

RESEARCH

Open Access



Large language models, social demography, and hegemony: comparing authorship in human and synthetic text

A. J. Alvero^{1*}, Jinsook Lee², Alejandra Regla-Vargas³, René F. Kizilcec², Thorsten Joachims^{2,4} and Anthony Lising Antonio^{5*}

*Correspondence:
ajalvero@cornell.edu;
aantonio@stanford.edu

¹ Center for Data Science for Enterprise and Society, Cornell University, Ithaca, USA

² Department of Information Science, Cornell University, Ithaca, USA

³ Department of Sociology, University of Pennsylvania, Philadelphia, USA

⁴ Department of Computer Science, Cornell University, Ithaca, USA

⁵ Graduate School of Education, Stanford University, Stanford, USA

Abstract

Large language models have become popular over a short period of time because they can generate text that resembles human writing across various domains and tasks. The popularity and breadth of use also put this technology in the position to fundamentally reshape how written language is perceived and evaluated. It is also the case that spoken language has long played a role in maintaining power and hegemony in society, especially through ideas of social identity and “correct” forms of language. But as human communication becomes even more reliant on text and writing, it is important to understand how these processes might shift and who is more likely to see their writing styles reflected back at them through modern AI. We therefore ask the following question: *who* does generative AI write like? To answer this, we compare writing style features in over 150,000 college admissions essays submitted to a large public university system and an engineering program at an elite private university with a corpus of over 25,000 essays generated with GPT-3.5 and GPT-4 to the same writing prompts. We find that human-authored essays exhibit more variability across various individual writing style features (e.g., verb usage) than AI-generated essays. Overall, we find that the AI-generated essays are most similar to essays authored by students who are males with higher levels of social privilege. These findings demonstrate critical misalignments between human and AI authorship characteristics, which may affect the evaluation of writing and calls for research on control strategies to improve alignment.

Keywords: Authorship characteristics, Large language models, Computational social science, Linguistic sociology, AI homogenization

Introduction

In the 2018 science fiction film *Sorry to Bother You*, a Black telemarketer in Oakland, California faces a dilemma. When people answer his calls and hear his African American Vernacular English (AAVE) inflected voice, they immediately hang up and ignore his sales pitch. With rising bills, debt, and desperation, he follows the advice of an older Black colleague to use a “White voice” (a US English dialect typically associated with upper-middle-class White Americans). When using this dubbed “White voice”

(exaggerated for comedic effect), people no longer hang up, and in fact, they become high-paying customers. The film uses this play on spoken language, social identity, and power to highlight the concept of hegemony, defined by Gramsci as the sociocultural norms that uphold structures of power and domination [36]. Linguistic hegemony, the focus of this particular scene in the film, operates in similar ways through social enforcement of particular ways to speak, write, and communicate.

Though the film is fictional, current technologies can manipulate voices to sound like specific individuals or reduce accents by replacing them with more “normalized” speech [67]. How these dynamics compare with written language is less known. Biases in natural language processing (NLP) techniques are well documented [31, 46], but applications of sociolinguistic perspectives on how large language model (LLM) communication styles track with specific social demographics could be instructive in determining whether or not the issues presented in the movie with spoken language could emerge with written language [2]. Educational systems in the US have a long demonstrated preference for the writing and speaking styles of those from higher socioeconomic backgrounds [15]; even if the preference is not explicitly stated, studies have found that writing patterns stratified by social class are highly predictive of standardized test scores and final college GPA [5, 69]. This raises questions about how modern society’s shift toward increased usage of text-based communication may affect these identity dynamics, something which remains underexplored. With widely and globally popular generative AI technology like ChatGPT that can write human-like text, examining what linguistic styles they adopt could reveal much about underlying biases in AI systems and the contexts in which they emerge.

The popularity of LLMs and the platforms they power (like ChatGPT, arguably the current most popular LLM application) is largely due to their ability to “write like real people” so convincingly that some have argued that the traditional Turing test has been inverted [76] (ie. LLMs test the intelligence of humans rather than humans test the intelligence of LLMs). However, the self-evident potential of LLMs has raised critical questions about their biases, tendency to emulate certain political and moral stances, and ability to fabricate references when used as a research assistant [1, 73]. These studies examine specific types of responses to structured sets of questions, whereas linguistic hegemony (as outlined by Gramsci and others) operates on more fundamental levels, such as word choice reflecting a universally understood “common sense” that does not consider sociolinguistic variation as a naturally occurring social phenomenon but rather something to “fix” [36]. Deviating from these linguistic norms (or at least being perceived as linguistically deviant) can put people at odds with the social order and subject them to hegemonic forces and pressures purely through their linguistic styles, tendencies, and preferences. Given the role of language in upholding social hegemony, it is vital to examine the linguistic styles and identities that LLMs adopt in the language they generate.

Most text on the internet was written by people (at least for now, eg. Bohacek and Farid [17]), so LLMs implicitly learn correlations between demographics, contexts, and communication styles from training data. If LLMs tend to write more like those dominant in the training data (ie. particular social strata) or if they are explicitly instructed by their designers to write in a particular way, this could perpetuate cultural and linguistic

hegemony by homogenizing expression [16, 50]. Some studies have found evidence of these trends with specific types of responses to survey questions [9, 43, 80, 89] and in terms of the limited range of content they produce for a given prompt or task [39, 66]. Investigating similarities and differences between human and AI writing across social groups can reveal biases in the demography of LLMs' training data, which has implications for widespread use of AI tools like ChatGPT in numerous contexts. It may also reveal sociolinguistically grounded forms of cultural hegemony if certain groups are over-represented in influencing LLMs' writing styles [85]. If LLMs adopt the writing patterns of privileged groups, this could shrink sociolinguistic variation or deflate it artificially in settings where AI is in active use, like educational contexts. New social divisions may emerge between those who write like AI versus those who do not, not unlike well documented sociolinguistic divisions that operate under similar parameters. However, these concerns are presently not grounded in scientific evidence of sociodemographic patterns in written language. Understanding these dynamics could point to deep social dimensions in broader issues like AI alignment (i.e., ensuring that AI systems behave in line with human values, norms, and intentions). The homogenization of writing and communication through LLMs' styles could have major implications for culture and communication. Further, linguistic hegemony may evolve through the popularization of LLMs based on correlations of whose writing is most reflective of the text generated by LLMs [84]. As LLM usage increases, comparative analyses between human and AI writing will only become more challenging as people use it more in their daily lives.

We start to answer these questions by examining a key social process: college admissions. College admissions serve as an insightful context since identity is salient in shaping how applicants present themselves, particularly in selective admissions where holistic review of transcripts, essays, and other applicant information is the norm [11, 79]. The personal information and written statements provided by applicants are highly correlated with their social identity and background context. For example, past studies find strong predictive relationships between essay content and variables like family income and standardized test scores [5]. Essays also show potential for use in replicating past admissions decisions, given trends toward test-optional and test-blind policies [49], and to infer personal qualities of the applicants that predict academic success [55]. In this analysis, we focus on stylistic features of writing as captured by a popular method to analyze text: Linguistic Inquiry and Word Count (LIWC). LIWC is a dictionary built on the basis of psycholinguistic research on the relationship between written language and psychological characteristics of the author [70]. Past studies have also found that LIWC features are strong predictors of many different dimensions of the authors, such as a college applicant's eventual grades after enrolling, SAT score, and household income [5, 69]. The high interpretability of the features is also useful: the metrics are based on relative and absolute frequencies of specific words and punctuation.

To compare the writing style of human-authored and AI-generated texts, we analyze a dataset of application essays submitted to the University of California system and to an engineering school at a private, selective university in the US. We generate AI-written essays using two popular LLMs responding to the same prompts from these two application contexts. We find that AI and human written essays have distinct distributions in how frequently they use words associated with individual LIWC features, such as verbs

and analytical thinking. Moreover, AI-generated essays had notably lower variation in individual LIWC features than the human-written essays, suggesting a more narrow linguistic style. We then compared these distributions split by applicant characteristics related to their identity and neighborhood context. We found the linguistic style of AI-generated essays to be closer to some identity groups, such as applicants with college-educated parents and those living in ZIP Codes with higher socioeconomic measures, than other students. Gender was prominent in two ways. First, among the style features where male and female applicants diverged the most, the synthetic text was more aligned with male usage. However, the opposite (female style features more similar to the AI style features) was more likely to be true when focusing on public school applicants or analyzing multiple features simultaneously.

Finally, we compared the essays using all of the LIWC features and found that these differences for individual features show strong social patterning when more information is provided to the model. In these analyses, we find that the AI-generated essays are most similar to students from areas with higher social capital compared to the average applicant. In situations where people use LLMs, the likelihood that users see writing styles similar to their own is somewhat dependent on their demographic information. For those with higher levels of social capital, they are likely to be presented with text they could have plausibly written. Regardless of whether individuals are aware of these social dynamics, the broad uptake of LLMs along with these patterns have the potential to homogenize written language to make people sound less like themselves or others in their communities (outside the upper middle class). Focusing on college admissions as a context for studying LLMs in this way provides unique and important perspectives. There is no clear standard for what constitutes a "good" or "bad" personal statement, and students and families spend much time and energy seeking support through various means (e.g., online forums, private admissions counselors). Within the college admissions ecosystem, modern AI systems may be viewed as oracles, providing seemingly authoritative guidance for applicants as they navigate opaque admissions processes. However, as colleges and universities continue their efforts to diversify their student bodies, LLMs might inadvertently shrink the pool of diverse experiences students have described in the past and make it more difficult to distinguish applicants or groups of applicants.

Related work

We organize our review of related work into two parts: first, we review the characteristics and biases of LLMs. Then, we discuss their diverse applications in both scholarly and general settings. This review establishes a foundation for our investigation, which aims to understand the sociolinguistic profiles of LLMs.

Profiles of LLMs

Recent advances in transformer-based architectures have enabled the development of LLMs that can generate impressively human-like text and respond to complex questions and prompts [19]. Models like OpenAI's GPT-3.5 and GPT-4 are trained on massive textual corpora extracted from the internet, allowing them to learn the statistical patterns of language and generate coherent new text given a prompt. Despite the notable advancements of LLMs compared to other NLP techniques, they demonstrate biases

that are reminiscent of past NLP approaches along with new forms of bias. The primary difference between bias in, say, word embeddings and LLMs is that people are able to directly interact with LLMs through platforms like ChatGPT. A better understanding of LLMs will therefore require consideration of the bias literature alongside studies of sociolinguistic profiles of LLMs.

