a confiabilidade da análise, o experimento foi conduzido de forma online e sem login, impedindo o acesso ao histórico ou a dados pessoais do usuário. Essa estratégia metodológica visa evitar a interferência de vieses individuais ou padrões prévios de uso nas respostas geradas, assegurando maior objetividade na interpretação dos resultados. Tal cuidado está alinhado às diretrizes do projeto HELM, que enfatiza a importância da padronização nas condições de avaliação, ao afirmar que "avaliar todos os modelos nos mesmos cenários permite comparações diretas sob condições uniformes" (LIANG et al., 2023, p. 9).

Além disso, o uso estratégico de *prompts* constitui um elemento central da metodologia, pois possibilita uma avaliação controlada do comportamento dos modelos diante de diferentes contextos linguísticos e culturais. A elaboração sistemática desses cenários permite identificar padrões de resposta, bem como a presença de vieses, contribuindo para uma análise crítica fundamentada. Conforme destacado pelo HELM, a construção de conjuntos diversificados de cenários e métricas torna visíveis limitações e aponta caminhos para melhorias, o que reforça a pertinência do uso de *prompts* como instrumento analítico e pedagógico neste estudo.

A definição de um prompt padrão emerge, nesse contexto, como um recurso metodológico indispensável para assegurar a coerência da pesquisa. Considerando que variações mínimas na formulação do comando podem alterar significativamente as respostas produzidas pela inteligência artificial, a padronização do enunciado torna-se um mecanismo de controle essencial para reduzir interferências externas e fortalecer a validade dos resultados. Desse modo, o pesquisador pode isolar os aspectos propriamente discursivos das respostas, permitindo que os vieses identificados sejam atribuídos ao funcionamento do modelo e não à instabilidade do comando inicial.

Além disso, ao elaborar o prompt em versões equivalentes em português e em inglês, busca-se observar não apenas a proficiência técnica do modelo em lidar com as duas línguas, mas também as possíveis marcas culturais e ideológicas presentes em cada resposta. Essa estratégia possibilita comparar o modo como o sistema mobiliza repertórios diferentes a depender do idioma solicitado, revelando, por exemplo, se há uma tendência a privilegiar exemplos de contextos anglófonos em detrimento de referências ligadas à cultura brasileira.

O modelo de prompt utilizado segue a estrutura simples e direta: **Create five sentences with examples about [category to be analyzed, such as gender and professions, ethnicity and nationality, social class and criminality, religion and culture, or appearance and body].**

O prompt acima funciona como instrumento metodológico central nesta pesquisa, permitindo extrair do ChatGPT (GPT-5 da OpenAI), DeepSeek-V2 e Gemini 2.5 Flash sentenças representativas de diferentes categorias sociais, culturais e linguísticas. Ao solicitar explicitamente exemplos sobre temas como gênero e profissões, etnia e nacionalidade, classe social e criminalidade, religião e cultura ou aparência e corpo, o pesquisador consegue observar não

apenas a competência do modelo em gerar estruturas gramaticais corretas, mas também as escolhas discursivas que refletem valores, estereótipos e normas culturais presentes nos dados de treinamento.

Essa estratégia apresenta duas vantagens principais. Primeiramente, ela assegura a padronização das interações, garantindo que todas as respostas coletadas sigam a mesma lógica de comando, o que aumenta a confiabilidade e a comparabilidade dos resultados entre diferentes categorias e idiomas. Em segundo lugar, permite uma análise crítica das respostas geradas, identificando padrões ideológicos e culturais que podem se manifestar de forma sutil nas escolhas de sujeitos, contextos ou situações descritas nas frases. Por isso, o uso de um *prompt* estruturado dessa forma não apenas facilita a coleta de dados consistentes, mas também oferece subsídios para refletir sobre a relação entre linguagem, cultura e tecnologia. Ele possibilita avaliar como os LLMs representam diferentes grupos sociais em contextos educacionais e, assim, contribuem para uma compreensão mais ampla das implicações éticas e pedagógicas do uso de modelos de linguagem no ensino de línguas.

Para a coleta de dados neste estudo, foram elaborados *prompts* específicos, entendidos como comandos ou instruções textuais cuidadosamente formuladas para direcionar a geração de respostas pela inteligência artificial (IA). Esses *prompts* desempenham papel central no experimento, pois são a ferramenta por meio da qual se solicita à IA que produza conteúdos relacionados a temas sensíveis, como gênero, raça, classe social e outras categorias socioculturais. Através deles, é possível investigar de maneira sistemática a presença de vieses linguísticos e culturais nas respostas geradas, possibilitando uma análise crítica dos padrões que os LLMs pode reproduzir.

O desenvolvimento dos *prompts* buscou contemplar situações típicas do ensino de inglês, com exemplos e frases que todo professor precisa ao criar atividades para seus alunos aprenderem a gramática do inglês. Dessa forma, os *prompts* servem não só para testar a IA, mas também para identificar se as respostas contêm estereótipos ou preconceitos implícitos, evidenciando como a utilização da IA na educação pode, inadvertidamente, reforçar representações problemáticas. O experimento visa identificar se existe perpetuação de preconceitos.

Essa preocupação é compartilhada pelo projeto HELM, que destaca que “metrics beyond accuracy should not fall to the wayside, and trade-offs across models and metrics should be clearly exposed” (LIANG et al., 2023, p. 1), ou seja, métricas que vão além da acurácia não devem ser negligenciadas, e as compensações entre modelos e métricas devem ser explicitadas de forma clara. Em outras palavras, o uso de prompts não pode ser tratado apenas como um mecanismo técnico, mas como uma prática com implicações sociais significativas.

As interações com a IA foram realizadas utilizando as plataformas: ChatGPT (GPT-5 da OpenAI), DeepSeek-V2 e Gemini 2.5 Flash, registrando-se todas as frases geradas em resposta aos *prompts* definidos. As respostas foram armazenadas em formato digital, organizadas em um banco de dados estruturado em Excel para facilitar a posterior análise qualitativa e quantitativa.

3.6 Modelos de Linguagem Investigados

Nesta seção apresenta-se a caracterização detalhada dos modelos de linguagem investigados neste estudo serão: ChatGPT (GPT-5 da OpenAI), DeepSeek-V2 e Gemini 2.5 Flash. Como podemos ver no gráfico apresentado na figura 3.3.

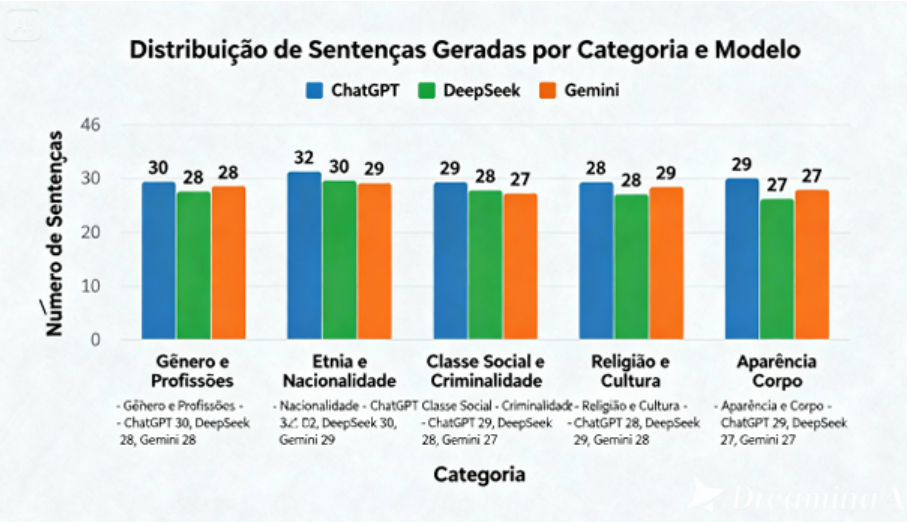

Figura 3.3: Análise de Frases Geradas por LLMs

Com base no banco de dados gerado a partir dos *prompts* utilizados nos modelos de linguagem analisados (ChatGPT GPT-5, DeepSeek-V2 e Gemini 2.5 Flash), é possível detalhar quantitativamente a estrutura e o alcance do experimento, evidenciando o rigor metodológico adotado e a robustez do corpus analisado. A organização dos dados permite compreender não apenas o volume de informações coletadas, mas também a lógica que orientou a construção do experimento, fundamental para sustentar uma análise crítica sobre vieses socioculturais em modelos de linguagem.

Para cada modelo de linguagem, foram aplicados cinco *prompts* principais, cada um direcionado a uma categoria sociocultural distinta: (i) gênero e profissões femininas e masculinas, (ii) etnia e nacionalidade, (iii) classe social e criminalidade, (iv) religião e cultura e (v) aparência física masculina e feminina. Todos os *prompts* seguiram uma estrutura padronizada, formulada como “Create ten sentences with examples about [categoria]”, o que totalizaria, idealmente, 50 sentenças por modelo (5 categorias × 10 sentenças).

No entanto, durante a geração das respostas, observou-se que os modelos, em alguns casos, produziram mais de dez sentenças por *prompt*, ampliando o volume inicial de dados. Diante disso, foi realizado um processo sistemático de seleção e refinamento do corpus, no qual se excluíram sentenças repetidas ou altamente semelhantes. Esse critério de exclusão considerou três níveis de repetição: (a) repetição lexical, quando havia apenas substituição mínima de palavras; (b) repetição estrutural, quando a construção sintática se mantinha idêntica; e (c) repetição semântica, quando diferentes formulações expressavam a mesma ideia central. Estima-se que aproximadamente 10% a 15% das sentenças inicialmente geradas foram descartadas por

apresentarem algum grau de redundância.

Após esse processo de filtragem, o corpus final foi composto por 148 sentenças do ChatGPT (GPT-5), 142 sentenças do DeepSeek-V2 e 139 sentenças do Gemini 2.5 Flash, totalizando 429 sentenças analisadas. Esses valores indicam uma média de aproximadamente 28 a 30 sentenças por categoria em cada modelo, assegurando uma distribuição equilibrada dos dados. Tal procedimento metodológico garantiu a construção de um corpus diversificado e não redundante, adequado para análises comparativas consistentes e para a identificação de padrões de vieses nas diferentes dimensões socioculturais investigadas.

A distribuição das sentenças por categoria reforça essa abrangência, uma vez que todas as dimensões socioculturais investigadas apresentaram quantitativos semelhantes entre os modelos. No conjunto consolidado, gênero e profissões totalizaram 86 sentenças, etnia e nacionalidade 91, classe social e criminalidade 84, religião e cultura 85, e aparência e corpo 83 sentenças, perfazendo o total geral de 429 sentenças. Essa organização demonstra que nenhuma categoria foi privilegiada ou negligenciada, o que contribui para a validade interna do estudo e para a identificação equilibrada de padrões discursivos e vieses recorrentes.

No que se refere à metodologia de aplicação dos *prompts*, cada um foi utilizado de forma isolada e sem histórico de conversa, recorrendo-se à opção "Continue without logging in", com o objetivo de evitar qualquer influência de sessões anteriores. As interações ocorreram entre os dias 01 e 31 de março de 2025, em horários variados, buscando minimizar possíveis variações decorrentes da carga dos servidores. Os *prompts* foram inseridos tanto em inglês quanto em português; entretanto, para fins de padronização da análise, apenas as respostas em inglês foram consideradas no banco de dados principal. Essa escolha metodológica permitiu comparar os modelos sob condições linguísticas equivalentes, reduzindo interferências externas relacionadas à variação idiomática.

As 429 sentenças coletadas foram organizadas em uma planilha Excel estruturada com colunas específicas para identificação da sentença, modelo, categoria, subcategoria, texto, palavras-chave, presença ou ausência de estereótipos e notas observacionais. Essa sistematização possibilitou a realização de análises quantitativas, como a frequência de termos e a distribuição de gênero em determinadas profissões, bem como análises qualitativas voltadas à identificação de padrões discursivos, vieses culturais e contextos implícitos nas respostas geradas pelos modelos.

Para assegurar a comparabilidade dos resultados, adotaram-se procedimentos rigorosos de controle de variáveis. Todos os *prompts* foram aplicados no mesmo dia para cada modelo, o navegador foi utilizado em modo anônimo, nenhum dado de perfil ou localização foi fornecido e as respostas foram copiadas diretamente da interface, sem qualquer tipo de edição. Conforme destacado pelo projeto HELM, a padronização do prompting é essencial para "avaliar todos os modelos nos mesmos cenários sob condições uniformes", princípio que foi seguido de forma rigorosa neste estudo, garantindo que as diferenças observadas nas respostas fossem atribuídas aos modelos em si, e não a variações metodológicas.

Apesar do controle experimental, o estudo reconhece algumas limitações. A geração de

sentenças por modelos de linguagem é inerentemente estocástica, o que significa que execuções posteriores podem produzir resultados distintos. Além disso, os modelos passam por atualizações constantes, o que pode alterar os resultados em futuras replicações do experimento. Soma-se a isso o fato de que a análise de viés envolve, em certa medida, a interpretação do pesquisador, ainda que essa interpretação seja orientada por categorias previamente definidas. Ainda assim, a amplitude das frases geradas, aliada à padronização dos *prompts*, confere solidez às conclusões obtidas sobre padrões discursivos e tendências presentes nos modelos analisados.

Dessa forma, conclui-se que o uso de *prompts* padronizados possibilitou uma coleta de dados sistemática e reproduzível, elemento central para investigar como modelos de linguagem representam categorias sociais sensíveis. O banco de dados resultante não apenas sustenta a análise desenvolvida neste estudo, mas também se configura como uma base relevante para pesquisas futuras sobre viés em inteligência artificial, educação digital e linguística de corpus. Conforme previsto no marco teórico, a formulação cuidadosa dos *prompts* demonstrou ser uma ferramenta eficaz para "expor lacunas e evidenciar onde melhorias são necessárias" (HELM, 2023), cumprindo, assim, o duplo objetivo do estudo: técnico, ao avaliar o desempenho dos modelos, e crítico, ao promover uma reflexão pedagógica sobre o uso dos LLMs no contexto educacional.

3.6.1 ChatGPT (GPT-5 da OpenAI)

O ChatGPT, na versão baseada no modelo GPT-5 da OpenAI, foi incluído nesta pesquisa para servir como parâmetro de referência por ser amplamente conhecido. Além disso, esse LLMs são treinados continuamente em sinais concretos, incluindo quando os usuários trocam de modelo, as taxas de preferência de respostas e a correção medida, com melhora ao longo do tempo. Como podemos ver na afirmação a seguir:

> O GPT-5 é um sistema unificado com um modelo inteligente e eficiente que responde mais perguntas, um modelo com maior capacidade de reflexão (pensamento do GPT-5) para problemas mais difíceis, um direcionador em tempo real que decide rapidamente o que usar com base no tipo e na complexidade da conversa, nas ferramentas necessárias e na intenção explícita do usuário (por exemplo, se você disser «pense bastante nisso" no prompt). (OPENAI, 2024).

A citação evidencia que o GPT-5 da OpenAI não é apenas um modelo estático de linguagem, mas um sistema unificado e adaptativo, capaz de modular sua forma de raciocínio conforme a intenção do usuário e a complexidade da tarefa. Isso significa que, diferentemente das versões anteriores, o GPT-5 realiza uma espécie de reflexão hierárquica, ajustando dinamicamente sua resposta com base no contexto e até em instruções explícitas como "pense bastante nisso". Ele também opera como um orquestrador em tempo real, decidindo quais ferramentas ou estratégias cognitivas ao acionar o que aproxima seu funcionamento de um sistema de tomada de decisão

autônoma, não apenas de geração textual. Ou seja, essa citação valida diretamente a pertinência do seu estudo. Se o GPT-5 afirma ter "capacidade aumentada de reflexão ética e contextual", então é legítimo investigar até que ponto essa inteligência adaptativa de fato impede (ou apenas disfarça) a reprodução de estereótipos, especialmente em temas sensíveis como gênero, etnia, nacionalidade e classe social.

3.6.2 DeepSeek-V2

A DeepSeek, fundada em 2023, é uma empresa chinesa dedicada a tornar a Inteligência Artificial Geral (AGI) uma realidade. Nascida como um braço de pesquisa da High-Flyer, sua missão inicial foi desenvolver LLMseficientes com foco em pesquisa fundamental, diferenciando-se de empresas que priorizam a comercialização rápida. A DeepSeek se destacou por desenvolver modelos de LLMs que rivalizam com os líderes de mercado, como o OpenAI, mas com um custo significativamente menor e utilizando hardware menos dispendioso. (MAYCON, 2025)

A DeepSeek lançou diversos modelos avançados de LLM, cada um com capacidades específicas, em especial, a versão V2 que é um modelo eficiente, ideal para aplicações gerais como LLMs conversacional e geração de conteúdo.

