Not Like Us, Hunty: Measuring Perceptions and Behavioral Effects of Minoritized Anthropomorphic Cues in LLMs

JEFFREY BASOAH*, University of Washington, USA
DANIEL CHECHELNITSKY*, Carnegie Mellon University, USA
TAO LONG, Columbia University, USA
KATHARINA REINECKE, University of Washington, USA
CHRYSOULA ZERVA, Instituto Superior Técnico, Portugal
KAITLYN ZHOU, Stanford University, USA
MARK DÍAZ, Google Research, USA
MAARTEN SAP, Carnegie Mellon University, USA

As large language models (LLMs) increasingly adapt and personalize to diverse sets of users, there is an increased risk of systems appropriating *sociolects*, i.e., language styles or dialects that are associated with specific minoritized lived experiences (e.g., African American English, Queer slang). In this work, we examine whether sociolect usage by an LLM agent affects user reliance on its outputs and user perception (satisfaction, frustration, trust, and social presence). We designed and conducted user studies where 498 African American English (AAE) speakers and 487 Queer slang speakers performed a set of question-answering tasks with LLM-based suggestions in either standard American English (SAE) or their self-identified sociolect. Our findings showed that sociolect usage by LLMs influenced both reliance and perceptions, though in some surprising ways. Results suggest that both AAE and Queer slang speakers relied more on the SAE agent, and had more positive perceptions of the SAE agent. Yet, only Queer slang speakers felt more social presence from the Queer slang agent over the SAE one, whereas only AAE speakers preferred and trusted the SAE agent over the AAE one. These findings emphasize the need to test for behavioral outcomes rather than simply assuming that personalization would leave to better and safer reliance outcome. They also highlight the nuanced dynamics of minoritized language in machine interactions, underscoring the need for LLMs to be carefully designed to respect cultural and linguistic boundaries while fostering genuine user engagement and trust.

CCS Concepts: • **Computing methodologies** → **Natural language processing**; • **Social and professional topics** → **Race and ethnicity**; **Sexual orientation**; *Cultural characteristics*; • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: Natural Language Processing, Linguistics, Large Language Models, Sociolect, User Perception, User Behavior, Reliance, Anthropomorphization, African American English, Queer Slang

*Both authors contributed equally to this research.

Authors' addresses: Jeffrey Basoah, jeffkb28@uw.edu, University of Washington, Seattle, USA; Daniel Chechelnitsky, dchechel@andrew.cmu.edu, Carnegie Mellon University, Pittsburgh, USA; Tao Long, long@cs.columbia.edu, Columbia University, New York, New York, USA; Katharina Reinecke, reinecke@cs.washington.edu, University of Washington, Seattle, Washington, USA; Chrysoula Zerva, chrysoula.zerva@tecnico.ulisboa.pt, Instituto Superior Técnico, Lisbon, Portugal; Kaitlyn Zhou, katezhou@stanford.edu, Stanford University, Stanford, California, USA; Mark Díaz, markdiaz@google.com, Google Research, Mountain View, California, USA; Maarten Sap, msap2@andrew.cmu.edu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Manuscript submitted to ACM

ACM Reference Format:
Jeffrey Basoah, Daniel Chechelnitsky, Tao Long, Katharina Reinecke, Chrysoula Zerva, Kaitlyn Zhou, Mark Díaz, and Maarten Sap. 2018. Not Like Us, Hunty: Measuring Perceptions and Behavioral Effects of Minoritized Anthropomorphic Cues in LLMs. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 50 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

With large language models (LLMs) being applied in many domains [6, 34, 42, 62, 78] and interacting with more diverse users from different backgrounds [86], there have been increased calls for them to adapt better to users and make them more human like [106]. For example, LLMs are increasingly being personalized to users' language style [90] and being designed to communicate in more polite ways [82]. However, while this increased adaptation and human-likeness can improve user experience [96], this causes various risks such as anthropomorphization and over-reliance [56, 118].

A significant challenge is that naively adapting LLMs to all users can significantly backfire, particularly in the context of users who are from a different culture or from a minority background. Many minority users speak *sociolects*, i.e., culturally specific dialects or speech forms [111], such as African American English (AAE) or Queer slang; their language style is deeply intertwined with their lived experience [36, 97]. For humans to naively adopt the speaking style and mannerisms of a minority interlocutor can be seen as cultural appropriation and offensive [75], raising the question of whether adapting LLMs to minority users poses similar risks [23, 80].