Extensive research has uncovered notable biases within LLMs [13, 25, 32, 47, 50, 63, 65, 80]. These biases emerge because LLMs are trained on extensive datasets collected from the internet, which generally mirror prevalent societal biases related to race, gender, and various attributes [25, 26, 47, 65]. For example, Omiye and colleagues [65] found that four LLMs (e.g., ChatGPT, GPT-4, Bard, & Claude) contained medical inaccuracies pertaining to different racial groups such as the belief that racial differences exist in pain tolerance and skin thickness. LLMs have been observed to reinforce gender stereotypes by associating women with professions like fashion models and secretaries, while assigning occupations like executives and research scientists to males [47]. Beyond the obvious risks related to misinformation and the perpetuation of socially harmful biases, LLMs are also widely marketed as being able to “adapt” to users given their input [44], making it possible for these same biases to be continually reinforced while also being difficult to properly audit [62]. For the millions of lay users of LLMs, receiving these kinds of messages repeatedly could reinforce linguistic hegemony by pointing to a more narrow set of possible outcomes in their use under the assumption that the models are learning to adapt to the user. Despite the clear advancements in sophistication and performance, LLMs still largely retain these well-documented forms of bias.

Apart from sociodemographic biases, recent research has noted the distinct characteristics or “profiles” of LLMs. Specifically, it has been observed that these models exhibit a left-leaning bias in their responses [57, 61] and an affinity for Western cultural values [80]. These models have also been found to consistently imitate personality traits such as openness, conscientiousness, and agreeableness in their generated outputs [68]. These and many other studies have focused on English language responses, but LLMs are able to generate non-English text as well [86] making it possible to test these results from many different sociocultural paradigms. LLMs are clearly capable of generating many different types and forms of language, but little work has taken a sociolinguistic perspective where the social characteristics of writers—human vs. AI—are the focus of textual comparisons. It is also the case that large platforms like ChatGPT are subject to internal tweaks and modifications that users may not be entirely aware of, though so far, this has only been observed with respect to specific types of question answering [20] rather than more holistic shifts to the ways that LLMs tend to use language and which people use language most similarly. Whether or not people want an LLM that can communicate like someone with the same social identities (e.g., African-American Vernacular English and Black Americans) is also an important question [56], but the capacity for LLMs to mimic different writing styles itself is underexplored. While some research is underway on alignment along macro perspectives [45], linguistic hegemony typically operates by presenting all speakers with an assumed standard way of using language. To take an example from the film mentioned in the introduction, the English associated with middle-to-upper-middle-class White people in the US is often considered the national standard

as opposed to the AAVE used by the protagonist. If AI alignment does not consider non-predominant linguistic styles and communities, it could further reinforce extant linguistic hegemony. Our study therefore complements ongoing research about the limitations of using LLMs to simulate human respondents due to their inabilities in portraying particular identity groups [83]. Generally, identifying sociolinguistic trends between humans and out-of-the-box LLMs would generate valuable insights to a broad range of computational social scientists, especially when considering the popularity of these tools.

Applications of LLMs

Scholarly community

LLMs are being rapidly adopted for a wide range of uses among scholars, including document annotation [28, 82], information extraction [21], and tool development with applications to healthcare, education, and finance [48, 58, 71, 87]. Social scientists have also been cautiously optimistic about potential uses for LLMs in research [23, 42]. The breadth of these use cases make it imperative to best understand their stylistic tendencies for language with respect to various social and scientific contexts.

For example, scholars are leveraging LLMs in text annotation tasks, including identifying the political party of Twitter users [82], automating document classification [28], and extracting counterfactuals from premise-hypothesis sentence pairs [21]. Concurrently, others are employing LLMs to create tools that are capable of generating personalized feedback from open-ended responses for students [58] and even collaborate with writers to co-author screenplays and scripts [60]. Aside from these academic uses, everyday people are more likely to engage in the text-generation dimension of LLMs. But this is especially the case for ChatGPT given its popularity among the world's population, due partly to its ease of use relative to other LLM tools and technologies early on [74]. Focusing on patterns in the types of text they generate allows us to imagine how people across social strata experience the tool.

General population

Existing research indicates that LLMs are being used among the general population for writing-related tasks [14, 60, 78]. For example, Wordcraft [88] employs a LLM [81] to assist writers with tasks such as proofreading, scene interpolation, and content generation. When writers were asked about their experience using Wordcraft [88], they indicated it had reduced their writing time and eased the writing experience. Similarly, Dramatron [60], an LLM-powered co-writing tool, helps industry professionals develop their scripts by generating content and offering suggestions for improving their work. These capacities also threaten creative labor, as seen with the writers' strike in the film and television industry. Recent findings from a Pew Research report highlight a notable trend in the use of ChatGPT among U.S. teens. According to the report, about 13% of all U.S. teens have used ChatGPT for assistance with their schoolwork, a number that is likely to increase over time.

Despite considerable research into the technical and ethical dimensions of LLMs, there remains a significant gap in understanding their linguistic profiles. For example, when students are using LLMs to help them with their homework and writing assignments, there is limited information about the ways that LLMs are more strongly correlated with writing styles and tendencies associated with particular groups of people. Conversely, students who do not speak English as a first language might find that their writing is more often labeled as “AI-generated” [52], pointing to additional ways that the current AI ecosystem could reinforce linguistic hegemony (in this case, English language hegemony). It is therefore crucial to examine their linguistic biases and tendencies as they may lead to the marginalization of non-standard dialects and expressions. Studies have mapped out such patterns in the context of academic writing and review [54], but here we consider a context that is more common and familiar to people living in the US: writing a personal statement for college admissions. Understanding LLMs’ linguistic tendencies is necessary to ensure that they do not perpetuate cultural hegemony [36, 85], potentially reinforcing biases against diverse linguistic practices. We address this directly by exploring the question: Who do these models mimic in their writing? This investigation is crucial, as the styles emulated by LLMs may influence the landscape of digital communication and by extension the outcomes of textual tasks (i.e. annotation, information extraction).

Research questions

We contribute to the literature on the social dimensions and implications of LLMs through a comparative analysis of human-authored and AI-generated text. Specifically, we organize our work around the following research questions:

1. How does the writing style of AI compare with the writing style of humans? Are there specific groups of people whose writing style is most closely imitated by AI?
2. What are the social characteristics of the humans whose essay is the most similar to a given AI generated essay?
3. What is the predicted social context of AI as an author?

Answering question one will generate insights into the ways that AI writing represents or not the type of variation seen in human writing as well as providing one perspective to the overarching question of “who does AI write like”. The second question takes this one step further by pairing human- and AI-written essays based on similarity to see which students are most likely to deploy writing styles similar to AI. Finally, we leverage past results showing the strong relationships between student essay writing and geographically distributed forms of social capital and mobility [6] to locate which communities are producing text most closely imitated by AI. Recent studies indicate that AI has more negative tendencies toward dialectal forms of language [38], suggesting that AI is likely to use higher and more formal registers used by people and communities with higher socioeconomic status. We hope these findings could spur more specified hypothesization that consider other social contexts as well as experimental and causal frameworks [30, 37]. For example, depending on the nature of the social alignment between human and AI writing style, a follow-up study could examine the effects of writing more or less like an AI on human evaluation (in our context:, evaluation of a college application).

Our analyses and questions are distinctive in another way. Many, if not most, of the studies we cite here examine LLMs in more controlled settings rather than connecting their behaviors to real world contexts and situations, something we explicitly try to address in our work. In the US, millions of people have gone through the ritual of crafting a personal statement describing themselves, their interests, and their goals as part of their college applications. Many more people have written similar types of documents, such as cover letters for job applications. We connect LLM behavior to human behavior¹ to posit how extremely popular models might communicate to people “right” and “wrong” ways to describe themselves when comparing authorship demographics with AI writing style tendencies.

Data and context

We analyze data from two higher education contexts. The first is the University of California (UC) system, one of the largest public research university systems in the world. The application process is streamlined: students who want to apply to any of the 9 campuses that provide undergraduate degree programs only need to complete one actual application. That one application is then submitted to the campuses that the students would like to attend, including highly selective, elite campuses (Berkeley, Los Angeles) along with less selective campuses (Riverside, Merced). Students are therefore unable to tailor their applications to specific campuses. Applicants have to write essays written to four of eight possible prompts (70 possible combinations). The essay prompts are similar to those used by the Common App, an independent, national college application platform (see Table A1 for a full list of the prompts). Our UC data come from every Latinx identifying in-state applicant who applied during the 2016–2017 academic year (well before ChatGPT was released). California has a long history of linguistic prejudice and bias against Latinx people and communities [10, 64], making these students particularly vulnerable to linguistic hegemony. As a population that is still under-represented in US higher education, especially so for the Mexican-American students who comprise the largest Latinx population in California and the rest of the country, they also represent a socially distinct group of students to the more elite students in our second education context.

The second context is a large private research university in the northeastern United States. Our data include undergraduate admissions essays that were submitted to the school of engineering during the 2022–2023 admissions cycle via the Common App (essays were submitted by November 1, 2022, right before the release of ChatGPT). Applicants wrote three essays: one in response to one of seven Common App prompts and two unique to the school of engineering. We analyze the Common App essays. This was done to make the cross prompt analyses as comparable as possible given the similarity between the Common App prompts and the public prompts. As is the case in this and other highly selective universities in the United States, students from elite social backgrounds are over-represented in the admissions pool despite the low probability of acceptance (below 10%). In this way, the essays submitted by these students help serve

¹ The human generated text we use was created prior to the release of ChatGPT, making it impossible that our data contained any synthetic or human-AI hybrid text.

Table 1 Descriptive statistics for the human-written and AI-generated essay samples

Source	Applicants	Essays	Format
Public University	35,789	143,156	4 essays (from 8 prompts), merged
Private University	10,619	10,619	1 essay (from 7 prompts)
GPT-3.5		10,000 (Public prompts) 2964 (Private prompt)	(Same as above)
GPT-4		10,143 (Public prompts) 2945 (Private prompt)	(Same as above)

as a counter-exemplary pool of students to the under-represented applicants to the large public university. The private school applicants we analyze represent the entire pool of engineering applicants to this particular university.