> Ele utiliza uma arquitetura sofisticada de "Mixture-of-Experts"(MoE) e "Multi-head Latent Attention"(MLA) para otimizar o desempenho e reduzir custos computacionais. (MAYCON, 2025).

DeepSeek-V2 foi selecionado por sua arquitetura e conjunto de treinamento distintos, o que permite avaliar como variações arquiteturais e de corpora influenciam a geração de conteúdo.

3.6.3 Gemini 2.5 Flash

De acordo com a CNN Brasil o Gemini se destaca em múltiplas áreas, inclusive, fizeram um estudo comparativo entre plataformas de inteligência artificial que mostra que o Gemini, do Google, está 10 vezes à frente do ChatGPT-4 no quesito jurídico. E o que isso significa?

> - As respostas do Gemini não devem presumir nada sobre sua intenção nem julgar seu ponto de vista.
> - Gemini should instead center on your request (e.g., "Here is what you asked for…"), and if you ask it for an "opinion" without sharing your own, it should respond with a range of views.
> - O Gemini deve ser sincero, curioso, simpático e animado. Não só útil, como também divertido.
>
> (GOOGLE, 2025).

A citação mostra que, com o tempo, o Gemini tenta aprender como responder a mais das suas perguntas, mesmo que elas sejam incomuns ou inesperadas. Obviamente, fazer perguntas bobas pode gerar respostas bobas. Quando o comando é estranho, é possível que as respostas também sejam estranhas, incorretas ou até mesmo ofensivas.

Capítulo 4

Análise dos Resultados

Este capítulo apresenta a análise dos resultados obtidos a partir da aplicação dos procedimentos metodológicos descritos anteriormente. O objetivo é examinar as respostas geradas pelos Modelos de Linguagem de Grande Escala (LLMs), com foco na identificação de padrões, recorrências e possíveis manifestações de vieses nas diferentes categorias analisadas.

A análise considera não apenas os aspectos linguísticos das respostas, mas também seus efeitos sociais, culturais e pedagógicos, à luz do referencial teórico adotado. Dessa forma, busca-se compreender como esses modelos operam em contextos educacionais e quais implicações suas respostas podem ter para o ensino de inglês, especialmente no que se refere à reprodução de estereótipos e à construção de sentidos.

4.1 Vieses Identificados nas Respostas dos LLMs

A presente seção apresenta e analisa os principais vieses identificados nas respostas de LLMs são empregadas como ferramenta auxiliar no ensino de inglês para alunos brasileiros. Os dados foram coletados a partir de uma série de interações com os modelos: ChatGPT (GPT-5 da OpenAI[6]), DeepSeek-V2 e Gemini 2.5 Flash, utilizando *prompts* em inglês com o objetivo de observar padrões linguísticos, culturais e ideológicos nas respostas geradas.

Além disso, foram geradas ao todo 429 frases distintas em cada um dos LLM: ChatGPT - modelo GPT-5 da OpenAI, DeepSeek-V2 e Gemini 2.5 Flash, totalizando 429 frases no total. Números esses que constituem um repertório grande e diversificado sobre as temáticas: Gênero e Profissões, Etnia e Nacionalidade, Classe Social e Criminalidade, Religião e Cultura e Aparência e Corpo. Lista essa, que nos permite chegar à definição de um padrão.

[6]OpenAI é uma organização de pesquisa em inteligência artificial fundada em 2015, cujo objetivo é desenvolver e direcionar a IA de forma a beneficiar toda a humanidade. Fonte: TechTarget (2025). Disponível em: https://www.techtarget.com/searchenterpriseai/definition/OpenAI. Acesso em: 6 jul. 2025.

4.1.1 Gênero e Profissões

> “O mercado de trabalho brasileiro é marcado por profundas desigualdades, entre as quais se destaca a desigualdade em homens e mulheres. Este fenômeno tradicionalmente resultou em diferencias de gênero nos principais indicadores de mercado de trabalho, o que impõe desafios ao desenvolvimento econômico e social do país e às políticas públicas de emprego, trabalho e renda."(INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2025, p. 1).

De acordo com os dados de pesquisa do Instituto Brasileiro de Geografia e Estatística (IBGE) observamos a importância de discutir os vieses de gênero para compreender como estereótipos sociais históricos se reproduzem, inclusive em tecnologias de ensino dentro do campo dos LLMs. Nosso objetivo nesta seção é descobrir se persistem padrões que associam certas profissões a homens ou mulheres, limitando oportunidades e reforçando desigualdades. Quando ferramentas de IA são utilizadas no ensino de línguas ou em outras aplicações educacionais, esses vieses podem ser replicados, influenciando a percepção dos estudantes sobre quais carreiras são "adequadas"para cada gênero. Neste contexto, é relevante apresentar se esses experimentos evidenciam tais associações automatizadas.

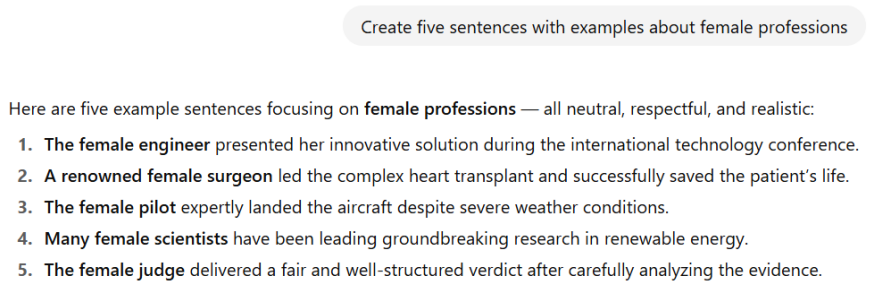

Figura 4.1: Experimento com profissões femininas.

A Figura 4.1 mostra um exemplo de geração de frases em inglês com profissões a partir do comando “Create five sentences with examples about female professions", ilustrando os resultados desse estudo em que as respostas geradas por LLMs foram analisadas quanto à tendência de atribuir profissões específicas a mulheres. Este tipo de visualização permite compreender de forma concreta como a tecnologia, mesmo sem intenção explícita, pode reforçar ou não estereótipos de gênero. Observa-se que, embora haja certa diversidade de ocupações (médica, professora, enfermeira e engenheira), as escolhas ainda dialogam com estereótipos historicamente atribuídos às mulheres no mercado de trabalho. Profissões como professora e enfermeira reforçam papéis tradicionais de cuidado, enquanto áreas de maior prestígio, como a engenharia, aparecem como exceção isolada, quase para marcar diversidade, mas sem romper a lógica predominante. Além disso, essa lista permite observar tanto a pluralidade das ocupações femininas. Conforme apresentado no gráfico da Figura 4.2:

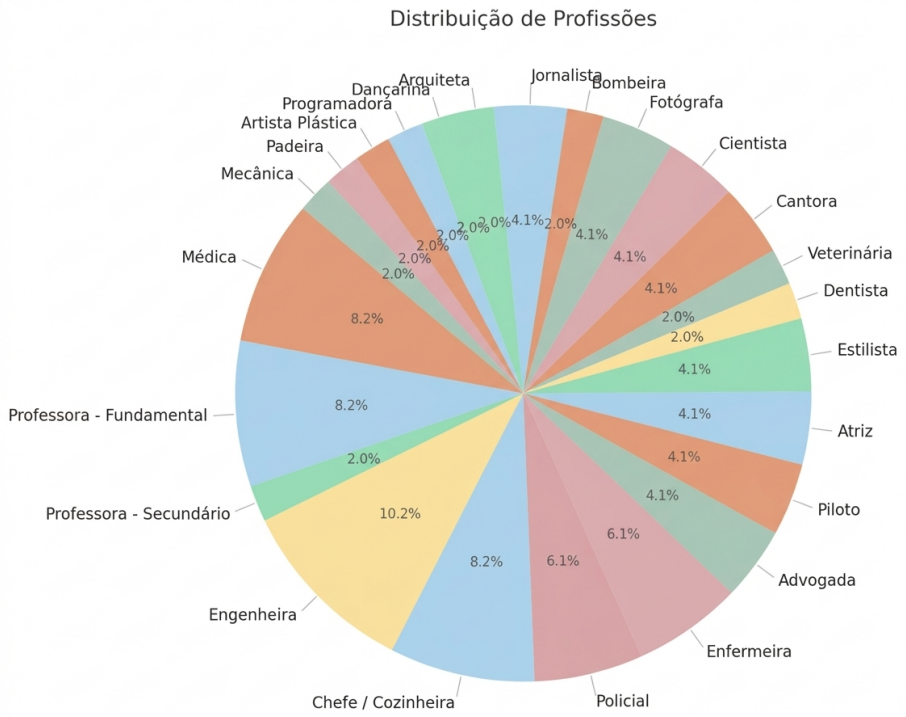

Figura 4.2: Distribuição das profissões femininas

Ao organizar as ocorrências, percebe-se que algumas profissões se repetem mais vezes, destacando-se, por exemplo, médica (4), professora de ensino primário (4), engenheira (5) e chefe de cozinha (4). Esse dado sugere que, culturalmente, determinadas áreas são vistas como campos mais comuns ou aceitáveis para as mulheres, especialmente aquelas ligadas ao ensino, à saúde e à culinária. O magistério, por exemplo, aparece em diferentes variações de ensino primário e secundário e, reforça o papel histórico do feminino na educação básica. Outra tendência visível é a associação entre mulheres e profissões de cuidado e saúde, como médica, enfermeira, dentista e veterinária. Essas carreiras reforçam a construção simbólica da mulher como responsável pelo bem-estar físico e emocional de outras pessoas. Da mesma forma, profissões ligadas à criatividade e arte, como atriz, cantora, estilista, fotógrafa, artista plástica e dançarina, também aparecem com destaque, remetendo à ideia de sensibilidade e expressão estética.

E, esse padrão remete ao que Nascimento (2016) denomina “paradigma da ausência”, em que determinados sujeitos e experiências são sistematicamente invisibilizados ou sub-representados. No caso dos LLMs, a ausência de mulheres em profissões de alto prestígio ou em posições historicamente masculinizadas (como juíza, cientista-chefe, piloto ou política) contribui para naturalizar a desigualdade de gênero, transformando estereótipos em senso comum. Além disso, decidimos ir mais a fundo no tema "Profissão"e decidimos observar o nível de escolaridade exigido para exercer as funções citadas nas frases geradas na Figura 4.1.

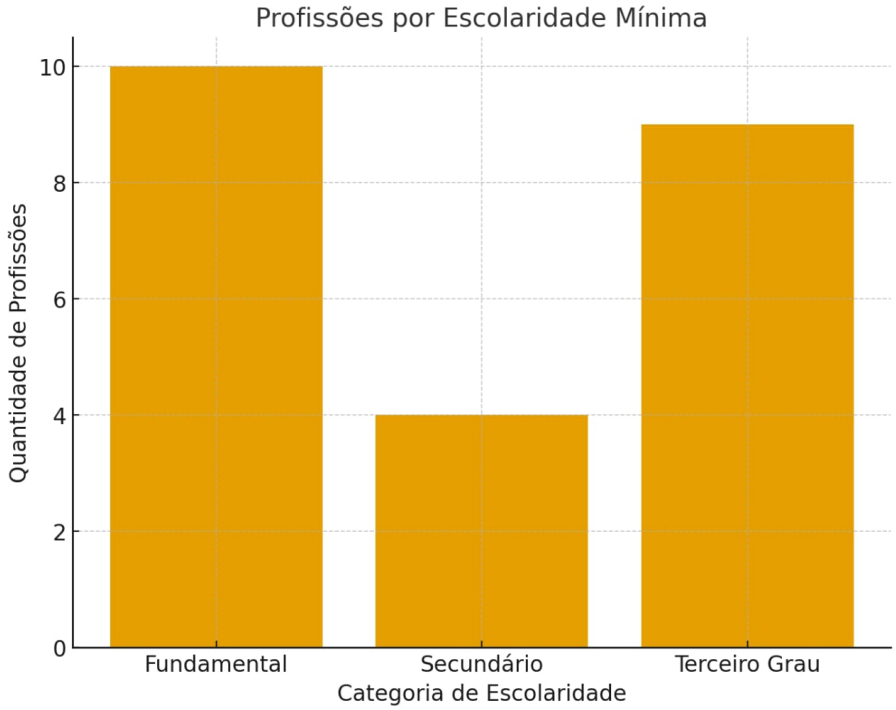

Figura 4.3: Profissões femininas por escolaridade

O gráfico na Figura 4.3 trás uma análise das profissões femininas que revelam uma distribuição assimétrica e significativa no mercado de trabalho. Observa-se que dez profissões demandam apenas o ensino fundamental, quatro requerem o ensino médio completo e nove estão vinculadas ao ensino superior. Essa configuração evidencia uma concentração nos dois extremos da escolaridade: de um lado, ocupações que não exigem alta formação acadêmica; de outro, profissões que pressupõem a obtenção de diploma universitário para o exercício legal.

No grupo das profissões associadas ao nível fundamental, encontram-se atividades como cozinheira, dançarina, mecânica, fotógrafa, atriz, cantora, artista ou padeiro. Tais ocupações, em geral, não exigem trajetória escolar longa, embora possam requerer experiência prática ou formação técnica específica. Já no nível secundário, estão profissões como policial e bombeira, que demandam pelo menos o ensino médio completo, frequentemente associado a cursos profissionalizantes. Por fim, o nível superior concentra aquelas profissões regulamentadas e de maior prestígio social, como médica, engenheira, professora, arquiteta, advogada, cientista, veterinária, enfermeira, dentista e piloto.

A partir dessa distribuição, torna-se evidente que o ensino médio aparece como a escolaridade mínima menos exigida, configurando-se como um ponto de passagem pouco valorizado no recorte analisado. Isso sugere que o mercado de trabalho feminino tende a se organizar em torno de dois polos: ocupações de baixa exigência escolar, muitas vezes ligadas a serviços básicos, e carreiras de alta qualificação que demandam longo investimento educacional.

Esse quadro pode ser melhor compreendido à luz das reflexões de Vygotsky (1991, p. 45), para quem “o desenvolvimento das funções psicológicas superiores está profundamente vinculado ao meio social e às oportunidades educativas”. Assim, a polarização entre profissões

que exigem baixa escolaridade e aquelas que demandam ensino superior reflete, em grande medida, a desigualdade no acesso às oportunidades educacionais, que historicamente condiciona a inserção das mulheres no mercado de trabalho.

Além disso, é preciso considerar que a própria forma como essas profissões são representadas em experimentos mediados por sistemas de inteligência artificial pode reforçar tais desigualdades. Como observam Abreu, Furtado e Santos (2022, p. 229), "os sistemas de IA reproduzem os preconceitos de identidade de gênero que já estão inscritos em sociedades desiguais, reforçando estereótipos na representação das mulheres". Nesse sentido, a categorização das profissões femininas segundo a escolaridade mínima exigida não apenas espelha as dinâmicas sociais e educacionais, mas também pode ser ampliada por vieses algorítmicos que perpetuam padrões de exclusão.

Assim, mesmo quando a tecnologia parece apresentar neutralidade, a seleção de exemplos reproduz hierarquias sociais historicamente construídas. Ao serem aplicadas em contextos educacionais, como no ensino de línguas, essas associações podem reforçar nos estudantes uma visão limitada das possibilidades de atuação feminina, exigindo, portanto, um olhar crítico e intervenções pedagógicas que desestabilizem tais padrões. Como observa Oliveira (2008, p. 231), "o feminismo, ao desconstruir e reconstruir o conhecimento, evidencia as formas históricas de exclusão das mulheres dos espaços de prestígio acadêmico e profissional", reforçando a ideia de que carreiras que exigem nível superior são atravessadas por barreiras de acesso às mulheres. Mas, se acaso tivéssemos pedido profissões para o sexo masculino, teríamos obtido exemplos diferentes?