Finally, we pair these human-written documents with essays generated using GPT-3.5 and GPT-4. We set the temperature (randomness-in-response hyperparameter²) to 1.20 (min = 0; max = 2.0). We generated 25 essays for each temperature setting of increments of 0.1 and found that 1.2 produced the best results across the essay prompts and formats. Focusing on GPT 3.5 and GPT-4 also gave us additional control by standardizing the temperature hyperparameter instead of estimating temperature settings across models. We found that texts generated with temperatures above 1.20 were more likely to include irrelevant or random text, and unicode errors (e.g., “haci\u8db3³” was included in an essay generated with a temperature of 1.6). Setting the temperature below 1.20 tended to generate highly repetitive texts (e.g., multiple essays generated with temperatures of 0.6 or below would begin with the sentence “As I reflect on my high school experience, one particular event stands out as a turning point in my personal growth and self-awareness.”). Testing the effects of these hyperparameters across models could be addressed in future studies. Our goal was to generate texts which reflected the wide variety of stories, experiences, and narratives that students included in their admissions essays [34] as a way to try and match human writing as closely as possible. We focus on OpenAI’s GPT models because of their usage in ChatGPT, currently the most popular LLM-powered chatbot platform. Future research could take a more purely technical perspective and analyze other LLMs, but we are interested in generating insights into the processes taking place when people create text and use the most popular tools to generate new text.

To match the distribution of the real applicants’ essays, we tailor the prompt based on the empirical distribution of students’ selected essay questions for each respective context. Specifically, we generate an essay by prompting GPT to respond to the same question that the applicant chose from the list of seven Common App essay options or four of the eight UC essay options. We also include the names of the schools (University of California and the anonymized private university) in the prompts. Future studies may consider including demographic information in the LLM prompt. Table 1 provides descriptive statistics for the human-authored and AI-generated essays. Note that we report the number of synthetic essays produced by the GPT models in terms of total documents generated: technically, the public prompts each contained four different essays prompts but we only analyze the merged documents.

² <https://platform.openai.com/docs/api-reference/authentication>.

³ \u8db3 might be a reference to a unicode character.

Table 2 Descriptive statistics of applicant characteristics

Source	Zip codes	Gender	First-generation status	Economic connectedness
Public University	806	F: 61.57% M: 38.43%	First-Gen: 56.35% Continuing-Gen: 43.70%	Min: 0.25 Mean: 0.84 Median: 1.07 Max: 1.70
Private University	3,147	F: 33.89% M: 66.11%	First-Gen: 20.02% Continuing-Gen: 79.98%	Min: 0.35 Mean: 1.06 Median: 1.08 Max: 1.70

For each application essay, we have a variety of metadata reflecting the sociodemographic attributes and contexts of the applicant. Following previous work, the sociodemographic attributes we focus on include first-generation status (whether or not the applicant has at least one parent who completed a college degree) and gender (recorded only as a binary: Male or Female) [4, 35, 49]. These data are important given longstanding barriers for women entering into engineering and the underrepresentation of lower-income students at selective universities. On a practical level, these were also two of the only pieces of information available to us for both the public and private school applicants. Future work should consider other authorship characteristics.

We complement these individual-level features with social context variables, specifically data from the Opportunity Insights Lab⁴. The primary research goal of the Lab is to understand socioeconomic patterning and mobility in the US and often pays particular attention to geography [22]. We connect geographic information for each applicant with socioeconomic data for their ZIP Code. We focus on one particular variable, economic connectedness (EC), for two reasons: (1) of all the metrics generated and collected by the Lab, they claim EC has the strongest relationship to socioeconomic mobility; and (2) previous research finds that application essay features are most strongly correlated with EC [6, 22]. EC for a given ZIP Code is a measurement of the proportion of friendship ties and networks across social classes, such as how many individuals from lower socioeconomic status backgrounds have friends from higher socioeconomic status backgrounds⁵. There is implicit information in EC that makes it a particularly useful metric for socioeconomic status: the amount of friendships containing anyone from high socioeconomic status backgrounds is contingent on how many live in a given ZIP Code (the opposite is also true with respect to lower socioeconomic status). Understanding how AI writing style tracks with economic connectedness could point to more complex, socially embedded ways that linguistic hegemony could be reshaped through LLMs. Table 2 provides descriptive statistics of the sociodemographic characteristics of applicants in our samples.

Analytical approach and measures

After generating essays based on the same prompts as the human applicant essays, we use a variety of statistical and visualization techniques to compare the text by source: public school applicants, private school applicants, and AI-generated text for each respective

⁴ <https://opportunityinsights.org/>.

⁵ The Opportunity Insights Lab provides the following non-technical explanation of economic connectedness: https://opportunityinsights.org/wp-content/uploads/2022/07/socialcapital_nontech.pdf.

writing prompt. We also compare essay features across different authorship sociodemographic characteristics and map them onto the AI-generated text to understand which students write most similarly to AI and vice versa. Hegemony functions by placing traits, preferences, and tendencies associated with one group above others, including language. Our approach will unearth patterns in LLM-produced texts and compare them to people from different sociodemographic backgrounds. If there is consistent alignment between LLM style and writing styles favored by those from higher social privileged backgrounds, the prospect of AI writing style contributing towards the existing machinery of linguistic hegemony would be more likely. The popularity of LLMs could also transform extant forms of linguistic hegemony in writing if the opposite were true. To account for these possibilities, we frame our analyses and results in terms of the predicted context of the AI author when compared to the actual context (e.g., social identities and geographically distributed forms of socioeconomic information) of the human authors. Framing these results in this way highlights the subtle and not so subtle ways that LLMs potentially homogenize language and culture toward a specific sociodemographic group and context.

There are many potential analytical techniques to describe a piece of text; we use the 2015 Linguistic Inquiry and Word Count (LIWC-15, shortened to LIWC hereafter) for multiple reasons. LIWC is a dictionary approach that counts the frequencies of writing features, such as punctuation and pronoun usage, and cross-references those counts with an external dictionary based on psycholinguistics research [70]. Generally, style is understood as the interplay between word selection, semantics, register, pragmatics, affect, and other linguistic dimensions that people use to communicate their ideas in ways that are reflective of their identity; in computational social science, style is usually calculated at the word level. LIWC models style in this way but with special attention to psychological dimensions, such as the ways that linguistic style can predict things like successful romantic matching [41] often in a way to correlate linguistic styles with personality. LIWC is a popular tool for text analysis across many domains, including psychology, social science, and computer science⁶, often in ways that use LIWC as a means to generate numerical features for a given document to use in some kind of predictive framework, including other studies of writing style and social demography [29]. LIWC has also been used in studies of college admissions essays, including studies showing that writing style in the essay is strongly correlated with household income, SAT score, and college GPA at graduation [5, 69]. Using LIWC also allowed us to directly compare the public and private applicant essays, whereas other methods would have violated data-use agreements. A limitation of LIWC in settings like ours is that it generates many variables (we use 76 in our analyses), and false positives are possible. To address this, we present multiple analyses and perspectives as a way to triangulate our results. Future studies with similar questions should consider other methods, but LIWC gives us the opportunity to be in direct conversation with other studies in the same domain of our data (college admissions) as well as many other contexts involving computational text analysis. Table A2 presents the 76 LIWC features we used in our analyses. We exclude several features, such as word count and dash (usage of the non-alphanumeric character “-”),

⁶ For more information on LIWC, see <https://liwc.app/static/documents/LIWC2015%20Manual%20-%20Operation.pdf>.

due to issues like incompatibility across text formats. The word count feature (literally the number of words in a given document) was found to be positively correlated with various sociodemographic features in past studies, but we excluded it because the essay prompts included explicit instructions about document length.

In this study, we take an agnostic approach in selecting which of the many LIWC features to examine since there is limited literature on the relationship between human-written and AI-generated text. After calculating LIWC features for each document, we compare them with respect to each set of documents (human or AI-generated, public or private school prompts) and human authorship characteristics. Our first set of analyses focuses on the distributions for each LIWC feature across each set of documents, drawing from sociolinguistic research that also analyzes variation and distributions of communicative practices. To compare the distributions, we use the Kolmogorov-Smirnov (KS) statistical test. The KS test, along with other analyses that compare distributions like Kullback-Liebler divergence, is widely used in data science and machine learning as a way to compare the likelihood that two continuous variables were drawn from the same distribution (in the two sample test case) [77]. Formally, we are comparing the empirical cumulative distribution functions of human written essays F and AI generated essays G for each LIWC feature X for each sample size n for humans and m for AI. The KS test compares the distribution using the following equation:

$$D = \sup_x |F_n(x) - G_m(x)| \quad (1)$$

D is then used as a test statistic. We use 0.05 as a standard threshold for statistical significance that the two samples came from different distributions (i.e., rejecting the null hypothesis that they came from the same distribution). We use the KS test to determine which LIWC features vary the most across samples based on their distributions. From this basic analytical framework, we will also compare the style features across sociodemographic characteristics of the human authors. Adding the social dimensions will then show which groups of student writing features (based on gender, first-generation status, and EC scores) are most similar to the AI. Combined, these analyses address our first research question. Linguistic hegemony functions through overt and covert associations that people make between themselves and other people based on idealized, standardized forms of communicating. The KS test comparing writing style distributions and the proximity to the AI-generated essay distributions are used to capture this dynamic.

We also compare essays as represented by the full set of their LIWC features using two methods. First, we use cosine similarities to find similar essays given the vector representation of all essays [8]. Here, cosine similarity is the metric used to find each AI-generated essay's most similar human-written essay (i.e, their "twin" essay). To find "twin" essays, we take each vector of LIWC features from human-written essays \mathbf{f} and AI-generated essays \mathbf{g} and calculate the cosine similarities for each human-AI pair. We then report which the demographic characteristics of the humans that had the highest cosine similarity for each AI generated essay. We compute cosine similarity as the dot product of the vectors divided by the product of the norm for each of the vectors:

$$\frac{\mathbf{f}\mathbf{g}}{\|\mathbf{f}\|\|\mathbf{g}\|} \quad (2)$$

We report the demographic breakdowns of the human authors for the twin documents to address research question two.

Next, we fit linear regression models where we regress information about the applicants, specifically the EC score for their self-reported ZIP Code on LIWC features. This model will then be applied to the AI-generated essays to impute the same characteristics [51] to answer our third research question. Past studies using LIWC features to predict sociodemographic characteristics found high predictive power (adjusted R^2 of 0.44 when predicting SAT score) [5]. Formally, after fitting the LIWC features to the human sociodemographic features, we use the coefficients to predict the social context of the AI (using the same notation as [51]):

$$\hat{y} = \beta X \quad (3)$$

To estimate the coefficients for predicted EC, we use 10-fold cross-validation with an 80/20 train and test split to prevent overfitting in training the model. The final measurement we report is the adjusted R^2 . Note that the AI-generated essays do not have the prediction outcome available (i.e., we do not have an EC score for them); rather, we are using a common prediction framework to estimate an EC score given the linguistic features of the essay (i.e., the “predicted context” for an AI-generated essay).

Results

Our results are organized as follows: First, we present direct comparisons and contrasts of LIWC features between human and AI-generated text. Second, we compare essays across sociodemographic variables, specifically first-generation status and gender. Third, we report findings of the “twins” analysis using cosine similarity. Finally, we present the results from the linear regression analysis on the predicted context of the AI-generated essays. Overall, the results for GPT-3.5 and GPT-4 were similar, though in various instances, GPT-4 exhibited stronger tendencies than those of GPT-3.5. This includes the relative similarity of statistical curves and distributions, less variance, and less variation in the human twin essay analysis (likely a product of the lower overall variation). GPT-4 is more expensive to use, making it less accessible, especially for high school students writing admissions essays. Therefore, we focus most (but not all) of the discussion of the results on the GPT-3.5 analyses.

Direct comparisons

Before comparing the writing styles of humans and AI along social dimensions, we compare the writing between essay prompts and applicant pools. The figures and findings we describe in this and the subsequent section present statistically significant differences, according to the KS test, between the human-written and AI-generated essay style features. These visual differences will also yield insights into basic questions on whether or not AI writes like humans, something that the rest of our analyses focus on with the rejoinder of “which humans.” The distributions for the AI stylistic features we present in this section will be used throughout the rest of the paper (as opposed to the disaggregated features for the human authors). Across many of the LIWC features, the AI distributions tended to be notably different from the human

distributions. These differences were also notable in visualizations of the distributions of these features. For example, Fig. 1 shows not only that AI uses longer words (six or more letters long, called Sixltr in LIWC) but also that the variation is much lower than in the human essays (though this is more pronounced with the essays generated with the public school prompts). We also noted that the distribution of this and other features for the private school applicants was slightly closer to the AI-generated essays. While this is representative of many of the LIWC features, some were much less clear. For example, humans and AI tend to write about affiliations (with groups, people, organizations, and friends) at similar rates despite the AI not actually having any affiliations (see Fig. 1). Future sociological studies might consider comparing the kinds of group memberships claimed by LLMs (as captured by the “affiliations” LIWC feature) with those of humans. In the case where the distributions were more distinctive (such as the Sixltr distributions), the private school applicants were slightly closer to the distributions for the AI-generated essays.

It was also the case that for many features, the distributions for AI-generated essays were more narrow (i.e., lower variance). These patterns of slight levels of similarity between some but not all applicants and lower levels of variance for the AI essay features also emerged in our social comparison analyses. Another way to interpret these results is that LLMs have a limited range in the text that they generate, which becomes obvious only when compared to human variation across a consistent set of stylistic dimensions. This would make LLMs ripe for use in hegemonic processes of standardization and homogenization, even if unintended. The slight similarity with some groups and not others, along with the low variance, may further contribute to these processes. Next, we examine applicant characteristics beyond the type of school to which they applied (though it is also the case that the public school applicants in our sample are, in the aggregate, underrepresented in US higher education).

Social comparisons

We compared essays between several sociodemographic characteristics of applicants, specifically their gender, first-generation status, and the EC score for their ZIP Code. We found two patterns: (1) the writing style features of AI-generated essays tend to have

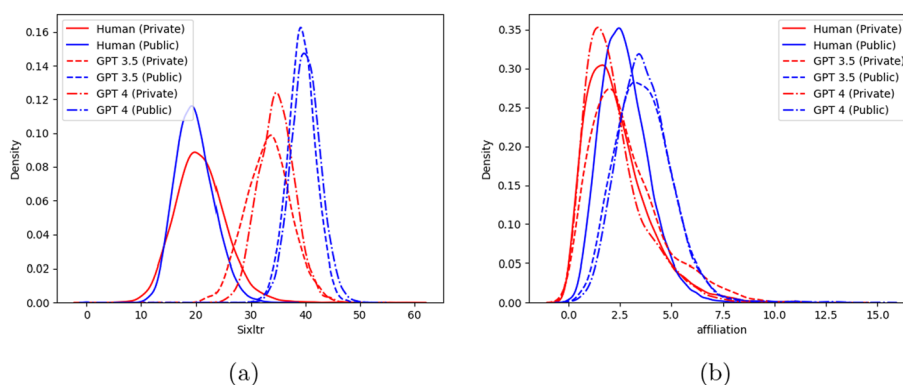


Fig. 1 Distributions of (a) Sixltr (usage of words with six or more letters) and (b) affiliation (affiliations) for each set of essays

different distributions (mean and variance) than human-written essays, and (2) when disaggregating the human-written essays based on the authors' social characteristics, some groups' writing style more closely resembles AI's writing style than that of other groups. First, we present a table of the most demographically distinctive features and whether AI tends to write like any particular group. Then, we examine the distributions of human and AI-generated text that were considered distinct. Both of these analyses used KS tests to measure and compare distinctiveness. We also note which groups of students used each LIWC feature most similarly to the LLM from among the independent distributions. We also present the results for the comparisons of the public and private school applicants separately.

First, we directly answer the question, "Who do LLMs write like?" while limiting the risk of false-positive results given the large number of LIWC features. We identified the LIWC features that were significantly different between the three demographic groups for both Public and Private school applicants using KS tests: gender, parental education level, and socioeconomic context (modeled using ZIP Code-level EC scores). For the identified LIWC features, we compared the subgroup distribution of human-written essays with those of AI-generated essays. Table 3 shows the number of distinctive features for each group as well as the proportion of features most similar to any particular group; for each set of features and demographic groups, the AI-generated essays' features most resembled male applicants with college-educated parents from high-EC ZIP Codes. What these results highlight are the ways that, in contexts where users are directly interacting with the most popular LLMs, there are stylistic features favored by people that are also favored by the models. The features favored by this particular group of students as well as the AI include those found to be predictive of college GPA (article, Analytic; Pennebaker et al. [69]), SAT scores (Sixltr, prep; Alvero et al. [5]), and is more reflective of usage patterns among enrolled college students (focusfuture, space; Alvero et al. [7]). The full list of features is included in the supplementary materials (Table A3). These traits, along with the demographic dimension of who in the data adopts these stylistic features more often, highlight how LLMs could reinforce linguistic hegemony in situations like college admissions.

We next describe our other sociodemographic analyses. For gender, we found more distinctiveness in the distributions of the individual LIWC features for the public school

Table 3 Among the most distinctive LIWC features based on gender, first-generation status, and EC (using a median split to create a low and high EC group), for both the public and private applicants, we present the proportions of features that are the most aligned between the AI-generated essays and each respective subgroup

	Sig. different LIWC features in human-written essay	GPT 4 (Public)	GPT 4 (Private)	GPT 3.5 (Public)	GPT 3.5 (Private)
Gender	29	79.3% male	65.5% male	72.4% male	75.9% male
First-gen. status	37	75.7%cont.-gen.	78.4% cont.-gen.	75.7% cont.-gen.	81.08%cont.-gen.
EC	25	80.0% high EC	88.0% high EC	80.0% high EC	92.0% high EC

For example, among the 29 LIWC features for which the distributions were significantly different in essays written by men versus women applicants, the feature distributions of the AI-generated essays were closer to those of male applicants for at least 65.5% of features (GPT-4; Private applicants) and up to 79.3% of features (GPT-4; Public applicants)

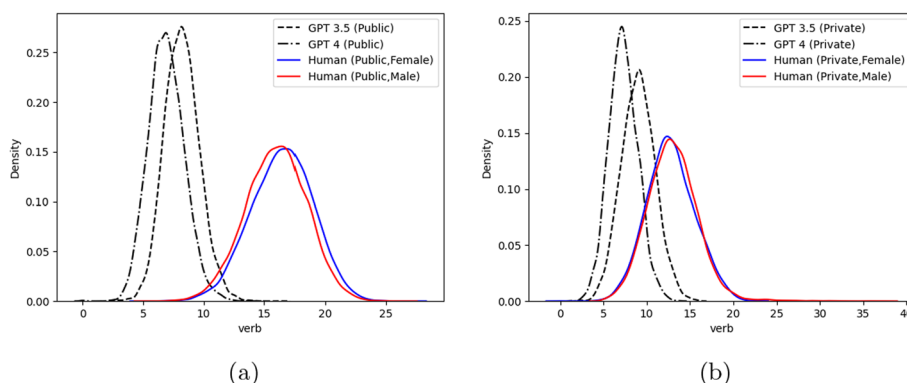


Fig. 2 Distribution of verb (verb usage) for each set of essays split by gender for (a) public school applicants and (b) private school applicants

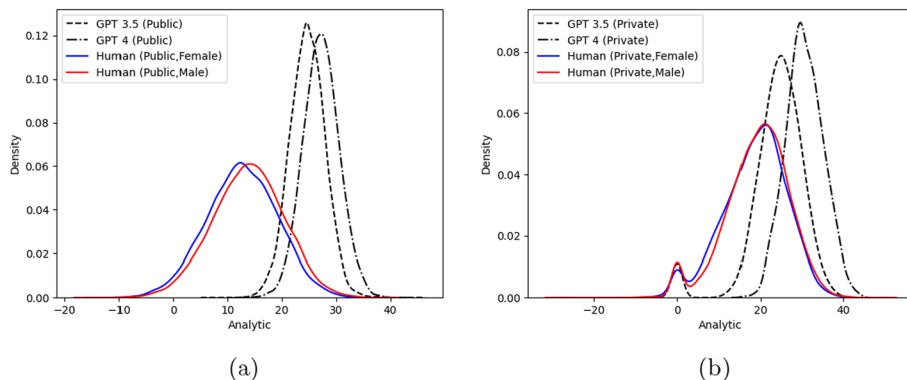


Fig. 3 Distributions of analytic (analytical thinking) for each set of essays split by gender for public school applicants (a) and private school applicants (b)

applicants’ essays than for private school applicants’ essays relative to those generated by the LLM (48 features for private school applicants, 75 for public school applicants). For many of these individual features, including common verb usage (see Fig. 2), the differences between the AI-generated and human-written essays were appreciably greater than the differences between essays written by men and women. The LIWC features for women were slightly more similar to those of LLMs for the public school prompt, but this pattern did not hold for private school applicants.