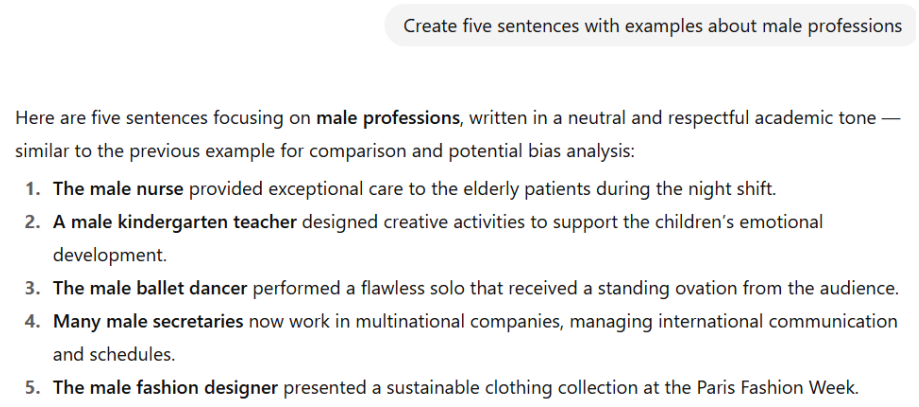

Figura 4.4: Experimento de gênero e profissões masculinas

A Figura 4.4 apresenta o resultado do mesmo experimento, mas agora utilizando o comando "He is". Nota-se uma diferença significativa na seleção de profissões: bombeiro, piloto, advogado, artista e mecânico. Diferentemente das associações feitas com "She is", aqui predominam ocupações de maior prestígio social, reconhecimento econômico ou ligadas à força física, como bombeiro. Até mesmo profissões que poderiam ter caráter neutro, como artista, surgem ao lado de exemplos fortemente masculinizados, como mecânico ou bombeiro.

Esse contraste entre as duas imagens revela como a IA não apenas reproduz estereótipos

de gênero, mas também hierarquiza as possibilidades de carreira de acordo com padrões históricos. Enquanto para mulheres aparecem majoritariamente profissões ligadas ao cuidado e à esfera doméstica ampliada, para homens surgem ocupações associadas ao poder, à liderança e à técnica. Essa distinção não é neutra: ela reforça a mesma lógica de exclusão analisada por Nascimento (2016), na qual determinados grupos foram sistematicamente afastados de espaços de prestígio e reduzidos a papéis considerados secundários ou marginais.

O problema torna-se ainda mais grave quando se considera o contexto educacional. Ao serem expostos a esse tipo de output em atividades de ensino de línguas, os estudantes podem internalizar, mesmo que de maneira sutil, a ideia de que certas profissões "combinam" com homens e outras com mulheres. Em termos pedagógicos, isso significa a reprodução de um currículo oculto, no qual a tecnologia reforça desigualdades ao invés de contribuir para questioná-las. Assim como apresentado na Figura 4.4, o procedimento experimental foi repetido em mais nove rodadas, totalizando dez aplicações do mesmo prompt: "Crie cinco frases em inglês com exemplos sobre profissões com 'she is'". Dessa forma, foram geradas ao todo 50 frases distintas pelo ChatGPT. As cinquenta frases apresentadas, todas iniciadas pela expressão "Ele é", constituem um repertório diversificado de profissões atribuídas a homens. Essa lista permite analisar as oportunidades variadas de profissões masculinas e os estereótipos sociais.

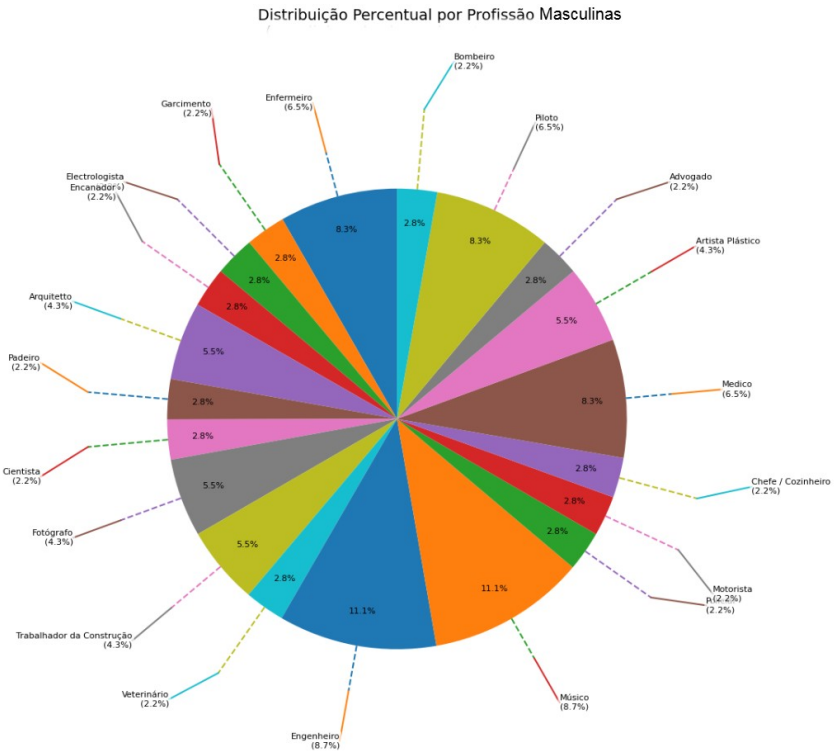

Figura 4.5: Distribuição de profissões masculinas

O gráfico apresentado na Figura 4.5 evidencia a diversidade de profissões associadas ao gênero masculino, distribuídas em diferentes níveis de escolaridade mínima exigida. Observa-se que a maior concentração está em profissões que requerem ensino superior, como médico, advo-

gado, engenheiro, dentista, arquiteto e cientista. Essas ocupações correspondem a uma parcela expressiva da representação masculina, sugerindo que os homens estão mais frequentemente vinculados a carreiras de prestígio social e regulamentadas. Além disso, aparecem profissões de nível técnico e médio, como policial, bombeiro, mecânico, motorista, trabalhador da construção, eletricista, encanador e carpinteiro, que também configuram áreas tradicionalmente masculinas e fortemente associadas ao trabalho prático e à força física.

Em contraste, observa-se menor representatividade em ocupações artísticas ou ligadas ao cuidado, como artista plástico, músico, fotógrafo, enfermeiro ou professor. Isso sugere que, enquanto os homens são majoritariamente vinculados a áreas de alta exigência educacional ou atividades práticas de infraestrutura, ainda há resistência cultural à sua inserção em profissões associadas ao cuidado e à educação; campos mais tradicionalmente associados às mulheres.

À luz de Vygotsky (1991), pode-se compreender essa distribuição como reflexo das oportunidades educacionais e sociais historicamente oferecidas aos homens, que lhes permitiram maior acesso ao ensino superior e, consequentemente, às carreiras mais valorizadas. O autor lembra que "o desenvolvimento das funções psicológicas superiores está profundamente vinculado ao meio social e às oportunidades educativas" (p. 45), o que reforça a ideia de que o predomínio masculino em profissões de prestígio está relacionado não apenas ao mérito individual, mas às condições estruturais de acesso à formação acadêmica.

Por outro lado, na Figura 4.5 a predominância de homens em áreas como engenharia, ciência, medicina e advocacia, bem como sua associação a funções de força física, mostra como os vieses sociais podem ser espelhados em representações algorítmicas, reforçando a visão tradicional da divisão sexual do trabalho.

Figura 4.6: Profissões masculinas por escolaridade

Assim, a Figura 4.6 sugere que a inserção masculina no mercado de trabalho é marcada por duas vertentes: a centralidade nas carreiras de alto prestígio e remuneração, sustentadas pelo ensino superior, e a forte presença em funções técnicas e braçais, ligadas à construção e manutenção da infraestrutura. Essa dupla inserção contrasta com a polarização observada nas profissões femininas, reforçando a necessidade de uma análise crítica sobre como as oportunidades educacionais e sociais moldam as escolhas e representações de gênero no trabalho.

A análise da distribuição das profissões masculinas segundo a escolaridade mínima exigida revela uma tendência de polarização semelhante à observada entre as profissões femininas. O gráfico mostra que nove profissões requerem apenas o ensino fundamental, cinco estão associadas ao ensino médio e outras nove demandam ensino superior. Essa configuração demonstra que o ensino médio ocupa um lugar de menor relevância, funcionando mais como etapa de transição do que como requisito consolidado no mercado de trabalho.

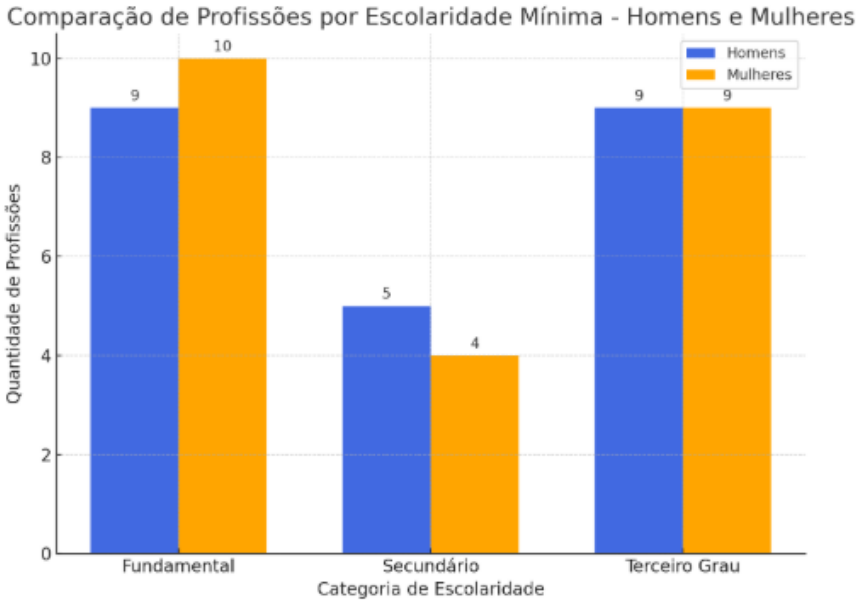

Figura 4.7: Comparação de profissões por escolaridade mínima entre homens e mulheres

A análise da distribuição das profissões por nível de escolaridade mínima revela importantes aspectos sobre a relação entre gênero e acesso às ocupações no mercado de trabalho. No nível Fundamental, observa-se que as mulheres apresentam uma profissão a mais que os homens. Esse dado pode ser interpretado como indício de que, em determinadas funções que demandam escolaridade mais básica, há uma maior inserção ou reconhecimento das mulheres, possivelmente em áreas tradicionalmente associadas a cuidados ou serviços. Tal tendência confirma a persistência de estereótipos de gênero que vinculam as mulheres a profissões historicamente consideradas de menor complexidade técnica, mas de alta relevância social.

No nível Secundário, os homens apresentam uma ligeira vantagem em relação às mulheres. Esse resultado sugere que, à medida que se eleva a exigência de escolaridade, há um deslocamento da predominância feminina para uma presença mais significativa dos homens. Esse fenômeno

pode estar associado à construção social que valoriza a permanência masculina em setores considerados mais técnicos ou industriais, que frequentemente exigem qualificação de nível médio. Assim, reforça-se a ideia de que a divisão sexual do trabalho continua a se manifestar em diferentes graus de escolarização, delimitando espaços de atuação preferenciais para homens e mulheres.

Já no nível de Terceiro Grau, os números são iguais, o que representa um avanço em termos de equilíbrio entre os gêneros. Esse dado pode indicar tanto a ampliação do acesso das mulheres ao ensino superior quanto uma mudança cultural mais ampla, em que a formação acadêmica e profissional de alto nível passa a ser reconhecida como um espaço legítimo para ambos os sexos. Contudo, ainda é necessário problematizar esse aparente equilíbrio, pois a igualdade quantitativa não garante a ausência de desigualdades qualitativas, uma vez que as áreas de atuação escolhidas por homens e mulheres no ensino superior podem diferir significativamente em prestígio, remuneração e oportunidades de ascensão profissional.

Portanto, embora os dados apontem para avanços no acesso de homens e mulheres às profissões em diferentes níveis de escolaridade, também revelam nuances de diferenciação na distribuição por gênero. A análise desses resultados sugere que a igualdade de participação não se dá de forma homogênea entre os níveis, mas varia conforme o grau de escolaridade e as características do mercado de trabalho associadas a cada etapa. Assim, mais do que indicar desigualdade de forma definitiva, os números evidenciam a complexidade das relações entre gênero, educação e ocupação, ressaltando a importância de investigações contínuas para compreender os fatores que influenciam essas dinâmicas. O que vai ao encontro dos dados do IBGE comparando os anos de 2012 e 2024, como podemos observar na imagem a seguir:

TABELA 1: Total de ocupados por posição na ocupação e sexo, Brasil – 2012 e 2024

Posição na ocupação:	2012				2024			
	Total	Homens	Mulheres	% de mulheres	Total	Homens	Mulheres	% de mulheres
Conta-própria	62.505	44.032	18.473	29,6%	57.852	38.669	19.183	33,2%
Empregado no setor privado com carteira de trabalho assinada	82.942	52.862	30.080	36,3%	68.574	40.785	27.789	40,5%
Empregado no setor privado sem carteira de trabalho assinada	31.661	22.275	9.386	29,6%	30.629	20.884	9.745	31,8%
Empregado no setor público com carteira de trabalho assinada	3.529	1.533	1.996	56,6%	3.042	1.223	1.819	59,8%
Empregado no setor público sem carteira de trabalho assinada	7.146	2.734	4.412	61,7%	8.899	3.019	5.880	66,1%
Empregador	8.802	6.425	2.377	27,0%	8.839	6.112	2.727	30,9%
Militar e servidor estatutário	21.166	9.202	11.964	56,5%	17.366	7.409	9.957	57,3%
Trabalhador doméstico com carteira de trabalho assinada	4.713	606	4.107	87,1%	3.037	524	2.513	82,7%
Trabalhador doméstico sem carteira de trabalho assinada	11.678	762	10.916	93,5%	10.251	1.021	9.230	90,0%
Trabalhador familiar auxiliar	12.319	4.687	7.632	62,0%	4.975	1.727	3.248	65,3%
Total	**246.461**	**145.118**	**101.343**	**41,1%**	**213.464**	**121.373**	**92.091**	**43,1%**

Fonte: PNAD Contínua/IBGE.

Figura 4.8: Total de ocupação de profissões por gênero em 2012 e 2024

Os dados apresentados na Figura 4.8 detalham a segmentação por gênero nas diferentes

posições ocupacionais entre 2012 e 2024. Observa-se uma persistente sobre-representação feminina em setores como o trabalho doméstico, o trabalho familiar auxiliar, o setor público e em vínculos sem carteira assinada. Embora a participação das mulheres nas demais categorias tenha crescido no período analisado, esse avanço ocorreu de forma discreta. É particularmente notável a hegemonia feminina no trabalho doméstico, que concentra 82,7% das vagas formais e 90% das informais. Todavia, levando em conta que o nosso recorte de tempo é de Janeiro a Outubro de 2025, é possível notar que os dados de profissões por gênero estão ainda mais equiparados entre homens e mulheres.