For example, while the difference between LLMs and humans in the use of Analytic language (Fig. 3) was smaller than Fig. 2, it was also the case that male applicants were slightly more similar to the LLM. Beyond their statistically significant differences, we highlight these LIWC features because of their associations with college GPA in previous work (negative for verb, positive for Analytic), suggesting potentially broader implications for these small differences [69]. Though these differences are not large, they could become magnified in different social contexts and situations.

The differences in writing style between first-generation and continuing-generation students were similar to those between gender groups: many stylistic features were distinct between the human-written and AI-generated essays, but many of the features were slightly more similar to one group than the other (Fig. 4). In this case,

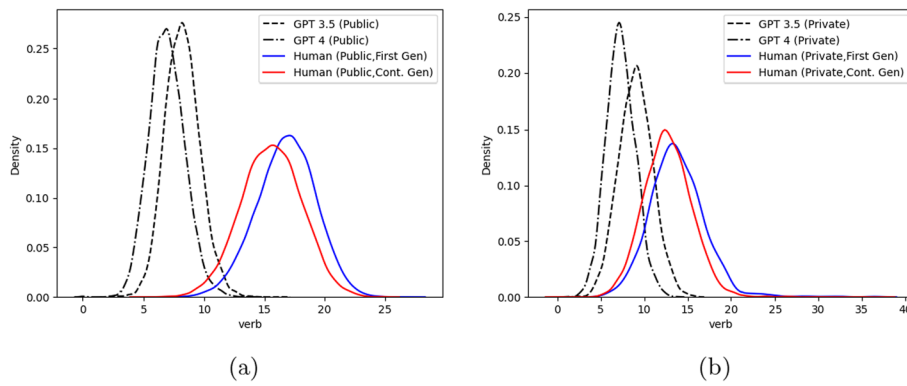


Fig. 4 Distribution verb (usage of common verbs) for each set of essays split by first-generation status for (a) public school applicants and (b) private school applicants

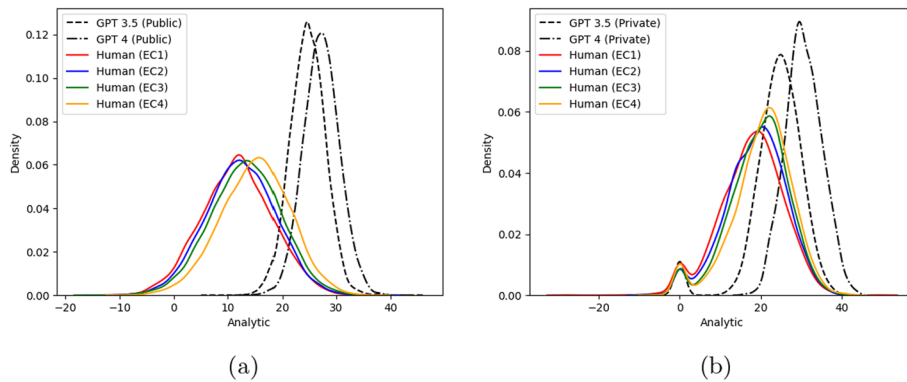


Fig. 5 Distribution of Analytic (analytical thinking) for each set of essays split by EC quartiles for (a) public school applicants and (b) private school applicants

many writing feature distributions of AI-generated essays were closer to the continuing-generation applicants.

Finally, we compare the style features for the AI and human essays but split the applicant essays into EC score quartiles. Like the other results, the distributions between the humans and AI were independent but some students had similar stylistic approaches to the AI than others. For example, the Analytic features for the students from the ZIP Codes with the highest EC scores (represented by the yellow curves in Fig. 5) were closest to the curve for the AI essays. It was also the case that the curves were roughly sequential in their patterning, with the highest EC scores closest to the AI curve and the lowest EC scores farthest away.

The individual writing style features showed differences across social groups. These differences reflected traditional forms of hegemony and social privilege, such as applicants with college-educated parents and those from areas with high levels of EC. For the comparisons of individual LIWC features, it was also the case that AI wrote more like male applicants among the most gendered stylistic features. But this was not universally the case, as some of the style features were more closely aligned with women. Although women are not typically associated with processes of hegemony, they tend to submit stronger overall applications than men [35], pointing to ways

that context could shape answers to the question of which students are writing most like LLMs. This, paired with the low variation in these same features for the AI, show how standardization and homogenization is likely but also likely to become associated in the writing styles of certain groups of people in hyperfocused ways. The next two subsections consider the entire set of variables simultaneously.

Cosine similarity derived twin

We generated cosine similarities for each essay pair in the public and private applicant data. Because the AI essays had low variation for the LIWC features, many of the most similar documents were other AI essays. See Table 4 for a breakdown of the characteristics of the authors of the twin essays and the respective applicant pools.

Regardless of the essay source (public or private school applicants), the authors of the twin essays were more likely to be women, continuing gen, and come from ZIP Codes with higher EC than their respective applicant pools. It was also the case that the differences for gender were more pronounced for the public school applicants and the differences for first-gen status were far more pronounced for the private applicants. Among the twin essays for the private school applicants, approximately 95% were written by students who had at least one college-educated parent compared with 82% in the actual applicant pool. Combined, these results suggest that AI writing styles (in a holistic sense) are more similar to students from more privileged backgrounds (continuing gen) but also from students who tend to perform better in other metrics in the same process (women). Conceptions of hegemony tend to focus more on the former, but more nuance might be needed with respect to the latter.

Authors of the twin essays also tended to come from ZIP Codes with higher EC scores than the applicant pools, but this trend was more than five times stronger for the public school applicants (difference of 0.16) than for the private school applicants (difference of 0.03). The types of communities reflected in these average scores were also different between the public and private school applicants. The three ZIP Codes closest to the average EC score for the public school applicants were located in San Marcos (92069, EC = 0.837); Anaheim (92801, 0.837), and Dixon (rural community southwest of Sacramento, 95620, 0.835). The public applicant twin essays were on average most similar to essays from Morro Bay (93442, 1.00); Coronado (adjacent to downtown San Diego, 92101, 0.99); and Auburn (a suburb of Sacramento, 95603, 1.02). While there is some variation across some locales, the most stark difference between these communities for

Table 4 Sociodemographic distribution of human-written essays and their AI-generated “twin” essay

Source	% Female	% Male	% First-Generation	% Continuing-generation	Mean EC (SD)
Public applicants	61.59	38.41	56.30	43.70	0.84 (0.25)
Public twin essay authors (GPT-3.5)	71.66	28.34	53.52	46.48	0.87 (0.23)
Public twin essay authors (GPT-4)	44.70	55.30	56.08	43.92	0.89 (0.19)
Private applicants	33.79	66.21	20.02	79.97	1.06 (0.25)
Private twin essay authors (GPT-3.5)	31.55	68.45	4.32	95.68	1.13 (0.25)
Private twin essay authors (GPT-4)	30.25	69.75	6.9	93.14	1.14 (0.24)

the applicants are the respective percentages of Hispanic residents⁷. The ZIP Codes in the actual applicant pool range from 40% (San Marcos) to 54% (Anaheim). In contrast, the Hispanic population percentages of the ZIP Codes and EC scores for the twin essays range from 13% (Auburn) to 19% (Coronado). These communities with higher EC also had less ethnic homophily for the entirely Latinx applicants in our dataset.

Although the private school applicants were not limited to California like the public school applicant data, we took the same approach to get a sense of the types of communities that are writing essays most similar to the AI-generated text and vice versa. Private school applicants tended to come from communities with higher EC than the public school applicants, and in the California context, these were also communities with higher socioeconomic status. The most similar ZIP Codes were located in San Pedro (90732, 1.07); Murrieta (92591, 1.067); and Pasadena (91106, 1.07). All of these communities were middle to upper-middle-class suburbs in the Los Angeles metropolitan area. When mapped onto California, the authors of the human twin essays come from well-known upper-middle to upper-class communities, including San Jose (downtown area, 95116, 1.10); Santa Monica (90404, 1.10); and Pacific Palisades (90272, 1.10). While more analysis would be needed on the specific communities where the private school applicants hail from, these California-based trends show how AI writing styles also reflect demographic and socioeconomic variation in ways that mirror segregation and social status.

Predicted social context

Finally, we present the results for the predicted context of the AI essays through predicted EC scores. The linear model fitted to essays and EC values of public school applicants had adequate predictive power with an adjusted R^2 value of 0.57, which matches past results using a similar approach [6]. However, the model for private school applicant essays achieved lower predictive power with an R^2 value of 0.15. To emphasize more reliable results, we report findings from public school applicants in this section. The discrepancy in the models' predictive power could be explained by the following two factors. First, the private school applicants came from communities with higher levels of EC and lower variation in EC; the lower variation would mean that the model would likely have a harder time distinguishing patterns. Second, the difference in variation might also be explained by which students are drawn to apply to highly selective universities and the most selective programs in those same schools (engineering). Many of the public school applicants had access to fee waivers to submit their materials given their socioeconomic status (including those who also applied to the private school), but if they felt as if their chances of acceptance were extremely low then they might not feel compelled to even apply.

Using the model trained on the public school applicant LIWC features, we predicted the EC scores for each of the AI essays⁸. Similar to the cosine twin analysis, the average imputed EC was higher (1.33) than the average EC for the applicant pool at the level of

⁷ See <https://www.census.gov/quickfacts/>.

⁸ We remind the reader that the AI generated essays did not have an actual EC, our model is predicting what the EC *would* have been given the style features.

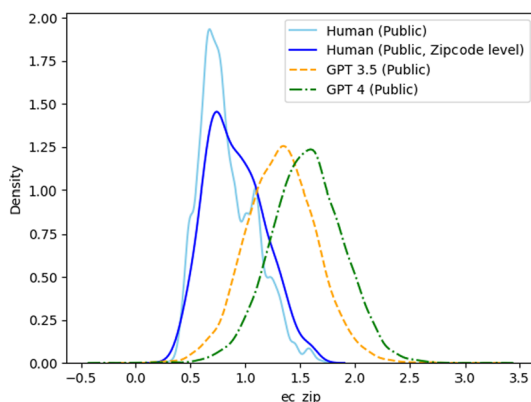


Fig. 6 Distributions of EC scores at the ZIP code level and predicted values for AI essays with individual-level EC

the individual applicant (0.84) and the average ZIP Code (0.91); Fig. 6 shows these distributions. Visually, these differences may not seem particularly large, but in terms of EC these differences were substantial. For example, among the ZIP Codes of public school applicants in our dataset, this difference in EC is the equivalent of slightly below median EC (0.88) with higher than the 90th percentile (1.28).

All of the results we just described are focused specifically on the analyses of essays generated using GPT-3.5. But as Fig. 6 shows, these trends become exaggerated with GPT-4 as it has an even higher average predicted EC score. Ironically, the higher cost to use GPT-4 mediates access to the tool based on income and then uses a writing style associated with people from the ZIP Codes with the highest levels of social mobility.

With the direct and social comparison analyses, we noted how many stylistic features were independent between humans and AI while also noting that some groups of students tended to write slightly more similarly to the AI. But the predicted context and cosine twin analyses, each of which incorporated all of the style features, show that these subtle differences quickly accumulate to produce text that resembles writing styles of students from certain backgrounds. The relatively low levels of variance for the individual features were not as prominent here as they were for the direct and social comparison findings, but future studies might consider which writing style features are most noticeable to human readers to see if only one or several style markers are prominent.

Discussion and conclusion

In this study, we considered the intersection of LLMs, social demography, and hegemony through analyses of college admissions essays submitted by applicants to a public university system and an engineering program at a private university. Using a popular dictionary-based method, LIWC, we compare the writing styles of human and AI-generated text in response to the same writing prompts for each applicant pool. Our findings were generated through direct comparisons of the LIWC features between the documents, social comparisons using demographic information of the applicants, and compositional analyses using all of the LIWC features (cosine similarity twins and the predicted context). We find that for individual stylistic features, LLMs are generally distinct from humans: they used various LIWC features either systematically more or less

than humans, and used them more consistently (i.e., with less variation across essays). The reduced variation could potentially narrow the scope of what is considered an appropriate way to present oneself in writing if the treatment of AI as an “oracle” that is always correct persists [59]. It was also the case that, when considering authorship identities and characteristics and individual features, the differences between the human-written and AI-generated essays were greater than the differences between groups of students (e.g., between men and women). However, in both sets of findings, it was also the case that some groups of students used features slightly more like the AI than other groups. In our analyses using all of the features, these differences became more acute as the essays resembled those from areas of higher EC than the average student.

However, the social comparison analyses also show that the writing styles of LLMs do not represent any particular group [27]. Though the feature distributions for the AI-generated essays were indeed closer to one sociodemographic group than another, as seen in Fig. 3 and Table 3 for example, they were also quite distinct from those human groups. This points to two reasonable interpretations of the similarity between human and AI-generated text: a distributional perspective and a sampling perspective (see Fig. 7 for a conceptual diagram). First, we might consider a distributional perspective (not to be confused with [12]). If the distribution of the writing style features is so distinctive and unaligned with any human style of writing, future studies might examine the extent to which people consider LLM-generated text unhuman and artificial. Second, we might consider a sampling perspective where we might focus less on the curves of the distributions of writing features and more on the peaks and expected characteristics. For example, $\mathbb{E}[Analytic|LLM\ author]$ would be on the higher ends of the distributions for a feature that is used more often by men (Fig. 3) or less (Fig. 2). Taking this interpretation further and considering the low variance we also noted, these results could be interpreted as indicating that the writing style features of LLMs are reflective of the most masculine (in the case of Fig. 3) or whatever group is most similar to the AI. Our analyses lend themselves more readily to the sampling perspective given our considerations of how individual students could be interacting with and comparing their writing

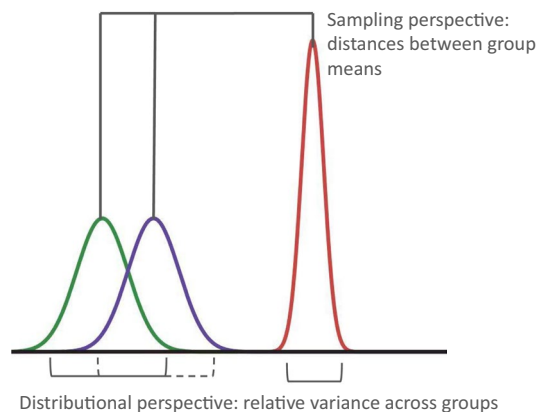


Fig. 7 A conceptual diagram of the distributional and sampling perspectives for comparing human (left-hand distributions) and AI (right-hand distribution) text. A sampling perspective might focus on closeness in means (the peaks of the distributions) whereas a distributional perspective might compare variance in the distributions (the widths of the distributions)

styles with AI-generated text, but future studies could better elucidate the nature of this homogenization and its impacts on society.

Beyond our findings here, future work comparing AI and human writing might benefit from specifying the type of perspective that is guiding a given study. For example, current work on AI alignment tends to focus on the sampling perspective (e.g., “can the AI respond like an average person with some set of demographic characteristics?”). Analyzing alignment from a distributional perspective might instead focus on questions that consider variation in human language and communication as a conceptual starting point. The importance of the latter will grow in prominence as AI increasingly enters into key social decision making processes that would require them to interact with a breadth of sociolinguistic variation. The early results are not promising, as a recent paper demonstrated that AI has strong social biases when given text that contains dialectal features [38]. Further, as people interact with LLMs in their daily lives across many different contexts there is a chance that broad understanding of “correct” ways to write and communicate will become more constrained. In this way, LLMs might undermine some of the ideals of college admissions where students are given a unique opportunity to highlight their experiences, ideas, and identity by shrinking the breadth of those details. Future studies might more explicitly consider how students are using LLM technology to better get at this issue.

Hegemony is a theory that encapsulates the myriad ways that that power is exerted through the supposedly common sense ways we understand the world and who deviates from these norms [36]. LLMs, as examples of extremely popular technology in terms of users and use cases, will undoubtedly play a complex role in modern digital hegemony. College admissions essays are unique in that they provide some creative flexibility for the authors while still being a primary data point used to either admit or reject students. There is a clear tension between sharing an authentic portrayal of one’s life and experiences with the norms and expectations to be able to demonstrate a certain level of writing ability and style. This tension has created an entire cottage industry that helps students balance these expectations [40]. The polished, “fancy” writing style of LLMs might give students enough of an incentive to put aside their own writing style and stories if it gives them an advantage in the admissions process. If the same advantages that come with writing in the same kinds of ways as those from higher social status backgrounds represent the linguistic disposition of LLMs, the many new tools and technologies relying on them could reinforce patterns where the language of some is structurally and systematically favored over the language of many. LLMs and platforms like ChatGPT might not be able to inflict direct control, violence, or power, but by reinforcing extant language standards would contribute to the ideologies people have about language as it relates to power. This is possible because of the many social mechanisms already in place which operate under similar logics and/or have similar outputs.

We hope our analyses can spur future studies that consider how everyday people might interact with LLMs in contexts like writing personal statements (an activity with many analogues in modern society). There is a robust research ecosystem focusing on improving LLMs, better understanding their capacities, and trying to prevent pernicious forms of bias from leaking. But there is room that consider how power in its current forms might be enacted through everyday use of these same tools. Consider the counterfactual

of our findings: “LLMs write like those from more vulnerable social positions”, not like those with (relatively) more power. It is possible that this would become a “problem” to fix, but in a practical sense it would be difficult to imagine LLMs permeating across social institutions as they have without sounding like the people who have historically moved through these same institutions with relative ease [75]. If LLMs did not write in the ways we described, they might not be as popular a tool for things like academic writing and research assistance, writing evaluation more generally, and the construction of the many types of documents in areas outside of education like the law, healthcare, and business.

College admissions is a high-stakes competition, but putting aside one’s language in order to achieve success is the exact kind of hegemonic process in the film “Sorry to Bother You” that we mentioned at the outset of this paper. It is not as if dialectal or non-academic forms of writing are not valuable either, such as the case of AAVE being so widely used on the internet despite its real-life stigmatization in formal contexts [18]. Unlike AAVE, the linguistic styles and markers of the upper middle class have long been held as the standard by which others are evaluated and compared, a trend unlikely to change given our findings. Along these lines, future studies might also compare the specific types of stories and lexical semantics in LLMs to extend our analyses on writing style features generated through LIWC. These might also include studies of the multilingual capacities of LLMs. Another study found that, despite well-documented social stigmas, approximately 20% of UC applicants include some form of Spanish in their admissions essays, whereas 0% of the synthetic text we examine in our study did the same [3]. Other studies might consider how not just the text generated by LLMs is stratified, but also things like access to the technology and perceptions of who is able to use it correctly.

If we assume that, like college admissions essays, there is a correspondence between writing style and social demographics, this paper might also shed some light on the demographics of the people who generate the text on the internet which form the training data used to create LLMs. Beyond training, our results also implicate patterns and preferences that go into the massaging and fine-tuning of LLMs prior to their deployment [50]. The last step in the pre-deployment process is reminiscent to how newscasters are trained to speak in a specific way during broadcasts that is intended to convey credibility and authority [33], though here we observe trends with writing and social demographics. Studies of the text on the internet has noted similar trends, such as 85% of Wikipedia’s editors (a major source of training data for LLMs) being White men⁹. Though we do not explicitly consider race, we do see similar trends in terms of specific stylistic features in writing (such as Analytic language) being used more often by male applicants *and* the AI-generated text. But it is also the case that in certain situations, the writing style of LLMs is more similar to the women in our dataset. These sociodemographic trends in writing point to future studies that examine how writing and language with AI could play a role in reproducing essentializing ideologies about not just gender and class but also about race, especially in educational contexts [72]. In this way, we

⁹ See <https://www.thejournal.ie/wikipedia-founder-gender-imbalance-3668767-Oct2017/>.

see our work as helping to lay the foundation for a sociologically oriented complement to ongoing work focused on LLMs and psychology [24]. In our text-saturated world, a demography of text could yield many crucial insights.