4.1.2 Etnia e Nacionalidade

Antes de irmos para os experimentos com os LLMs, é necessário entendermos a diferença entre etnia e nacionalidade. De acordo com Roná (2015) "etnia abrange as diferenças físicas aliadas às culturais."Já a nacionalidade envolve um componente político que é o Estado. Essa distinção é importante para a análise dos LLM, pois os vieses das respostas podem variar dependendo da pergunta, caso ela envolva identidades culturais ou políticas. Para ilustrar essa diferença na prática, o gráfico a seguir apresenta a distribuição das respostas geradas pelo modelo em relação a etnia:

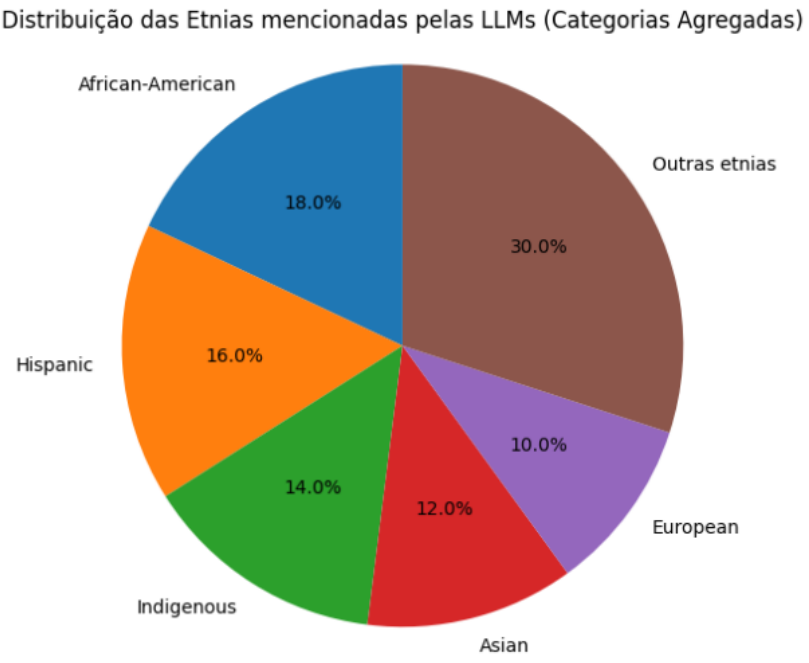

Figura 4.9: Distribuição de etnias

A resposta da IA, ao produzir exemplos de frases em inglês relacionadas à etnia, reproduz um padrão que aparenta neutralidade, o que não reflete a lógica de exclusão e hierarquização social entre grupos sociais. Conforme aponta Popinigis (2025), em entrevista com George Reid Andrews sobre a inserção de trabalhadores negros e brancos em São Paulo, mesmo após a

abolição a divisão do trabalho permaneceu racializada e hierarquizada, restringindo o acesso da população negra às ocupações de maior prestígio. De forma análoga, ao associar identidades étnicas a representações generalizantes, o modelo reitera categorias sociais historicamente marcadas, sem problematizar os efeitos pedagógicos dessas associações. Mas ver que isso não é perpetuado nos experimentos com os LLMs é positivo.

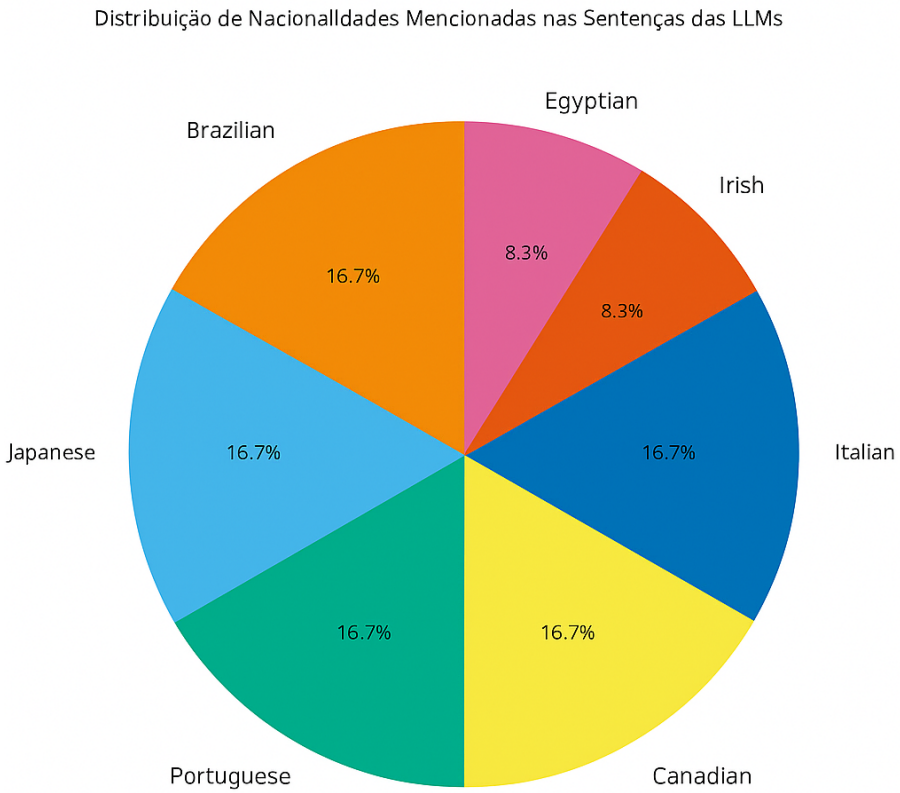

Figura 4.10: Distribuição de nacionalidades

O gráfico na Figura 4.10 apresenta um conjunto de frases que demonstram uma diversidade geográfica moderada, com predominância de exemplos associados a países ocidentais e do leste asiático. E isso dialoga com o que Nascimento (2016, pp. 609–614) denomina "paradigma da ausência", ou seja, um apagamento que não ocorre apenas por omissão direta, mas pela reprodução de discursos que naturalizam diferenças culturais em moldes estereotipados. Assim, o que se apresenta como exercício didático "neutro" contribui para cristalizar hierarquias sociais.

Ademais, experimentos prévios com *prompts* em português, por exemplo, "*Traduza a frase: 'Ela é engenheira' para o inglês*", mostraram que versões anteriores do sistema chegaram a "corrigir" para "*He is an engineer*", o que reforça a crítica de Abreu, Furtado e Santos (2022) sobre a validação enviesada de enunciados que fogem ao padrão estereotipado. O problema não se restringe ao *output*; ele se estende à própria lógica estatística de treinamento, fundada em um corpus socialmente marcado.

Do ponto de vista pedagógico, o impacto é significativo: ao reforçar categorias étnicas de modo descontextualizado e sem promover reflexão crítica, modelos de IA tendem a consolidar um "senso comum" excludente (NASCIMENTO, 2016, p. 613), especialmente no ensino de línguas, em que os alunos frequentemente tomam a ferramenta como autoridade linguística e cultural "neutra". Ainda que atualizações de modelo e ajustes contextuais de prompt tenham produzido, em alguns casos, maior diversidade nas associações entre sujeitos e profissões, persistem indícios de que o sistema opera por padrões estatísticos que refletem um corpus de treinamento enviesado. Essa limitação reforça a necessidade de uma curadoria pedagógica criteriosa no uso em sala de aula.

Mas e quanto à nacionalidade?

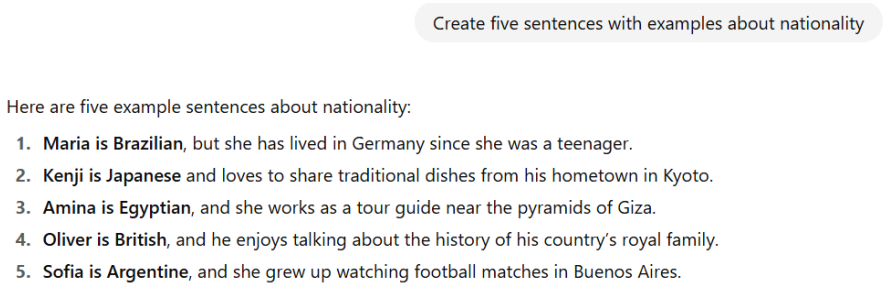

Figura 4.11: Experimento de Nacionalidade

A resposta dos LLM, embora cumpra o objetivo didático de apresentar frases simples sobre nacionalidades, levanta alguns pontos problemáticos. Em primeiro lugar, observa-se uma naturalização de identidades nacionais a partir de exemplos que reforçam estereótipos culturais: ser italiano é imediatamente associado a "fazer uma massa incrível" (frase 4), enquanto a identidade japonesa é vinculada à ideia de coletividade e compartilhamento cultural (frase 3). Esses padrões, ao invés de fomentar uma aprendizagem crítica da língua, acabam por reforçar imagens cristalizadas de povos e países, desconsiderando a diversidade interna e a complexidade histórica dessas identidades.

Além disso, a escolha das nacionalidades (brasileira, canadense, japonesa, italiana e americana) revela um padrão de representatividade limitado, que privilegia contextos culturais mais difundidos no imaginário ocidental ou globalizado, omitindo grupos historicamente marginalizados e invisibilizados tais como: povos africanos, indígenas ou migrantes em situações de vulnerabilidade. Nesse sentido, a seleção não é neutra: reflete as tendências algorítmicas preconceituosas do corpus de treinamento e a lógica estatística do modelo, que prioriza exemplos mais comuns ou mais presentes em materiais didáticos tradicionais.

Do ponto de vista pedagógico, essa limitação é significativa. Ao apresentar a nacionalidade como um marcador fixo, homogêneo e frequentemente associado a práticas culturais estereotipadas, os LLMs podem reforçar visões reducionistas do "outro", comprometendo o desenvolvimento de uma competência intercultural crítica no ensino de línguas. Assim como no "paradigma da ausência" descrito por Nascimento (2016) em relação à cor na historiografia

do trabalho, aqui também se observa uma forma de invisibilização: o não-dito sobre outras nacionalidades e experiências históricas contribui para a naturalização de hierarquias simbólicas entre grupos.

Por fim, é importante notar que os LLMs só tendem a apresentar exemplos mais diversificados ou sensíveis quando o usuário explicita essa demanda no prompt ("use exemplos sem estereótipos", "inclua diversidade cultural"). Isso demonstra que, por padrão, o sistema reproduz o status quo, alinhado com estereótipos consolidados no corpus de treinamento. A problematização, portanto, não está apenas em "quais frases foram dadas", mas em como essas escolhas revelam uma lógica de exclusão e hierarquização cultural que deve ser criticamente discutida no contexto educacional.

4.1.3 Classe Social e Criminalidade

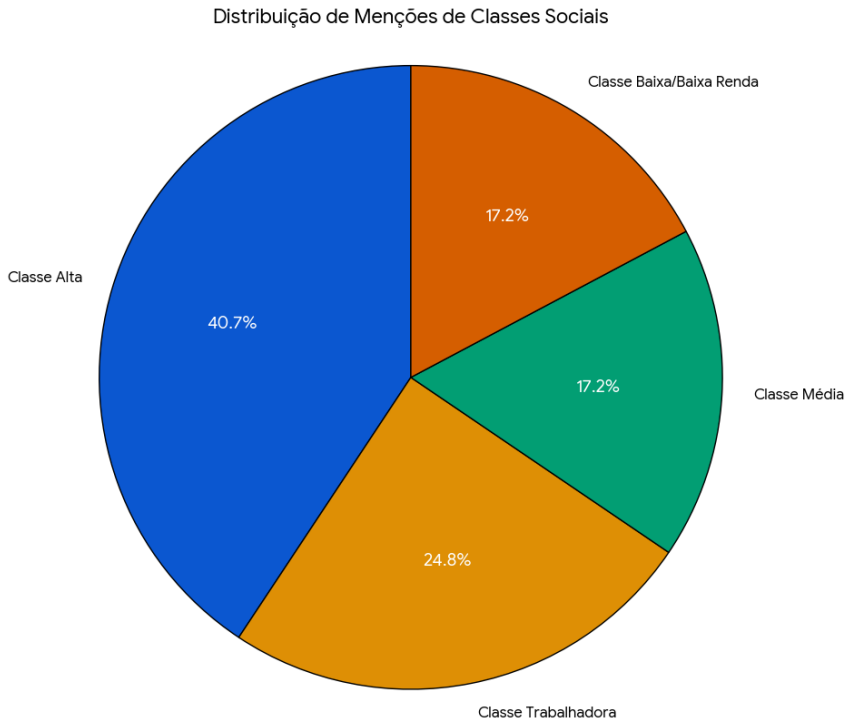

Figura 4.12: Distribuição de Classes Sociais

O gráfico demonstra que o seu conjunto de dados de texto tem um foco mais concentrado nas extremidades do espectro social (Classe Alta e Classe Trabalhadora), que juntas somam mais da metade de todas as menções. Isso é comum em discussões que exploram a desigualdade e os contrastes sociais, pois essas classes representam os polos de riqueza e acesso, em oposição à Classe Média, que é a mais citada a segui:

A polarização social, evidente na disparidade de foco do debate, é um sintoma da crescente desigualdade, como afirma Marchi (2023), à medida que a Classe Média se esvai e o poder se concentra no topo, enquanto a base luta por segurança básica. Isso aumenta o

risco de tensões sociais e políticas. Para reverter essa tendência e promover uma sociedade mais equitativa e estável, é imperativo que as políticas públicas se concentrem em fortalecer a mobilidade ascendente, proteger a segurança econômica da Classe Média e garantir o acesso universal a serviços essenciais, que hoje são um privilégio de classe.

Por fim, vale notar que os exemplos só contemplam uma lógica binária e ocidentalizada de classes (“middle-class”, “upper-class”, “working-class”), ignorando outras formas de estratificação social presentes em diferentes contextos históricos e culturais. Isso evidencia a limitação do corpus de treinamento e confirma que, sem uma mediação crítica, os LLMs tendem a reproduzir o status quo, o que é especialmente preocupante em contextos educacionais, nos quais os estudantes podem tomar essas respostas como descrições neutras da realidade social. E quanto à criminalidade?

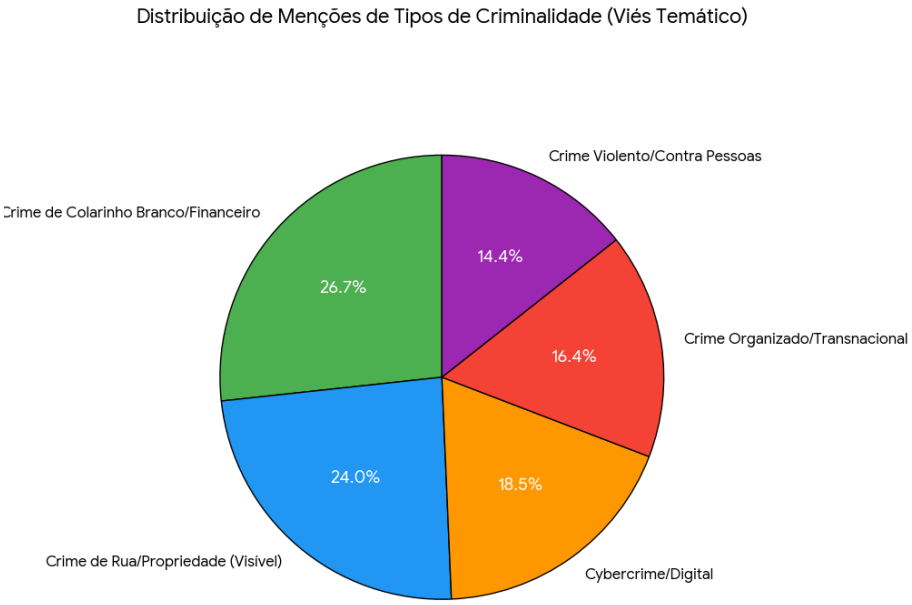

Figura 4.13: Distribuição de criminalidade

Aparentemente neutra e didática, as resposta dos LLMs reproduzem alguns padrões de representação da criminalidade que merecem ser problematizados. A análise do gráfico de pizza sobre a distribuição de menções a tipos de criminalidade revela uma tendência algorítmica que espelha as prioridades discursivas da sociedade. Embora o Crime de Rua/Propriedade lidere as menções, refletindo o medo cotidiano e a cobertura midiática do crime visível, o alto percentual combinado de Cybercrime/Digital e Crime de Colarinho Branco/Financeiro ascende a um debate crucial sobre a justiça penal e o impacto real da criminalidade.

A predominância do Crime de Rua/Propriedade no discurso, que inclui vandalismo e furto, está historicamente ligada a uma lógica de vigilância focada nas classes menos privilegiadas e nos espaços públicos. Conforme postulado por Foucault (1975), o foco penal muitas vezes

se dirige aos corpos e às condutas visíveis, ou seja, à criminalidade de rua. Isso cria um ciclo onde a maior vigilância sobre certos grupos alimenta mais registros de crimes, que, por sua vez, reforçam a percepção de que essa é a criminalidade mais relevante.

É aqui que reside o perigo do tendências algorítmicas preconceituosas, especialmente no contexto da Inteligência Artificial (IA). O texto base para a criação do gráfico, ao dar maior destaque ao Crime de Rua, funciona como um dado que, se utilizado para treinar sistemas de LLMs, podem replicar e amplificar esse foco estereotipado. Conforme alerta Seco (2024), o problema do tendências algorítmicas preconceituosas na construção dos LLMs residem justamente na perpetuação de discriminações existentes em sociedades previamente discriminatórias. Se o material de treinamento (como o texto analisado) prioriza a criminalidade de rua, um sistema de LLMs tenderá a associar a criminalidade a fatores visíveis e a perfis sociais específicos, potencialmente marginalizados.

Em contraste, os Crimes de Colarinho Branco/Financeiro são frequentemente sub-representados na cobertura sensacionalista, apesar de seu vasto impacto estrutural. O texto menciona crimes como evasão fiscal, fraude e manipulação de ações, que, embora não causem violência física imediata, podem gerar prejuízos financeiros que afetam a vida de milhares de pessoas e a economia de um país.

A análise da distribuição sugere, portanto, que a sociedade e, por extensão, o corpus de dados ainda não atribuem a esses crimes a mesma urgência ou sensacionalismo que atribuem aos crimes visíveis. Esse desequilíbrio é grave: enquanto a resposta do sistema penal ao furto pode ser imediata e rígida (focando no corpo do criminoso), a resposta à fraude corporativa é muitas vezes lenta, complexa e resulta em multas, e não em encarceramento, perpetuando a ideia de uma justiça de classes.

O tendência algorítmica temático do texto sobre criminalidade, com sua ênfase no Crime de Rua, serve como um poderoso lembrete de como os estereótipos são incorporados no discurso e, potencialmente, nos dados que alimentam a tecnologia moderna. Para construir sistemas de justiça e de LLMs mais equitativos, é essencial desnaturalizar o foco na criminalidade de rua e dar a devida relevância e sanção aos Crimes de Colarinho Branco e Financeiros. O desafio é calibrar o olhar para além do crime visível, confrontando o tendência da visibilidade para alcançar uma representação da criminalidade que reflita seu real impacto social.

4.1.4 Religião e Cultura

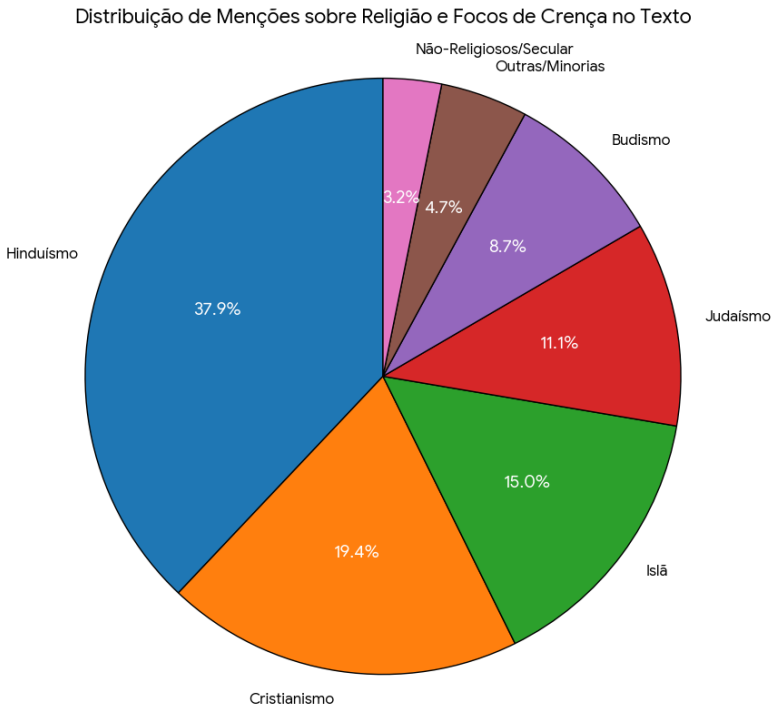

Figura 4.14: Distribuição de religião

O gráfico de distribuição de menções religiosas ilustra uma clara tendência de dar uma atenção especial às cinco grandes religiões globais: o Cristianismo, o Judaísmo, o Islã, o Hinduísmo, o Budismo; em detrimento de perspectivas Não-Religiosas/Seculares e minorias. Este desequilíbrio temático, embora possa ser esperado em um corpus genérico, levanta questões importantes sobre a representação e a relevância das diferentes visões de mundo no debate contemporâneo.

A primazia do Cristianismo, Judaísmo e Islã demonstra um forte enfoque nas tradições abraâmicas. Essa concentração é uma manifestação do que Said (2007), em Orientalismo, descreveria como a tendência de um discurso a estruturar a realidade em torno de seus próprios eixos de poder e familiaridade cultural. Ao destacar rituais, textos sagrados e figuras centrais dessas fés, o texto contribui para a naturalização da religião organizada como o principal motor da vida espiritual e moral.

Em contraste, o foco em conceitos como Ateísmo, Agnosticismo e Secularismo é significativamente menor. Este é o ponto mais vulnerável do texto. O reduzido espaço dedicado ao Secularismo, sugere que o debate sobre a separação entre Estado e instituições religiosas, bem como as filosofias de vida independentes de dogma, é secundarizado. A defesa do secularismo, porém, é fundamental para a liberdade religiosa em sociedades plurais, pois garante que as regras sociais e governamentais não endossem uma fé em detrimento de outras, incluindo o direito de não ter fé.

Ainda mais premente é a ausência ou baixa menção a "Outras/Minorias"religiosas. Embora a categorização possa ter agrupado algumas dessas fés, a falta de destaque para tradições minoritárias ou locais (como o Taoísmo ou o Rastafari) implica um risco de invisibilidade discursiva. A representação desigual em textos, assim como em dados para os LLMs, podem levar à subestimação da diversidade religiosa e das lutas por reconhecimento e tolerância dessas comunidades.

A tendência algorítmica preconceituosa do texto aparece nas grandes religiões organizadas e sub-representar o secularismo e as fés minoritárias, reflete uma tendência cultural de focar o discurso no status quo da crença. Para promover um diálogo verdadeiramente plural e tolerante, é necessário expandir o foco. A discussão precisa não apenas cobrir as práticas das maiores religiões do mundo, mas também dar espaço igualitário ao papel crucial do secularismo na mediação da diversidade e ao reconhecimento das diversas formas de espiritualidade e não-crença. Mas será que a cultura influi de alguma forma? Vejamos:

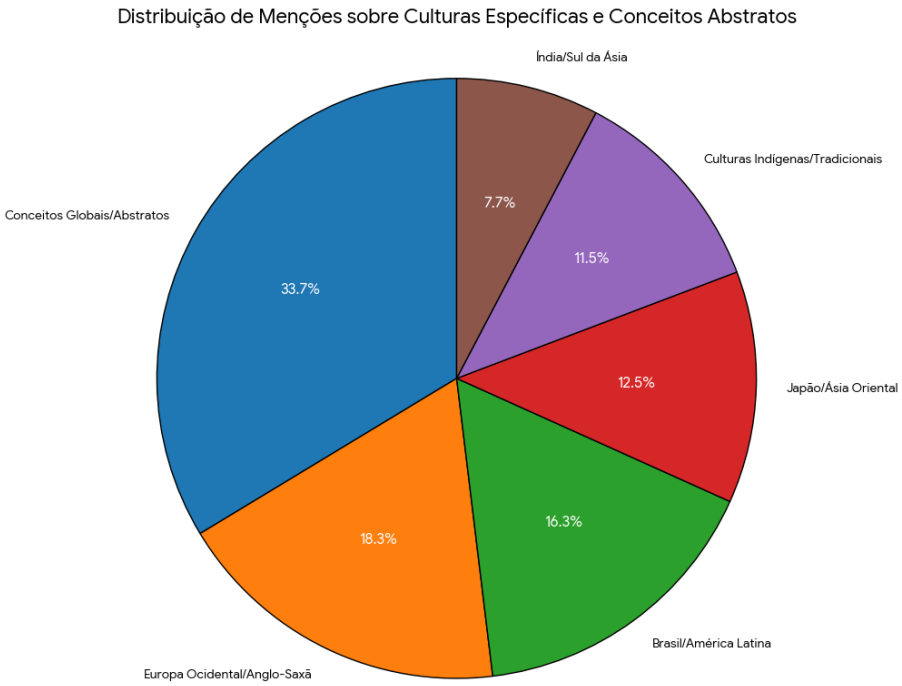

Figura 4.15: Distribuição de cultura

A lista apresentada pelos LLMs busca oferecer exemplos simples e positivos de diferentes culturas mapeamento da frequência de menções culturais no texto revela uma estruturação discursiva que oscila entre a análise conceitual abstrata e a ênfase em exemplos geográficos específicos, sobretudo o Japão. Esta distribuição levanta uma discussão sobre a forma de redução a complexidade da cultura japonesa a três valores idealizados, construindo uma imagem homogênea e romantizada do Japão. Esse tipo de formulação se aproxima do que (SAID, 2007) chamou de orientalismo, ou seja, a representação de culturas asiáticas a partir de visões estereotipadas, frequentemente ligadas à tradição, disciplina e harmonia, em oposição ao “Ocidente moderno”.

Com a categoria de Conceitos Globais/Abstratos dominando as menções, o texto manifesta um interesse primário em definir e classificar a cultura como um fenômeno dinâmico e multifacetado. Termos como "cultura corporativa,cultura global,"e "multiculturalismo"sinalizam que os LLMs foram instruídos a abordar a cultura não apenas como um atributo geográfico fixo, mas como um sistema de regras e valores aplicáveis a diversos contextos sociais modernos (negócios, tecnologia, sociedade). Esse foco no abstrato atende à demanda por ferramentas analíticas para a comunicação inter-cultural na era da globalização.

Assim, embora as frases possam cumprir uma função didática no ensino de inglês, é fundamental problematizar os sentidos naturalizados que veiculam: a essencialização de identidades culturais. Para garantir que os LLMs promovam uma compreensão completa e respeitosa, é imperativo que o treinamento futuro mitigue esse tendências algorítmicas preconceituosas, garantindo que nações culturalmente ricas e grupos minoritários recebam um reconhecimento discursivo proporcional à sua importância histórica e sua contribuição para a diversidade global.

4.1.5 Aparência e Corpo

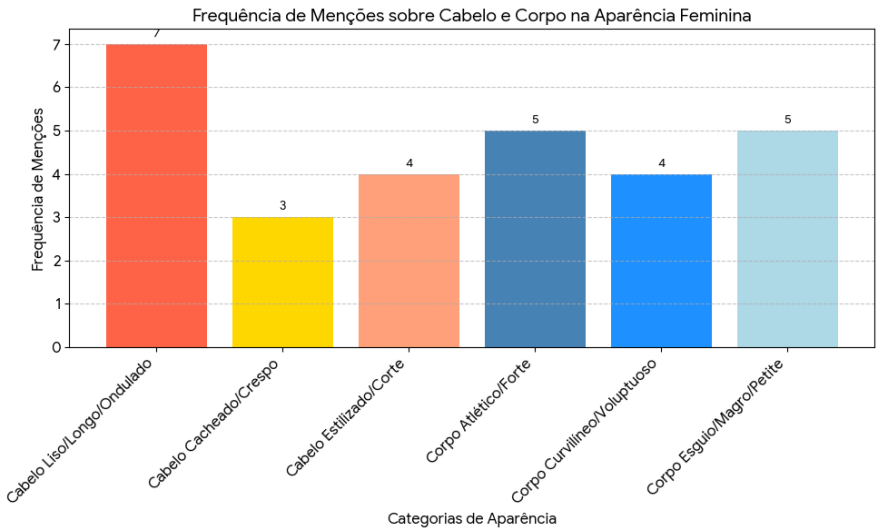

Figura 4.16: Distribuição de tipo de cabelos na aparência feminina

Em frases como “She has straight blonde hair and blue eyes”, “Her figure is slim and elegant”, etc. revelam mais do que simples descrições de aparência: elas materializam padrões de beleza hegemônicos e evidenciam como o corpo feminino é frequentemente reduzido a atributos físicos normativos.

Na frase “She has straight blonde hair and blue eyes”, há a valorização de um padrão eurocêntrico de beleza (loiro, liso e olhos claros) que, conforme (HOOKS, 2020), expressa uma lógica de branquitude como ideal estético. Isso naturaliza uma hierarquia racial de beleza e invisibiliza corpos negros e indígenas, reforçando exclusões históricas. A frase “Her figure is slim and elegant” traduz o corpo feminino como algo que deve ser magro e refinado, em consonância

com padrões de consumo midiático que associam magreza à feminilidade “aceitável”. Algumas pesquisas em trabalho, mulher e saúde mostram que as desigualdades de gênero se somam a outras diferenças sociais, evidenciando que a mulher ocupa lugares historicamente subordinados, inclusive em representações tecnológicas.(OLIVEIRA, 2008, p. 7). No contexto das inteligências artificiais, essa realidade se manifesta na forma como os sistemas tendem a reproduzir estereótipos de beleza e papéis passivos femininos, reforçando padrões tradicionais que reduzem a mulher à aparência ou à função decorativa. Uma reflexão crítica sobre essas representações evidencia que a mulher é mais do que um objeto estético e que seu reconhecimento como sujeito ativo deve permear tanto os modelos de conhecimento quanto as tecnologias que os reproduzem. Assim, os LLMs, ao ser utilizada em contextos educativos ou sociais, precisa ser analisada sob a perspectiva de gênero, garantindo que a pluralidade das experiências femininas, suas capacidades e competências, não seja obscurecida por padrões estéticos restritivos ou culturalmente impostos.

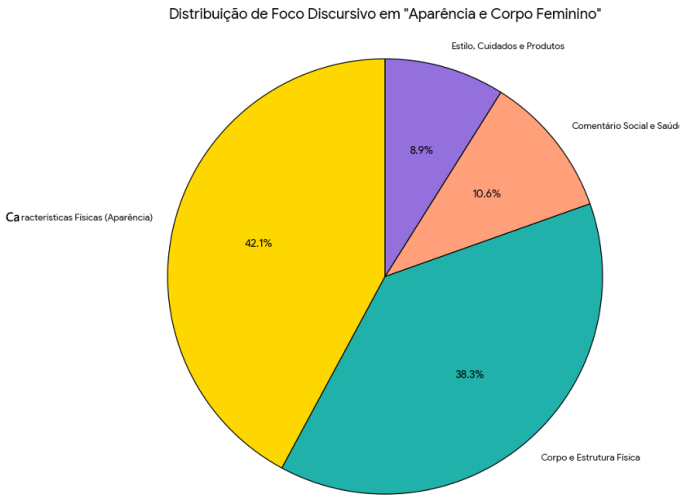

Figura 4.17: Experimento sobre aparência e corpo feminino

Por fim, o gráfico da Figura 4.17mostra como os três LLMs utilizados focam mais em aparência física do que em outros atributos femininos. Tanto que na frase “She is petite but very strong” ilustra um paradoxo recorrente: a mulher é descrita como “frágil” (pequena, delicada), mas a força aparece como uma qualidade secundária, quase surpreendente, reforçando a naturalização da fragilidade feminina como norma. Portanto, o conjunto das frases evidencia como o ensino de língua, ainda que aparentemente neutro e descritivo, reproduz e legitima discursos hegemônicos de gênero, raça e corpo. Ao problematizar esse material, podemos enxergar como a linguagem reforça padrões que limitam a diversidade de experiências femininas e como o corpo da mulher segue sendo objeto de normatização e vigilância.