Our analysis focused on “out-of-the-box” AI-generated essays without prompts that specify demographic information about the applicants (e.g., write this essay from the perspective of a first-generation college applicant). Although this points to future directions for this kind of work, it is possible that many or most people using LLMs in their daily lives do not include demographic information in their prompts, such as “write this from the perspective of a first-generation, under-represented minority student”. It is also unclear if LLMs can be prompted or manipulated to such a degree that they are able to mimic the lower-level stylistic trends we identify here without explicit instruction to do so (e.g., “use more commas in the output”). As a comparison, one study of similar data found that approximately 20% of the entire University of California applicant pool uses Spanish words or phrases in their admissions essays [3]. Though multilingualism was not the primary focus of this analysis, the out-of-the-box model did not include any Spanish words in the generated text. It would be easy to include these types of instructions, but figuring out which types of instructions and specific linguistic features to include would not be obvious (aside from possibly prompting the model to use “big words”, though that is already the case). It might also be the case that the types of stylistic features most notable to students are stratified in specific ways or shaped through other hegemonic processes.

Extensions of our analyses could focus on different elements of the relationship between humans and LLMs. For example, our out-of-the-box approach based on text generated for the same responses might be contrasted with a study on the types of prompts people use in their everyday lives. These could include comparisons of how people craft prompts for the same goal or task (such as writing a college admissions essay). The text generated from the slightly different prompts could then be analyzed using similar approaches and methods we adopted here. To address the issue of discrepancies in writing style, other studies could consider fine-tuning on custom training data. These studies could evaluate the controllability of LLMs to generate text and writing styles outside the low variance distributions we describe in this paper. Similar results have been observed in the shrinking vocabulary of peer review [53]. Outside of these more computationally focused studies, social scientists might also begin to analyze the trends where LLMs write both like those with traditional social privilege (such as having college-educated parents) as well as those who tend to do well in specific domains, contexts, and processes (such as women in education and college admissions). There are clear tensions in terms of hegemony: if writing like people who have privilege or are generally more successful, should other students adopt LLMs to assist in their writing in earnest? How might this exacerbate or ameliorate social inequality as it pertains to language and writing? These questions could be used to guide future studies not just on authors but also evaluators of text in a given situation, such as college admissions officers who read essays and evaluate applicants. Given the way that text is so widely used to evaluate people, the stakes of these answers are quite high, and the trends we describe in our analyses point to plausible hypotheses in many different domains.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40537-024-00986-7>.

Supplementary material 1.

Acknowledgements

We are grateful for the comments and feedback received at (1) the Generative AI and Sociology Workshop at Yale University organized by Daniel Karell and Thomas Davidson; (2) the 2024 Monash-Warwick-Zurich Text-as-Data Workshop organized by Elliott Ashcall, Sascha O. Becker, and Philine Widmer; (3) the Pens & Pixels Generative AI in Education Conference organized by Mark Warschauer and Tamara Tate.

Author contributions

All authors contributed to the writing, analysis, and interpretation for the paper.

Funding

Not applicable.

Availability of data and materials

Raw essays are not available. Public school data (LIWC features for the essays) are available on Harvard dataverse; private school data (respective LIWC features) will be made available soon. Data for the public essays is available here: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FJSE4&version=DRAFT>. We are finalizing the data availability for the private applicant essays. The same goes for the synthetic essays. A preprint of this work is available here: <https://osf.io/preprints/socarxiv/qfx4a>.

Declarations

Ethics approval and consent to participate

IRB approval for all data used.

Consent for publication

Authors agreed on the final version. Consent was not necessary for applicants.

Competing interests

Authors declare no competing interests.

Received: 20 November 2023 Accepted: 26 August 2024

Published online: 27 September 2024

References

- Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*. 2023;15(2):e35179.
- Alvero A. Sociolinguistic perspectives on machine learning with text data. In: *The Oxford handbook of the sociology of machine learning*. Oxford University Press; 2023. <https://doi.org/10.1093/oxfordhb/9780197653609.013.15>.
- Alvero A, Pattichis R. Multilingualism and mismatching: Spanish language usage in college admissions essays. *Poetics*. 2024;105: 101903.
- Alvero A, Arthurs N, Antonio al, et al. AI and holistic review: informing human reading in college admissions. In: *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*; 2020. pp. 200–206.
- Alvero A, Giebel S, Gebre-Medhin B, et al. Essay content and style are strongly related to household income and SAT scores: Evidence from 60,000 undergraduate applications. *Sci Adv*. 2021;7(42):eabi9031.
- Alvero AJ, Luqueño L, Pearman F, Antonio AL. Authorship identity and spatiality: social influences on text production. *SocArXiv*. 2022. <https://doi.org/10.31235/osf.io/pt6b2>.
- Alvero A, Pal J, Moussavian KM. Linguistic, cultural, and narrative capital: computational and human readings of transfer admissions essays. *J Comput Soc Sci*. 2022;5(2):1709–34.
- Amer AA, Abdalla HI. A set theory based similarity measure for text clustering and classification. *J Big Data*. 2020;7(1):74.
- Atari M, Xue M, Park P, et al. Which humans? 2023. <https://doi.org/10.31234/osf.io/5b26t>.
- Barrett R, Cramer J, McGowan KB. *English with an accent: language, ideology, and discrimination in the United States*. London: Taylor & Francis; 2022.
- Bastedo MN, Bowman NA, Glasener KM, et al. What are we talking about when we talk about holistic review? Selective college admissions and its effects on low-ses students. *J High Educ*. 2018;89(5):782–805.
- Bellemare MG, Dabney W, Munos R. A distributional perspective on reinforcement learning. In: *International conference on machine learning*, PMLR, pp 449–458. 2017.
- Bender EM, Gebru T, McMillan-Major A, et al. On the dangers of stochastic parrots: can language models be too big?. In: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. ACM, Virtual Event Canada; 2021. pp 610–623. <https://doi.org/10.1145/3442188.3445922>.
- Berkovsky S, Hijikata Y, Rekimoto J, et al. How novelists use generative language models: an exploratory user study. In: *23rd International Conference on Intelligent User Interfaces* 2020.

15. Bernstein B. Elaborated and restricted codes: their social origins and some consequences. *Am Anthropol*. 1964;66(6):55–69.
16. Boelaert J, Coavoux S, Ollion E, Petev ID, Präg P. Machine bias. Generative large language models have a view of their own. 2024. <https://doi.org/10.31235/osf.io/r2pnb>.
17. Bohacek M, Farid H. Nepotistically trained generative-AI models collapse. arXiv preprint. 2023. [arXiv:2311.12202](https://arxiv.org/abs/2311.12202).
18. Brock Jr A. Distributed blackness. In: *Distributed blackness*. New York University Press, New York; 2020.
19. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. arXiv preprint. 2020. [arXiv:2005.14165](https://arxiv.org/abs/2005.14165).
20. Chen L, Zaharia M, Zou J. How is ChatGPT's behavior changing over time? *Harvard Data Science Review*. 2024. <https://hdsr.mitpress.mit.edu/pub/y95zitmz>.
21. Chen Z, Gao Q, Bosselut A, et al. DISCO: distilling counterfactuals with large language models. In: *Proceedings of the 61st annual meeting of the association for computational linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Toronto, Canada; 2023. pp 5514–5528, <https://doi.org/10.18653/v1/2023.acl-long.302>.
22. Chetty R, Jackson MO, Kuchler T, et al. Social capital I: measurement and associations with economic mobility. *Nature*. 2022;608(7921):108–21.
23. Davidson T. Start generating: Harnessing generative artificial intelligence for sociological research. *Socius*. 2024. 10.23780231241259651.
24. Demszky D, Yang D, Yeager DS, et al. Using large language models in psychology. *Nat Rev Psychol*. 2023;2:688–701.
25. Dev S, Monajatipoor M, Ovalle A, et al. Harms of gender exclusivity and challenges in non-binary representation in language technologies. In: Moens MF, Huang X, Specia L, et al (eds) *Proceedings of the 2021 conference on empirical methods in natural language processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021. pp 1968–1994, <https://doi.org/10.18653/v1/2021.emnlp-main.150>.
26. Dhingra H, Jayashanker P, Moghe S, et al. Queer people are people first: deconstructing sexual identity stereotypes in large language models. <http://arxiv.org/abs/2307.00101>, [cs] 2023.
27. Dominguez-Olmedo R, Hardt M, Mender-Dünner C. Questioning the survey responses of large language models. arXiv preprint [arXiv:2306.07951](https://arxiv.org/abs/2306.07951). 2023.
28. Egami N, Hinck M, Stewart B, et al. Using imperfect surrogates for downstream inference: design-based supervised learning for social science applications of large language models. *Adv Neural Inf Process Syst*. 2024;36.
29. Eichstaedt JC, Schwartz HA, Kern ML, et al. Psychological language on Twitter predicts county-level heart disease mortality. *Psychological science*. 2015;26(2):159–69.
30. Feder A, Keith KA, Manzoor E, et al. Causal inference in natural language processing: estimation, prediction, interpretation and beyond. *Trans Assoc Comput Linguist*. 2022;10:1138–58.
31. Field A, Blodgett SL, Waseem Z, et al. A survey of race, racism, and anti-racism in nlp. In: *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (Volume 1: Long Papers)* 2021.
32. Gallegos IO, Rossi RA, Barrow J, et al. Bias and fairness in large language models: a survey. <http://arxiv.org/abs/2309.00770>, [cs] 2023.
33. Gasser E, Ahn B, Napoli DJ, et al. Production, perception, and communicative goals of american newscaster speech. *Lang Soc*. 2019;48(2):233–59.
34. Gebre-Medhin B, Giebel S, Alvero A, et al. Application essays and the ritual production of merit in us selective admissions. *Poetics*. 2022;94: 101706.
35. Giebel S, Alvero A, Gebre-Medhin B, et al. Signaled or suppressed? how gender informs women's undergraduate applications in biology and engineering. *Socius*. 2022;8:23780231221127536.
36. Gramsci A. *Prison notebooks*, vol. 1. New York: Columbia University Press; 1992.
37. Gururangan S, Card D, Dreier S, et al. Whose language counts as high quality? Measuring language ideologies in text data selection. In: *Proceedings of the 2022 conference on empirical methods in natural language processing*, pp 2562–2580. 2022.
38. Hofmann V, Kalluri PR, Jurafsky D, et al. Ai generates covertly racist decisions about people based on their dialect. *Nature*. 2024; 1–8.
39. Hovy D, Bianchi F, Fornaciari T. “You sound just like your father” commercial machine translation systems include stylistic biases. In: *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp 1686–1690, 2020.
40. Huang TJ. Translating authentic selves into authentic applications: private college consulting and selective college admissions. *Sociol Educ*. 2023; p 00380407231202975.
41. Ireland ME, Slatcher RB, Eastwick PW, et al. Language style matching predicts relationship initiation and stability. *Psychol Sci*. 2011;22(1):39–44.
42. Karell D, Sachs J, Barrett R. Synthetic duality: a framework for analyzing natural language generation's representation of social reality. *SocArXiv*, 2023.
43. Kim J, Lee B. Ai-augmented surveys: leveraging large language models for opinion prediction in nationally representative surveys. arXiv preprint [arXiv:2305.09620](https://arxiv.org/abs/2305.09620). 2023.
44. Kirk HR, Vidgen B, Röttger P, et al. The benefits, risks and bounds of personalizing the alignment of large language models to individuals. *Nat Mach Intell*. 2024;1–10.
45. Kirk HR, Whitefield A, Röttger P, et al. The PRISM alignment project: what participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models. arXiv preprint. 2024. [arXiv:2404.16019](https://arxiv.org/abs/2404.16019).
46. Koenecke A, Nam A, Lake E, et al. Racial disparities in automated speech recognition. *Proc Natl Acad Sci*. 2020;117(14):7684–9.
47. Kotek H, Dockum R, Sun D. Gender bias and stereotypes in large language models. In: *Proceedings of the ACM collective intelligence conference*. ACM, Delft Netherlands, pp 12–24, 2023. <https://doi.org/10.1145/3582269.3615599>.
48. Koyuturk C, Yavari M, Theophilou E, et al. Developing effective educational chatbots with ChatGPT prompts: insights from preliminary tests in a case study on social media literacy (with appendix) 2023.