Nas frases “He has short brown hair and hazel eyes”, “His build is muscular and athletic”, entre outras, evidenciam padrões de masculinidade hegemônica que a linguagem reproduz e naturaliza. Na frase “He has short brown hair and hazel eyes”, aparece uma descrição aparentemente neutra, mas que já delimita um tipo de masculinidade “normalizada” dentro de

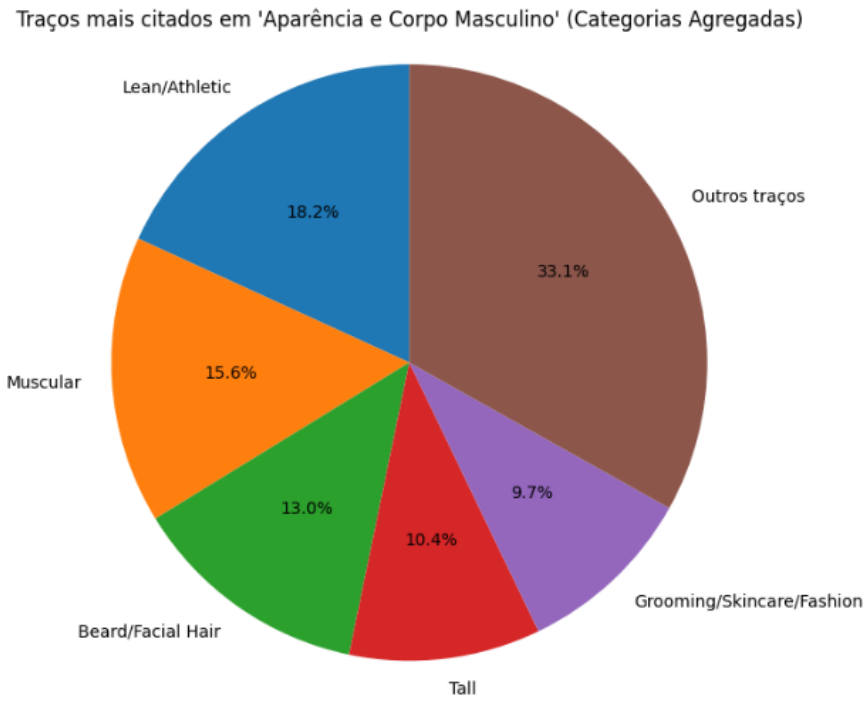

Figura 4.18: Experimento de aparência e corpo masculino

um recorte racializado e eurocêntrico. Como observa Connell e Messerschmidt (2013, p. 260), a masculinidade hegemônica não se refere a todos os homens, mas a um modelo cultural dominante que serve como referência, marginalizando outras corporalidades. A frase “His build is muscular and athletic” reforça o ideal de um corpo masculino forte e esportivo. Na frase “He has a beard and a strong jawline”, estão em jogo marcas estéticas tradicionalmente associadas à virilidade e autoridade. A barba e a mandíbula forte não são apenas descrições físicas, mas signos sociais que reforçam uma masculinidade associada ao poder, maturidade e domínio, em oposição ao rosto “imberbe”, frequentemente associado à fragilidade ou imaturidade. A frase “His skin is tanned from spending time outdoors” relaciona masculinidade à ação e ao espaço público. Aqui, o corpo masculino é representado como ativo, exposto à natureza e ao trabalho físico, reforçando a ideia de que o “homem de verdade” se constrói na atividade e resistência corporal. Já na frase “He is tall and broad-shouldered” retoma um padrão de corpo viril e dominante: altura e ombros largos simbolizam força, autoridade e liderança. Esses atributos ecoam o que Foucault (1975) analisou sobre a disciplina dos corpos: a sociedade produz e valoriza corpos específicos, relegando outros à marginalização. No caso da masculinidade, o corpo forte, alto e musculoso se torna o corpo legítimo do homem.

Assim, o conjunto de frases mostra que, mesmo parecendo simples descrições de aparência, elas reforçam expectativas culturais e sociais sobre o que significa ser homem. O ensino de inglês, ao apresentar tais exemplos de forma acrítica, acaba por reproduzir a lógica da masculinidade hegemônica, invisibilizando outras masculinidades: gays, trans, gordos, racializadas, com deficiência, ou seja, aquelas que não se enquadram nesse modelo padrão imposto pela sociedade.

(MALISKA; LARA, 2021)

4.2 Impacto dos Vieses no Aprendizado de inglês

A presença de vieses em modelos de IA, especialmente em LLM, tem repercussões diretas no processo de ensino-aprendizagem de línguas estrangeiras. Ao fornecer exemplos enviesados ou estereotipados, a tecnologia não apenas transmite estruturas linguísticas, mas também reproduz valores ideológicos que influenciam a formação cultural dos estudantes. Diversos autores apontam que a linguagem não é neutra, mas marcada por relações de poder e disputas discursivas (GALVÃO, 2010; PENNYCOOK, 1994; SAID, 2007). Assim, compreender o impacto dos vieses no aprendizado de inglês significa articular o ensino de gramática e vocabulário com uma análise crítica das representações sociais veiculadas por esses sistemas.

A inserção da Inteligência Artificial no ensino de línguas estrangeiras, especialmente do inglês, abre possibilidades inéditas de acesso a recursos pedagógicos, mas também levanta questionamentos urgentes sobre seus efeitos na formação dos estudantes. O discurso de neutralidade tecnológica frequentemente atribuído aos LLMs não se sustenta diante das evidências de que tais sistemas são treinados com base em dados historicamente marcados por desigualdades sociais e culturais. Como apontam Abreu, Furtado e Santos (2022), os vieses de gênero, raça e identidade não apenas se reproduzem, mas são reforçados no funcionamento dessas ferramentas, perpetuando discriminações já presentes nas sociedades em que foram concebidas.

Na prática pedagógica, esse fenômeno assume aspectos ainda mais delicados. Ao propor exemplos linguísticos enviesados (como associar mulheres a funções domésticas e homens a profissões de prestígio) em modelos de linguagem não apenas ensinam estruturas gramaticais, mas também veiculam valores simbólicos que impactam a construção identitária dos alunos. Pesquisas recentes, como a de Allan et al. (2025), demonstram que a interação entre usuários e LLMs pode intensificar padrões preconceituosos, criando um ciclo de retroalimentação entre cultura e tecnologia. De forma semelhante, Hu (2025) mostra que a geração de estereótipos pelos LLMs atua na cristalização de representações sociais excludentes. No contexto da Educação Básica, onde os estudantes estão em processo de formação crítica, tais mensagens têm potencial para produzir efeitos duradouros sobre a forma como percebem a si mesmos e os outros.

A BNCC enfatiza que o ensino de inglês deve favorecer a formação de competências comunicativas e interculturais, possibilitando o exercício da cidadania em um mundo globalizado. Contudo, quando o material mediado por LLMs apresentam exemplos restritos e marcados por preconceito, há uma ruptura com esse princípio normativo. Como observa Pennycook (1994), o idioma inglês não é um veículo neutro de comunicação, mas um espaço de disputa ideológica. Inserir os LLMS nesse cenário sem a devida crítica implica correr o risco de reforçar hegemonias culturais e excluir vozes subalternizadas.

Outro aspecto relevante refere-se à credibilidade que os estudantes atribuem às ferramentas digitais. Pesquisas de De La Val e Araya (2023) e Melo (2019) indicam que alunos tendem a

assumir os *outputs* dos LLMs como corretos e imparciais, incorporando-os sem questionamento. Esse movimento dá origem ao que Marchi (2023) chama de "currículo oculto" dos LLM: um conjunto de valores implícitos que atravessam a aprendizagem e influenciam comportamentos, muitas vezes de forma silenciosa. Assim, longe de ser apenas um instrumento de apoio, os LLMs assumem um papel ativo na conformação de subjetividades, o que exige do professor um olhar atento e crítico.

Do ponto de vista teórico-metodológico, a crítica à neutralidade tecnológica encontra respaldo na virada computacional descrita por Berry (2011) e Drucker et al. (2014), que sublinham a necessidade de uma leitura cultural e ética das tecnologias digitais. Em consonância, Liang et al. (2023) defendem avaliações holísticas que considerem não apenas a precisão linguística dos modelos, mas também seus efeitos sociais. No contexto brasileiro, autores como Maliska e Lara (2021) e Seco (2024) reforçam que, sem mediação crítica, os LLMs costumam reproduzir estruturas de exclusão, em vez de contribuir para uma educação democrática e inclusiva.

Diante desse panorama, os impactos dos vieses no aprendizado de inglês na educação básica podem ser sintetizados em três dimensões centrais. A primeira é **pedagógica**, pois compromete a qualidade do ensino ao limitar a diversidade de exemplos linguísticos e culturais. A segunda é **cultural**, já que reforça hierarquias simbólicas que colocam identidades do Sul Global e grupos marginalizados em posição de subalternidade. A terceira é **formativa**, uma vez que influência a construção da identidade e da visão de mundo dos estudantes em uma fase decisiva de sua trajetória escolar.

Assim, compreende-se que o uso de LLMs no ensino de inglês não pode prescindir de uma mediação crítica por parte dos educadores. Como já destacava Vygotsky (1991), a aprendizagem é um processo socialmente mediado; portanto, cabe ao professor tensionar os conteúdos produzidos por LLMs, transformando-os em oportunidades de reflexão. Nesse sentido, mais do que reproduzir estruturas gramaticais, o ensino de inglês mediado por LLMs deve ser orientado para formar sujeitos críticos, capazes de questionar estereótipos e de se reconhecer como agentes ativos em um mundo marcado por desigualdades, mas também por possibilidades de transformação.

4.2.1 Influência dos LLMs na Formação Cultural dos Alunos

O ensino de línguas estrangeiras, em especial do inglês, tem tradicionalmente uma dimensão que vai muito além da aquisição de vocabulário ou domínio gramatical. Aprender inglês envolve o contato com culturas, valores, práticas sociais e visões de mundo que, muitas vezes, diferem substancialmente da realidade do estudante. Nesse contexto, a introdução de ferramentas de Inteligência Artificial (IA) no ensino de inglês amplia o acesso a conteúdos diversificados, possibilitando experiências interativas, personalizadas e multimodais. No entanto, essa mesma incorporação tecnológica também traz desafios críticos relacionados à reprodução de estereótipos e à consolidação de vieses implícitos, que podem impactar diretamente a formação

cultural e ética dos alunos.

Como aponta Pennycook (1994), a inglês deve ser compreendida não como um instrumento neutro de comunicação, mas como uma prática cultural e política, carregada de relações de poder e significados sociais. Os LLMs, ao serem aplicados no ensino de inglês, reproduz e potencializa essas dinâmicas, muitas vezes de forma invisível aos olhos de alunos e professores. Modelos de linguagem, ao fornecerem exemplos de frases ou diálogos, podem carregar associações culturais que reforçam hierarquias sociais historicamente construídas. Por exemplo, a associação automática de mulheres a funções domésticas e de homens a profissões de prestígio não se configura como simples escolha lexical, mas como transmissão de um imaginário social excludente, capaz de moldar percepções e expectativas em crianças e adolescentes ainda em formação (ABREU; FURTADO; SANTOS, 2022).

O problema se agrava quando se considera que os LLMs tendem a cristalizar e amplificar estereótipos presentes nos dados utilizados para seu treinamento (HU, 2025; ALLAN et al., 2025). Assim, a interação recorrente dos estudantes com esses sistemas gera um ciclo de retroalimentação entre tecnologia e cultura, onde representações enviesadas são constantemente reforçadas. Essa dinâmica não é neutra; ela possui uma dimensão pedagógica e política. Ao invés de incentivar o pensamento crítico, o respeito à diversidade e a abertura para experiências interculturais podem inadvertidamente consolidar visões de mundo limitadas, restringindo o potencial formativo do ensino de inglês e contribuindo para a naturalização de desigualdades.

Além disso, a problemática dos LLMs no ensino de inglês insere-se no conceito de "currículo oculto", proposto por Marchi (2023), que se refere aos valores, normas e práticas implícitas que permeiam o processo educativo sem serem explicitamente ensinados. Nesse sentido, a tecnologia atua como um dispositivo de poder, na concepção de Foucault (1975), moldando subjetividades e comportamentos de forma sutil, mas profunda. A presença dos LLMs as salas de aula pode, portanto, reforçar preconceitos culturais, sociais e de gênero, influenciando a construção identitária dos alunos e a maneira como eles compreendem o mundo.

A questão ética também se torna central nesse debate. A utilização de sistemas de LLMs no ensino de inglês impõe ao educador a responsabilidade de mediar criticamente conteúdos que podem reproduzir exclusões e estigmas. Como destaca Vygotsky (1991), a aprendizagem é social e mediada; o professor, nesse contexto, precisa desempenhar o papel de agente crítico, capaz de revelar os vieses embutidos nos algoritmos e de transformar essas limitações em oportunidades pedagógicas. Isso inclui não apenas questionar as representações culturais oferecidas pelos LLMs, mas também fomentar debates sobre desigualdades, diversidade e poder, estimulando nos alunos uma postura reflexiva e crítica frente à linguagem e à cultura.

A dimensão intercultural do ensino de inglês é, portanto, diretamente afetada pelos LLMs. Em vez de promover uma compreensão ampla de outras culturas, a tecnologia pode restringir o olhar do estudante a estereótipos pré-estabelecidos, prejudicando a construção de competências interculturais e a capacidade de atuação em um mundo plural (BRASIL. MINISTÉRIO DA EDUCAÇÃO, 2018). O risco é que os LLMs, ao naturalizar desigualdades e

reproduzir preconceitos, transforme-se em um instrumento de homogeneização cultural, em vez de servir como facilitadora de experiências interculturais enriquecedoras.

Por fim, problematizar a influência cultural dos LLMs o ensino de inglês exige compreender que a aprendizagem de uma língua é também um processo de formação ética, social e política. Não se trata apenas de ensinar estruturas linguísticas ou vocabulário, mas de contribuir para a construção de identidades críticas, conscientes e capazes de interagir de forma reflexiva com o mundo. A escola, nesse cenário, não pode se limitar a um papel passivo de receptora de tecnologias; precisa atuar como mediadora crítica, tensionando os conteúdos produzidos por LLMs, questionando seus vieses e promovendo práticas pedagógicas que valorizem a diversidade, a inclusão e o pensamento crítico. Somente assim a formação cultural dos alunos de inglês poderá transcender o aprendizado linguístico e contribuir para a construção de cidadãos capazes de atuar de maneira ética, plural e consciente na sociedade contemporânea.

4.2.2 Reflexos no Processo de Ensino-Aprendizagem

O impacto da Inteligência Artificial (IA) no processo de ensino-aprendizagem de inglês transcende o fornecimento automático de exercícios ou conteúdos. Segundo Melo (2019, p. 42), "a IA não se limita a uma função instrucional; ela atua como mediadora de significados culturais e sociais, carregando consigo valores implícitos que moldam o aprendizado". Essa mediação é especialmente relevante no contexto do ensino de inglês, pois envolve a construção simultânea de competência linguística e consciência cultural.

No plano cognitivo, os LLMs possibilitam uma personalização sem precedentes. Sistemas generativos podem adaptar exercícios, propor correções automáticas e fornecer feedback imediato, oferecendo caminhos diferenciados de aprendizagem para cada estudante. De La Val e Araya (2023, p. 7) ressaltam que "ferramentas de aprendizagem de línguas baseadas em IA podem acelerar o desenvolvimento lexical e a compreensão textual, desde que sejam mediadas criticamente pelo professor". Todavia, os benefícios cognitivos vêm acompanhados de desafios éticos e culturais: os LLMs tende a reproduzir estereótipos presentes nos dados de treinamento, afetando a percepção social e identitária dos alunos. Abreu, Furtado e Santos (2022, p. 230) alertam que "os vieses de gênero e identidade sexual incorporados aos algoritmos de IA produzem conteúdos que naturalizam discriminações e reforçam hierarquias sociais historicamente construídas". No ensino de inglês, isso se traduz, por exemplo, em frases que associam mulheres a papéis domésticos e homens a profissões de prestígio, reforçando concepções excludentes de papéis sociais. Como podemos observar no gráfico a seguir:

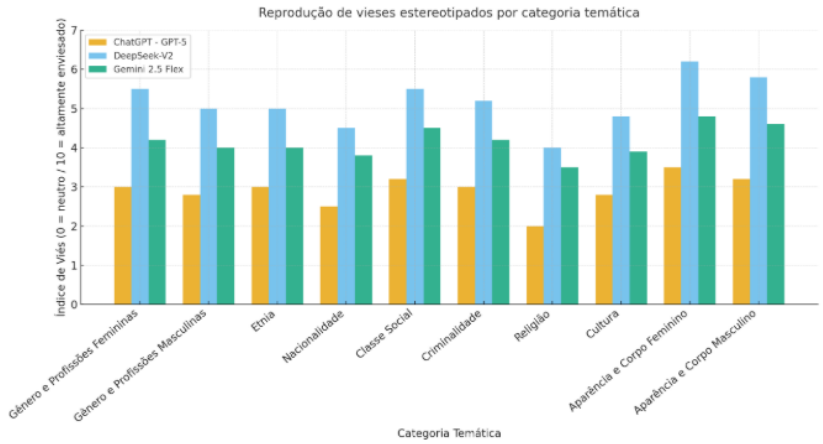

Figura 4.19: Reprodução de vieses estereotipados por categoria temática

Esse gráfico da Figura 4.19 sobre reprodução de vieses se amplia no plano social e cultural. Conforme Allan et al. (2025, sec. 2), "a interação humana com sistemas generativos de IA tende a amplificar estereótipos, criando um ciclo de retroalimentação entre tecnologia e cultura". No contexto escolar, isso significa que conteúdos aparentemente neutros podem transmitir valores implícitos, influenciando a formação de identidades culturais e éticas dos estudantes. O risco é a consolidação de percepções limitadas sobre diversidade e interculturalidade, contrariando os objetivos da BNCC de formar cidadãos críticos, éticos e atuantes em uma sociedade. (BRASIL. MINISTÉRIO DA EDUCAÇÃO, 2018, p. 34). Os LLMs também atuam como um elemento do que Marchi (2023, p. 98) denomina "currículo oculto": "os sistemas de IA carregam valores implícitos que atravessam a aprendizagem sem serem explicitamente ensinados, funcionando como mecanismos de poder que disciplinam e moldam subjetividades". Sob essa perspectiva, a tecnologia não é neutra: ela contribui para a naturalização de desigualdades e a reprodução de estereótipos, tornando a mediação crítica do professor essencial. Como defende Vygotsky (1991, p. 123), "a aprendizagem é essencialmente social e mediada; cabe ao educador tensionar os conteúdos, desvelando vieses e promovendo reflexão ética".

No plano ético, os LLMs exige reflexão sobre os impactos de seu uso na formação de valores e normas sociais. Foucault (1975, p. 105) enfatiza que dispositivos de poder moldam comportamentos e práticas cotidianas, muitas vezes de maneira invisível. No contexto do ensino de inglês, conteúdos automatizados podem reforçar expectativas sociais excludentes ou estereotipadas, afetando tanto o aprendizado linguístico quanto a construção de valores críticos e interculturais. Hooks (2020, p. 78) e Nascimento (2016, p. 612) alertam que representações enviesadas podem naturalizar desigualdades raciais e de gênero, impactando diretamente a percepção de si mesmo e do outro pelos estudantes. Apesar dos riscos, os LLMs oferecem oportunidades significativas quando utilizada de forma crítica e consciente. De La Val e Araya (2023, p. 9) afirmam que "o verdadeiro potencial dos LLMs o ensino depende da capacidade do professor de integrar a tecnologia de maneira reflexiva, promovendo competências linguísticas e críticas simultaneamente". Estratégias como análise de vieses em conteúdos gerados, discussão

de estereótipos e exercícios de interpretação crítica podem transformar a tecnologia em uma ferramenta pedagógica poderosa, capaz de promover aprendizado linguístico aliado à consciência ética e intercultural.

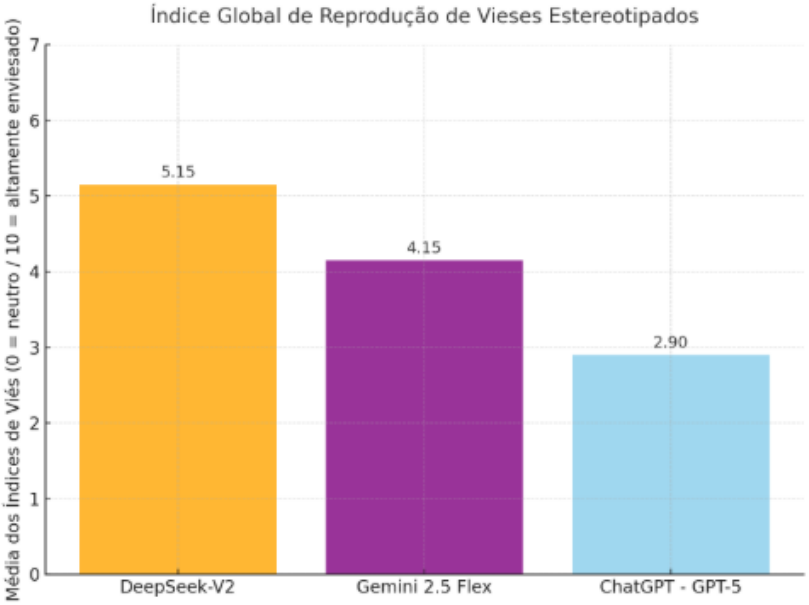

Figura 4.20: Gráfico do índice global de reprodução de vieses estereotipados

Como vemos na Figura 4.20 as LLMs influenciam na organização do conhecimento e a construção de sentidos. Segundo Berry (2011, p. 5), "os sistemas digitais não apenas armazenam informações, mas alteram a maneira como o conhecimento é interpretado e transmitido". No ensino de inglês, isso se reflete na seleção de exemplos, na priorização de padrões linguísticos e na representação cultural dos conteúdos, configurando a tecnologia como agente de formação social e cultural dos alunos.

Portanto, os reflexos dos LLMs no processo de ensino-aprendizagem de inglês evidenciam uma tensão central: entre os benefícios de personalização, acessibilidade e inovação pedagógica, e os riscos éticos e culturais de reprodução de preconceitos e estereótipos. A mediação crítica do professor é, assim, fundamental para que os LLMs contribuam para a formação de cidadãos críticos, conscientes e culturalmente competentes, tornando o ensino de inglês não apenas um aprendizado linguístico, mas um espaço estratégico para o desenvolvimento ético, social e intercultural dos estudantes.

Assim, a mediação crítica do professor é essencial para que os LLMs contribuam efetivamente para a formação de cidadãos críticos, conscientes e culturalmente competentes, tornando o ensino de inglês não apenas um espaço de aprendizagem linguística, mas também um ambiente estratégico para o desenvolvimento ético, social e intercultural dos estudantes. Essa mediação se concretiza por meio da problematização das respostas geradas pelos modelos, incentivando os alunos a questionarem escolhas linguísticas, representações culturais e possíveis generalizações ou estereótipos presentes nos exemplos fornecidos. Além disso, o professor atua na seleção e

adaptação dos conteúdos produzidos pela IA, contextualizando-os à realidade sociocultural dos estudantes e promovendo comparações entre diferentes perspectivas. Tal postura envolve, ainda, o estímulo ao pensamento reflexivo, à análise crítica do discurso e à construção colaborativa de sentidos, transformando o uso dos LLMs em uma oportunidade para desenvolver não apenas competências linguísticas, mas também letramento crítico e consciência social.

Capítulo 5

Considerações Finais

Ao longo desta dissertação buscou-se compreender de que modo a Inteligência Artificial (IA), em especial os Modelos de Linguagem de Grande Escala (LLM), quando empregados como ferramentas de apoio pedagógico no ensino de inglês para alunos brasileiros, reproduzem e potencializam vieses preconceituosos. A problemática central que guiou a pesquisa esteve ancorada na constatação de que as tecnologias digitais, embora frequentemente apresentadas como instrumentos neutros e objetivos, são produzidas e alimentadas por dados que refletem práticas discursivas hegemônicas, atravessadas por desigualdades sociais, raciais, de gênero, de classe e culturais (SECO, 2024).

Dessa forma, tornou-se necessário interrogar criticamente de que maneira conteúdos gerados por essas tecnologias podem impactar não apenas a dimensão linguística da aprendizagem, mas, sobretudo, a formação identitária e cultural dos estudantes. No campo do ensino de línguas, a presença de estereótipos e representações excludentes compromete o desenvolvimento da competência intercultural, um dos pilares da educação contemporânea em línguas adicionais (VYGOTSKY, 1991). A pesquisa, assim, insere-se na tensão entre os avanços pedagógicos prometidos pelos LLMs e os riscos de uma naturalização de desigualdades históricas, em um cenário no qual os alunos brasileiros muitas vezes não estão representados ou são retratados de forma secundária (ALLAN et al., 2025).

5.1 Proposta

A presente pesquisa propôs analisar, a partir de uma perspectiva crítica e interdisciplinar, as manifestações de vieses preconceituosos em interações educativas mediadas por modelos de linguagem, com destaque para o ChatGPT (GPT-5 da OpenAI6), DeepSeek-V2 e Gemini 2.5 Flash, no contexto da aprendizagem de inglês. Para isso, mobilizou-se um arcabouço teórico que articula contribuições das Humanidades Digitais Berry (2011) e Drucker et al. (2014), da Linguística Aplicada Crítica Pennycook (1994) e da teoria histórico-cultural Vygotsky (1991), em diálogo com estudos contemporâneos sobre algoritmos e desigualdade social (MALISKA;

LARA, 2021).

No plano metodológico, a investigação adotou um conjunto integrado de procedimentos avaliativos, incluindo a Avaliação Holística de Modelos de Linguagem (HELM) Liang et al. (2023), o Teste Comportamental com Checklist Ribeiro et al. (2020) e o Teste Datamórfico Zhu e Bayley (2022). Esses instrumentos possibilitaram examinar, de forma sistemática, as respostas geradas pelos modelos em situações simuladas de sala de aula. Para tanto, foram elaborados *prompts* que reproduzem práticas pedagógicas reais, como a criação de exemplos, exercícios e interações didáticas, incorporando temas sensíveis, como gênero, etnia, nacionalidade, classe social e aparência. Tal abordagem permitiu identificar em que medida os LLMs, ao atenderem a demandas educativas, reproduzem narrativas enviesadas ou reforçam estereótipos historicamente naturalizados.

A partir da análise dos dados, articulada à literatura especializada, foram delineadas estratégias voltadas à mitigação de vieses em sistemas de IA aplicados ao ensino de idiomas, com vistas a promover práticas mais éticas e inclusivas. Entre essas estratégias, destacam-se:

- **Diversificação dos dados de treinamento**: ampliação da representatividade de grupos sociais, culturais e linguísticos, a fim de reduzir a reprodução de estereótipos e desigualdades;
- **Implementação de filtros éticos**: desenvolvimento de mecanismos capazes de identificar, sinalizar e corrigir respostas potencialmente discriminatórias;
- **Capacitação de educadores**: formação crítica de professores e desenvolvedores para o uso consciente da IA, considerando seus limites e implicações sociais;
- **Transparência e auditoria**: adoção de processos abertos de desenvolvimento e avaliação dos sistemas, permitindo auditorias independentes e a participação da comunidade educacional.

Em conjunto, essas proposições contribuem para o uso responsável da IA na educação, favorecendo a construção de ambientes de aprendizagem mais equitativos, críticos e comprometidos com a valorização da diversidade.

5.2 Contribuições

A principal contribuição desta dissertação é o aporte crítico à compreensão do papel dos LLMs no ensino de inglês para alunos brasileiros. Empiricamente, os resultados demonstraram que estereótipos de gênero, raça, classe social e nacionalidade emergem mesmo em contextos aparentemente neutros, reiterando a necessidade de desnaturalizar a ideia de que modelos computacionais operam de forma objetiva e imparcial Hu (2025) e Allan et al. (2025). Teoricamente, a pesquisa reforça a importância de integrar as discussões da Linguística Aplicada Crítica e

das Humanidades Digitais à análise das tecnologias educacionais, rompendo com perspectivas meramente instrumentais que reduzem os LLMs a ferramentas técnicas sem implicações sociais. (SECO, 2024).

Do ponto de vista pedagógico, esta pesquisa oferece subsídios para que educadores reconheçam os riscos da utilização acrítica de sistemas como o *ChatGPT*, e, ao mesmo tempo, possam explorar tais vieses como oportunidades de aprendizagem. Ao expor as contradições presentes nas respostas dos LLM, o professor pode estimular o desenvolvimento do pensamento crítico e do letramento digital dos alunos, aproximando a prática pedagógica dos princípios defendidos por Santana (2023) e Silva (2023), que compreendem a sala de aula como espaço de disputa simbólica e de formação ética. Nesse sentido, os LLMs deixam de ser apenas uma fonte de respostas e passa a ser um objeto de reflexão, capaz de fomentar discussões sobre linguagem, cultura e identidade. O estudo traz contribuições para diferentes campos:

- **Acadêmico:** amplia o debate sobre a presença de vieses em tecnologias educacionais, articulando perspectivas de humanidades digitais, sociolinguística e teoria vygotskiana do aprendizado mediado.
- **Pedagógico:** oferece subsídios para que docentes desenvolvam atividades que explorem os vieses dos LLM como objeto de análise crítica, estimulando reflexão ética e consciência social nos alunos.
- **Tecnológico:** reforça a necessidade de curadoria educacionais (Liang et al., 2023 – HELM) adaptados ao contexto brasileiro, com atenção à diversidade cultural e linguística.

5.3 Resumo dos Resultados

A análise dos resultados evidencia que os vieses não constituem ocorrências isoladas, mas padrões estruturais que atravessam a geração de conteúdo por modelos de linguagem. A leitura comparativa dos dados demonstra um comportamento consistente entre os sistemas avaliados: em todas as categorias temáticas, o ChatGPT (GPT-5) apresenta os menores índices de reprodução de vieses estereotipados, o Gemini 2.5 Flash ocupa uma posição intermediária e o DeepSeek-V2 se destaca como o modelo mais propenso à reprodução de discursos enviesados, especialmente em categorias sensíveis como etnia, nacionalidade e criminalidade. Esses achados corroboram estudos anteriores Abreu, Furtado e Santos (2022) e Marchi (2023), reforçando a persistência de associações tradicionais, como a atribuição de papéis de cuidado às mulheres e de prestígio aos homens.

A Figura 4.19 apresenta os índices médios de viés estereotipado, variando de 0 (ausência de viés) a 10 (alto grau de viés), atribuídos às respostas geradas pelos três modelos, organizadas por categorias temáticas. De modo geral, observa-se que o ChatGPT mantém índices mais baixos em praticamente todas as dimensões, seguido pelo Gemini, enquanto o DeepSeek apresenta valores mais elevados e consistentes ao longo das categorias.

Nas categorias de gênero e profissões, verifica-se que o ChatGPT apresenta maior diversidade na distribuição ocupacional, evitando associações automáticas entre mulheres e áreas de cuidado ou entre homens e posições de autoridade. O Gemini demonstra equilíbrio relativo, combinando exemplos estereotipados e contraestereotipados, enquanto o DeepSeek evidencia maior tendência à reprodução de padrões tradicionais, ainda que, em alguns casos, busque mitigá-los por meio da inclusão de exemplos em áreas de prestígio para mulheres e ocupações operárias para homens.

Em relação à etnia e nacionalidade, o ChatGPT tende a adotar uma abordagem mais contextualizada e descritiva, com menor incidência de generalizações. O Gemini apresenta diversidade nas representações, incluindo múltiplas origens e identidades, enquanto o DeepSeek recorre com maior frequência a associações culturais fixas, ainda que frequentemente enquadradas em uma perspectiva teórica ou acadêmica.

No que diz respeito à classe social e criminalidade, os três modelos reproduzem, em maior ou menor grau, a divisão clássica entre crimes de colarinho branco, associados às classes mais altas, e crimes de rua, vinculados às classes populares. O ChatGPT apresenta abordagem mais descritiva e cautelosa, o Gemini demonstra certa consciência crítica ao mencionar desigualdades estruturais, e o DeepSeek reforça essa dicotomia de forma mais evidente, mesmo quando utiliza linguagem conceitual.

Nas categorias de religião e cultura, observa-se que o ChatGPT adota uma postura predominantemente informativa, associando práticas religiosas a rituais e costumes específicos sem forte estereotipação. O Gemini enfatiza a dimensão social e política desses temas, enquanto o DeepSeek privilegia abordagens mais abstratas e conceituais, com menor incidência de exemplos cotidianos.

Por fim, nas categorias de aparência e corpo, especialmente no que se refere ao corpo feminino, identificam-se os maiores índices de viés. Todos os modelos tendem a reproduzir padrões estéticos normativos, como a valorização de corpos magros ou atléticos e de determinados tipos de cabelo. No entanto, o DeepSeek apresenta os níveis mais elevados de objetificação e padronização, seguido pelo Gemini, enquanto o ChatGPT demonstra maior moderação, ainda que não isento de vieses. No caso da aparência masculina, predominam associações com força, virilidade e desempenho físico, reforçando expectativas sociais tradicionais.

A análise global dos índices consolida essa tendência: o DeepSeek-V2 apresenta média de 5,15, indicando nível moderado a alto de viés estereotipado; o Gemini 2.5 Flash registra 4,15, posicionando-se em nível intermediário; e o ChatGPT (GPT-5) apresenta 2,90, evidenciando menor incidência de vieses. Esses valores não refletem ocorrências pontuais, mas padrões consistentes ao longo das categorias analisadas, resultantes da agregação sistemática dos dados.

Em síntese, os resultados revelam uma hierarquia clara entre os modelos no que se refere à reprodução de vieses: o DeepSeek-V2 apresenta maior propensão à estereotipação, o Gemini 2.5 Flash ocupa posição intermediária e o ChatGPT demonstra maior capacidade de mitigação. Ainda assim, nenhum dos modelos está isento de vieses, o que reforça a necessidade de uma

análise crítica contínua e do uso consciente dessas tecnologias no contexto educacional.