49. Lee J, Thymes B, Zhou J, et al. Augmenting holistic review in university admission using natural language processing for essays and recommendation letters. arXiv preprint [arXiv:2306.17575](https://arxiv.org/abs/2306.17575). 2023.
50. Lee J, Hicke Y, Yu R, et al (in press) The life cycle of large language models in education: a framework for understanding sources of bias. *Br J Educ Technol*.
51. Leiby BD, Ahner DK. Multicollinearity applied stepwise stochastic imputation: a large dataset imputation through correlation-based regression. *J Big Data*. 2023;10(1):23.
52. Liang W, Yuksekgonul M, Mao Y, et al. Gpt detectors are biased against non-native English writers. *Patterns*. 2023;4(7):100779.
53. Liang W, Izzo Z, Zhang Y, et al. Monitoring AI-modified content at scale: a case study on the impact of chatgpt on ai conference peer reviews. arXiv preprint. 2024. [arXiv:2403.07183](https://arxiv.org/abs/2403.07183).
54. Liang W, Zhang Y, Wu Z, et al. Mapping the increasing use of LLMs in scientific papers. arXiv preprint 2024. [arXiv:2404.01268](https://arxiv.org/abs/2404.01268).
55. Lira B, Gardner M, Quirk A, et al. Using artificial intelligence to assess personal qualities in college admissions. *Sci Adv*. 2023;9(41):eadg9405.
56. Lucy L, Blodgett SL, Shokouhi M, et al. "One-size-fits-all"? Observations and expectations of NLG systems across identity-related language features. <https://arxiv.org/abs/2310.15398>, [cs] 2023.
57. Martin JL. The ethico-political universe of ChatGPT. *J Soc Comput*. 2023;4(1):1–11. <https://doi.org/10.23919/JSC.2023.0003>, <https://ieeexplore.ieee.org/document/10184066/> 2023.
58. Matelsky JK, Parodi F, Liu T, et al. A large language model-assisted education tool to provide feedback on open-ended responses. <https://arxiv.org/abs/2308.02439>, [cs] 2023.
59. Messeri L, Crockett M. Artificial intelligence and illusions of understanding in scientific research. *Nature*. 2024;627(8002):49–58.
60. Mirowski P, Mathewson KW, Pittman J, et al. Co-writing screenplays and theatre scripts with language models: evaluation by industry professionals. In: *Proceedings of the 2023 CHI conference on human factors in computing systems*, pp. 1–34, 2023.
61. Motoki F, Pinho Neto V, Rodrigues V. More human than human: measuring ChatGPT political bias. *Public Choice*. 2023. <https://doi.org/10.1007/s11127-023-01097-2>.
62. Mökander J, Schuett J, Kirk HR, et al. Auditing large language models: a three-layered approach. *AI Ethics*. 2023. <https://doi.org/10.1007/s43681-023-00289-2>.
63. Navigli R, Conia S, Ross B. Biases in large language models: origins, inventory, and discussion. *J Data Inf Qual*. 2023;15(2):1–21. <https://doi.org/10.1145/3597307>.
64. Olson W. The shame of Spanish: cultural bias in English first legislation. *Chicano-Latino L Rev* 11:1. Publisher: HeinOnline 1991.
65. Omiye JA, Lester JC, Spichak S, et al. Large language models propagate race-based medicine. *NPJ Digit Med*. 2023;6(1):1–4. <https://doi.org/10.1038/s41746-023-00939-z>.
66. Padmakumar V, He H. Does writing with language models reduce content diversity? In: *The Twelfth International Conference on Learning Representations*. 2023.
67. Payne AL, Austin T, Clemons AM. Beyond the front yard: the dehumanizing message of accent-altering technology. *Appl Linguistics*. 2024; amae002.
68. Pellert M, Lechner CM, Wagner C, et al. AI psychometrics: assessing the psychological profiles of large language models through psychometric inventories. *Perspect Psychol Sci*. 2023;19(5):808–26.
69. Pennebaker JW, Chung CK, Frazee J, et al. When small words foretell academic success: the case of college admissions essays. *PLoS One*. 2014;9(12): e115844.
70. Pennebaker JW, Boyd RL, Jordan K, et al. The development and psychometric properties of liwc2015. Tech. rep. 2015.
71. Rajashekar NC, Shin YE, Pu Y, et al. Human-algorithmic interaction using a large language model-augmented artificial intelligence clinical decision support system. In: *Proceedings of the CHI conference on Human Factors in Computing Systems*. 2024. pp. 1–20.
72. Roth WD, van Stee EG, Regla-Vargas A. Conceptualizations of race: essentialism and constructivism. *Annu Rev Sociol*. 2023;49.
73. Rozado D. The political biases of ChatGPT. *Soc Sci*. 2023;12(3):148.
74. Sakirin T, Said RB. User preferences for chatgpt-powered conversational interfaces versus traditional methods. *Mesop J Comput Sci*. 2023;2023:24–31.
75. Schneider B. Multilingualism and AI: the regimentation of language in the age of digital capitalism. *Signs Soc*. 2022;10(3):362–87.
76. Sejnowski TJ. Large language models and the reverse turing test. *Neural Comput*. 2023;35(3):309–42.
77. Severiukhina O, Kesarev S, Bochenina K, et al. Large-scale forecasting of information spreading. *J Big Data*. 2020;7:1–17.
78. Shen H, Wu T. Parachute: evaluating interactive human-LM co-writing systems. [arXiv:2303.06333](https://arxiv.org/abs/2303.06333). [cs] 2023.
79. Stevens ML. *Creating a class*. Cambridge: Harvard University Press; 2009.
80. Tao Y, Viberg O, Baker RS, et al. Auditing and mitigating cultural bias in llms. arXiv preprint [arXiv:2311.14096](https://arxiv.org/abs/2311.14096). 2023.
81. Thoppilan R, De Freitas D, Hall J, et al. LaMDA: language models for dialog applications. <https://arxiv.org/abs/2201.08239>, [cs] 2022.
82. Törnberg P. ChatGPT-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning 2023.
83. Wang A, Morgenstern J, Dickerson JP. Large language models cannot replace human participants because they cannot portray identity groups. arXiv preprint [arXiv:2402.01908](https://arxiv.org/abs/2402.01908). 2024.
84. Warschauer M, Tseng W, Yim S, et al. The affordances and contradictions of AI-generated text for writers of english as a second or foreign language. *J Second Lang Writing*. 2023;62: 101071. <https://doi.org/10.1016/j.jslw.2023.101071>.
85. Woolard KA. language variation and cultural hegemony: toward an integration of sociolinguistic and social theory. *Am Ethnol*. 1985;12(4):738–48. <https://doi.org/10.1525/ae.1985.12.4.02a00090>.

86. Workshop B, Scao TL, Fan A, et al (2022) Bloom: a 176b-parameter open-access multilingual language model. arXiv preprint [arXiv:2211.05100](https://arxiv.org/abs/2211.05100).
87. Wu S, Irsoy O, Lu S, et al. BloombergGPT: a large language model for finance. <https://arxiv.org/abs/2303.17564>, [cs, q-fin] 2023.
88. Yuan A, Coenen A, Reif E, et al. Wordcraft: story writing with large language models. In: 27th International conference on intelligent user interfaces. ACM, Helsinki Finland, pp 841–852, <https://doi.org/10.1145/3490099.3511105>, 2022.
89. Zhang S, Xu J, Alvero A. Generative AI Meets Open-Ended Survey Responses: Participant Use of AI and Homogenization. 2024. Retrieved from <https://osf.io/preprints/socarxiv/4esdp>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.