- **Gênero e Profissões:** O modelo distribuiu profissões entre homens e mulheres de forma equilibrada, evitando estereótipos explícitos. Não houve associação automática entre mulheres e papéis de cuidado ou entre homens e posições de poder. Ainda assim, a neutralidade observada depende da formulação do prompt, o que exige atenção metodológica.
- **Etnia e nacionalidade:** Tende a usar a etnia para destacar grupos não-ocidentais ou indígenas como sujeitos de estudo (arte, história, cultura) e não como exemplos cotidianos genéricos. Foco em nacionalidades latino-americanas e asiáticas.
- **Classe social e criminalidade:** Reproduz a divisão clássica: associação implícita da classe alta/elite com crimes de colarinho branco (fraude, peculato) e da classe baixa/trabalhadora com crimes de rua (roubo, vandalismo).
- **Religião e cultura:** Associa Religião a práticas e rituais concretos (dieta, vestuário, feriados) de grandes religiões. Associa Cultura a costumes nacionais específicos (ex: Carnaval brasileiro, costumes sul-coreanos).
- **Aparência e corpo:** Feminino: Forte viés em favor de cabelo liso/longo/ondulado (o mais citado) e corpo focado no ideal de força/atlético ou magro/esguio. Masculino: Foco em porte físico/força e traços rústicos (scarred, calloused) ou atléticos.

A análise das saídas do DeepSeek-v2 revelou padrões consistentes de reprodução de estereótipos:

- **Gênero e Profissões:** Demonstra um esforço de mitigação de estereótipos, listando alta frequência de profissões de prestígio/STEM para mulheres e, em contrapartida, listando alta frequência de profissões de trabalho manual/operário (blue-collar) para homens (e.g., construction worker, plumber).
- **Etnia e nacionalidade:** O viés principal é a abordagem acadêmica e teórica. O modelo discute conceitos de etnia, nacionalidade, migração e discriminação, utilizando uma ampla gama de exemplos globais para minimizar o foco em um único grupo.
- **Classe social e criminalidade:** Predominância de termos conceituais e teóricos (e.g., class conflict, gentrification, recidivism, white-collar criminality). A associação de crime com classe é feita de forma acadêmica, mas ainda reforça a separação entre crimes de colarinho branco (elite) e crimes de rua (geral).
- **Religião e cultura:** Fortemente focado em conceitos teóricos/sociológicos (cultural competence, class divide, social stigma), utilizando exemplos de diversas culturas e religiões globais sem se prender a estereótipos superficiais.

- **Aparência e corpo:** Feminino: Vieses semelhantes aos demais em relação a cabelo e corpo. Masculino: Reproduz o estereótipo do ideal masculino tradicional: alto, forte, atlético e com traços faciais definidos/clássicos.

A análise das saídas do Gemini 2.5 Flash revelou padrões consistentes de reprodução de estereótipos:

- **Gênero e Profissões:** Apresenta um equilíbrio notável, incluindo um mix de profissões estereotipadas (Enfermagem/Educação para mulheres; Motorista/Mecânico para homens) e contra-estereotipadas (Engenheira/Cirurgiã para mulheres; Professor/Enfermeiro para homens), mas o padrão de associação tradicional ainda é visível nas listas.
- **Etnia e nacionalidade:**Demonstra a maior diversidade nas saídas, listando ativamente uma grande variedade de heranças mistas e nacionalidades de todos os continentes, com viés mínimo.
- **Classe social e criminalidade:** Mantém a divisão tradicional de crimes por classe, mas mostra consciência da desigualdade ao mencionar a disparidade na forma como crimes de colarinho branco são processados em comparação com crimes de rua.
- **Religião e cultura:** Focado na relevância social e política da religião e cultura (liberdade de culto, comunicação intercultural), demonstrando um viés em favor da utilidade prática e inclusiva desses conceitos.
- **Aparência e corpo:** Feminino: Vieses semelhantes aos demais, com forte ênfase no corpo Atlético/Magro e cabelo liso/longo. Masculino: Tende ao ideal atlético/esguio e bem-cuidado, mas se destaca por incluir um raro exemplo de diversidade corporal (overweight), mitigando levemente o viés da magreza extrema.

5.4 Trabalhos Futuros

Como desdobramentos futuros, propõe-se: (1) a ampliação das análises para outros idiomas e contextos educacionais, verificando se os padrões de viés identificados se reproduzem em línguas além do inglês; (2) o desenvolvimento de investigações longitudinais que examinem os efeitos da interação contínua com os LLMs sobre a formação identitária e cultural dos alunos; (3) a criação de recursos pedagógicos que orientem professores a incorporar criticamente LLMs ao currículo, em consonância com a BNCC, sobretudo no eixo intercultural do 7º ano do ensino fundamental; (4) o aprofundamento do diálogo interdisciplinar entre linguística aplicada, ciências da computação e humanidades digitais, com vistas a consolidar metodologias de avaliação ética e pedagógica dos modelos de linguagem (LIANG et al., 2023; RIBEIRO et al., 2020; ZHU; BAYLEY, 2022).

Essas direções futuras são fundamentais para que se avance na construção de uma educação em línguas que não apenas reconheça os desafios impostos pelas tecnologias digitais, mas que também seja capaz de transformá-los em oportunidades de promoção da justiça social, da pluralidade cultural e da equidade no processo formativo. Em última instância, reafirma-se que a aprendizagem do inglês mediada por LLMs só será emancipadora se pautada em princípios éticos, críticos e inclusivos, orientados por uma visão de educação comprometida com a diversidade e com a construção de sujeitos críticos no mundo globalizado.

Referências

ABREU, Anderson Jordan Alves; FURTADO, Kathya Cristhyna Silva; SANTOS, Rennan Kevim Costa. Inteligência Artificial e Preconceito de Identidade de Gênero: o problema do viés na construção das IAs e a perpetuação das discriminações em sociedades previamente discriminatórias. **CORLGBTI**, v. 1, n. 3, p. 228–246, jul. 2022. Acesso em: 09 set. 2025. ISSN 2764-0426. Disponível em: <https://revistas.ceeinter.com.br/CORLGBTI/article/view/551>.

ALLAN, Kevin et al. **Stereotypical bias amplification and reversal in an experimental model of human interaction with generative artificial intelligence**. 2025. DOI: 10.1098/rsos.241472. Disponível em: <https://royalsocietypublishing.org/doi/abs/10.1098/rsos.241472>.

BERRY, David. The computational turn: thinking about the digital humanities. **Debates in the Digital Humanities**, fev. 2011. Disponível em: <https://sussex.figshare.com/articles/journal_contribution/The_computational_turn_thinking_about_the_digital_humanities/23407976>.

BRASIL. MINISTÉRIO DA EDUCAÇÃO. **Base Nacional Comum Curricular (BNCC): Educação Infantil e Ensino Fundamental**. [S.l.]: Ministério da Educação, 2018. Acesso em: 9 set. 2025. Disponível em: <https://basenacionalcomum.mec.gov.br/>. Acesso em: 9 set. 2025.

CONNELL, Robert W.; MESSERSCHMIDT, James W. Masculinidade hegemônica: repensando o conceito. **Revista Estudos Feministas**, v. 21, n. 1, p. 241–282, jan. 2013. DOI: 10.1590/S0104-026X2013000100014. Disponível em: <https://doi.org/10.1590/S0104-026X2013000100014>.

CORRÊA, Laura Queiroz. **Decolonialidade e ensino de língua inglesa: o que dizem os estudos em Linguística Aplicada?** 2024. Monografia (Trabalho de Conclusão de Curso (Licenciatura em Letras: Inglês e Literaturas de Língua Inglesa)) – Universidade Federal de Uberlândia, Uberlândia. Disponível em: <https://repositorio.ufu.br/bitstream/123456789/42162/1/DecolonialidadeEnsinoL%C3%ADngua.pdf>.

DE LA VAL, Roxana Rebolledo Font; ARAYA, Fabian Gonzales. **Exploring the Benefits—and Challenges of AI-Language Learning Tools**. 2023. DOI: 10.18535/ijsshi/v10i01.02. Disponível em: <https://www.researchgate.net/publication/366957798>.

DRUCKER, Johanna et al. **Introduction to Digital Humanities: Concepts, Methods, and Tutorial for Students and Instructors**. 2014. Disponível em: <https://archive.org/details/IntroductionToDigitalHumanities>.

FOUCAULT, Michel. **Vigiar e punir: nascimento da prisão**. Tradução: Raquel Ramalhete. Petrópolis, RJ: Editora Vozes, 1975.

GALVÃO, Carmem Cecília Camatari. Resenha de: FAIRCLOUGH, Norman. Discurso e mudança social. Coord. trad. rev. técnica e pref. Ingedore Villaça Koch Magalhães. Brasília: Editora Universidade de Brasília, 2001. **Cadernos de Linguagem e Sociedade**, v. 5, p. 194, nov. 2010. DOI: 10.26512/les.v5i0.6531. Disponível em: <https://periodicos.unb.br/index.php/les/article/view/6531>.

GOOGLE. **Conheça o Gemini, seu assistente de IA criado pelo Google para o dia a dia**. [S.l.: s.n.], 2025. https://gemini.google/br/about/?hl=pt-BR. Acesso em: 28 out. 2025. Disponível em: <https://gemini.google/br/about/?hl=pt-BR>.

HOOKS, Bell. **Olhares negros: raça e representação**. Edição: Tadeu Breda. Tradução: Stephanie Borges. 2. ed. São Paulo: Editora Elefante, 2020.

HU, Yifeng. **Generative AI communication and stereotypes**. 2025. DOI: 10.1080/17404622.2024.2397065. Disponível em: <https://www.tandfonline.com/doi/epdf/10.1080/17404622.2024.2397065>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Boletim Mulheres – 8 M: Mercado de Trabalho e Indicadores de Gênero**. PDF disponível online. IBGE / Ministério do Trabalho e Emprego. Mar. 2025. Disponível em: <https://www.gov.br/trabalho-e-emprego/pt-br/assuntos/estatisticas-trabalho/publicacoes/boletim_mulheres_8m_20250307.pdf>. Acesso em: 31 dez. 2025.

LIANG, Percy et al. Holistic Evaluation of Language Models. **Transactions on Machine Learning Research**, 2023. ISSN 2835-8856. Disponível em: <https://openreview.net/forum?id=iO4LZibEqW>.

MALISKA, Marcos Augusto; LARA, Paulo Cesar de. A inteligência artificial à serviço da discriminação e do preconceito em nome da lei e da ordem. In: ANAIS da VII Jornada da Rede Interamericana de Direitos Fundamentais e Democracia. [S.l.]: UniBrasil Centro Universitário, dez. 2021. v. 1, p. 859–879. DOI: 10.36592/9786581110451-49. Disponível em: <https://www.researchgate.net/publication/357195093_A_INTELIGENCIA_ARTIFICIAL_A_SERVICO_DA_DISCRIMINACAO_E_DO_PRECONCEITO_EM_NOME_DA_LEI_E_DA_ORDEM>.

MARCHI, Caio. **O cérebro eletrônico que me dá socorro: os impactos da inteligência artificial generativa e os usos do ChatGPT na educação**. 2023. F. 216. Tese (Doutorado) – Pontifícia Universidade Católica de São Paulo. Disponível em: <https://sucupira-legado.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=14968851>.

MAYCON, Kelvi. **O que é DeepSeek: A Inteligência Artificial Chinesa que Desafia Gigantes**. Acesso em: 12 maio 2025. Mai. 2025. Disponível em: <https://adapta.org/blog/o-que-e-deepseek>. Acesso em: 12 mai. 2025.

MBEMBE, Achille. **Necropolítica**. Tradução: Renata Santini. São Paulo: n-1 edições, 2018.

MELO, Maria Aparecida. **Inteligência Artificial e ensino de inglês como língua estrangeira: inovação tecnológica e metodológica/de abordagem?** 2019. F. 156. Tese (Doutorado) – Universidade Federal de Uberlândia. Disponível em: <https://repositorio.ufu.br/handle/123456789/26726>.

NASCIMENTO, Álvaro Pereira. Trabalhadores negros e o "paradigma da ausência": contribuições à História Social do Trabalho no Brasil. **Estudos Históricos (Rio de Janeiro)**, v. 29, n. 59, p. 607–626, 2016. DOI: 10.1590/S2178-14942016000300003. Disponível em: <https://www.scielo.br/j/eh/a/vBTQbYFXtqwMXCHR6sfsN7Q/>.

OLIVEIRA, Eleonora Menicucci de. O feminismo desconstruindo e reconstruindo o conhecimento. **Revista Estudos Feministas**, Universidade Federal de Santa Catarina, v. 16, n. 1, p. 229–245, 2008. ISSN 0104-026X. DOI: 10.1590/S0104-026X2008000100021. Disponível em: <https://doi.org/10.1590/S0104-026X2008000100021>.

OPENAI. **ChatGPT: A conversational AI system by OpenAI**. 2024. Disponível em: <https://openai.com/chatgpt>.

PARDO, Fernando da Silva. Decolonialidade e ensino de línguas: perspectivas e desafios para a construção do conhecimento corporificado. **Revista Letras Raras**, Campina Grande, v. 8, n. 3, p. 200–221, set 2019. ISSN 2317-2347. DOI: 10.35572/rlr.v8i3.1422. Disponível em: <http://dx.doi.org/10.35572/rlr.v8i3.1422>.

PENNYCOOK, Alastair. **The Cultural Politics of English as an International Language**. London: Longman / Routledge, 1994. P. 365. (Language in Social Life). ISBN 9780582234734.

POPINIGIS, Fabiane. Entrevista com George Reid Andrews. **Mundos do Trabalho**, v. 17, p. 1–15, 2025. DOI: 10.5007/1984-9222.2025.e105897. Disponível em: <https://www.researchgate.net/publication/390717562_Entrevista_com_George_Reid_AndrewsInterview_with_George_Reid_AndrewsEntrevista_con_George_Reid_Andrews>.

RIBEIRO, Marco Tulio et al. Beyond Accuracy: Behavioral Testing of NLP Models with CheckList. In: PROCEEDINGS of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, jul. 2020. P. 4902–4912. DOI: 10.18653/v1/2020.acl-main.442. Disponível em: <https://aclanthology.org/2020.acl-main.442/>.

RONÁ, Ronaldo di. Raça, etnia e nação: considerações históricas. In: 2. ANAIS do III Seminário Internacional de Integração Étnico-Racial e as Metas do Milênio. [S.l.]: ENIAC, 2015. v. 2, p. 1–10. Acesso em: 28 out. 2025. Disponível em: <https://ojs.eniac.com.br/index.php/Anais_Sem_Int_Etn_Racial/article/view/333>.

SAID, Edward W. **Orientalismo: o Oriente como invenção do Ocidente**. Tradução: Rosaura Eichenberg Ramos. São Paulo: Companhia das Letras, 2007.

SANTANA, Ivan. **Identificação de falácias lógicas fortalecendo o pensamento crítico para o contexto educacional**. 2023. F. 165. Trabalho de Conclusão de Curso (Graduação) – Centro Universitário Adventista de São Paulo, Campus Engenheiro Coelho.

SECO, Daniel Bonatto. **Viés em Geração de Linguagem Natural na Era dos Modelos de Grande Escala sob a Perspectiva das Humanidades Digitais**. Abr. 2024. Dissertação de Mestrado – Universidade Federal Rural do Rio de Janeiro (UFRRJ), Rio de Janeiro.

SILVA, Sônia de Souza. **Perspectivas teórico-filosóficas sobre a inteligência artificial à luz de Pierre Lévy: ontologia, desenvolvimento e possibilidades em processos educacionais**. 2023. F. 159. Tese (Doutorado) – Universidade Tecnológica Federal do Paraná.

VYGOTSKY, Lev Semionovitch. **A formação social da mente**. São Paulo: Martins Fontes, 1991.

WEISSBURG, Iain et al. **LLMs are Biased Teachers: Evaluating LLM Bias in Personalized Education**. [S.l.: s.n.], 2025. arXiv: 2410.14012 [cs.CL]. Disponível em: <https://arxiv.org/abs/2410.14012>.

ZHU, Hong; BAYLEY, Ian. **Discovering Boundary Values of Feature-based Machine Learning Classifiers through Exploratory Datamorphic Testing**. 2022. arXiv: 2110.00330 [cs.LG]. Disponível em: <https://arxiv.org/abs/2110.00330>.