In this work, we empirically explore this question by analyzing perceptions of sociolect adaptation in LLMs and examining the behavioral impacts these adaptations have on users, adding to the growing body of FAccT and adjacent work that highlights the negative consequences of overlooking individual identity and cultural understanding in technology design [10, 40]. We specifically examine how LLM agents that use minoritized anthropomorphic cues, i.e., sociolects, impact the user experience, investigating:

- **RQ1**: Whether a user's *reliance* on the agent depends on whether the LLM agent speaks the user's sociolect or not, to examine possible risks of overreliance due to anthropomorphization [118].
- **RQ2**: Whether a user's *perception* varies depending on whether the agent uses sociolect or not, to quantify possible risks of cultural appropriation [75].
- **RQ3**: Whether *perception* factors (e.g., trust, frustration) are associated with user *reliance*, to explore variation and association between reliance and other factors.

To explore these questions, we devised a within-subjects user study where participants answered factual questions about two videos with the help of suggestions given either by a sociolect-using LLM agent or by a standard American English (SAE) LLM agent, inspired by setups from HCI and natural language processing (NLP) [9, 118]. We assessed users' reliance on each agent behaviorally. Also, we evaluated users' perceptions of each agent across four key dimensions: trust, satisfaction, frustration, and social presence along with a pairwise judgment and rationale of which agent they preferred. We focused our investigations on AAE and Queer slang, motivated by the history of negative experiences Black and LGBTQIA+ users with artificial intelligence (AI) systems [29, 43, 83], as well as the authors' own positionality.

Our results from 498 AAE speakers and 487 Queer slang speakers show that sociolect usage by LLMs influences both reliance and user perception in most cases, highlighting the nuanced relationship between language, behavior, and user perceptions in machine interactions. Notably, AAE participants preferred and relied more on the SAE-using LLM, while, surprisingly, Queer slang speakers showed no significant preference. As expected, AAE participants trusted, were more satisfied with, and experienced less frustration with the SAE-using LLM. However, while Queer slang speakers reported

less frustration with the SAE-using LLM, they felt greater social presence with the Queer slang-using LLM. Thematic coding of the rationales highlights how participant comments mirrored key trends from the quantitative findings, while offering nuanced insights into user perceptions of sociolect-using LLMs.

We conclude by discussing the implications and risks of sociolect usage by LLMs, emphasizing the need to examine both user perceptions and behaviors. Our findings reveal that users prefer SAE outputs, yet sociolect-using LLMs may be perceived as warmer and friendlier in some cases, underscoring the importance of understanding the historical and social nuances of each sociolect. Furthermore, our work addresses broader issues of linguistic bias and inclusivity, highlighting the challenges of sociolectal adaptation and developing inclusive LLMs that respect and do not appropriate minority cultures.

2 BACKGROUND

To contextualize our investigations into the effect of sociolect usage by LLMs, we first provide background on AAE and Queer slang sociolects, then provide an overview of related work examining minority experiences with LLMs and anthropomorphism in artificial intelligence (AI).

2.1 Sociolects and Biases in LLMs

Sociolects are a dialect, or variant, that differs from the standard form of a language that is primarily associated with a specific social group [54, 65]. Sociolects can reflect an intersection of several different identities, such as 'Valley Girl' spoken predominantly by women of a certain age range and geographical location [64]. Within NLP, studies that look at sociolect are often also grouped into dialect or multi-language studies, and there are many applications within socially-aware and low-resource language NLP contexts [82, 121].

In our study, we examined two specific sociolects: AAE and Queer slang, chosen because of the history of biases against their speakers in NLP, their interconnectedness as sociolects, as well as the positionality of many of the authors who identify as speakers of either AAE, Queer slang, or both.

African American English (AAE) is a widely studied sociolect of American English spoken by Black people in the United States (U.S.) [7, 36, 38].[1] AAE has its own distinct grammatical and phonological features that distinguish it from SAE. Despite existing literature on AAE, it is still an underrepresented both in society as well as online spaces, which is associated with the marginalization of Black AAE speakers in predominantly White spaces [1]. Biases against AAE have been widely documented in NLP systems [91, 95] and specifically LLMs [29, 43, 117, 120]. For instance, a recent study [26] found that current LLMs have difficulty both generating and interpreting AAE due to the lack of Black American representation in LLMs. The authors emphasized the potential harm to the AAE-speaking community and called for further development of LLMs that can effectively interact with and understand AAE speakers [26].

Queer slang, unlike AAE, is a relatively less well-documented sociolect which can be defined by words, phrases, or metaphors predominantly used by LGBTQ+ individuals in the U.S. [e.g., in queer or drag spaces; 73, 97].[2] It is closer to SAE, yet could be defined as a minority sociolect due to the its ties with Queer individuals who speak it [16]. Importantly, Queer slang is heavily interconnected with AAE, as Drag and Ballroom culture as well as prominent Black Queer and femme folks (e.g., RuPaul, T.S. Madison) have historically influenced mainstream Queer culture [21, 60, 61]. Recent NLP

[1]Although some refer to the language variety as African American (Vernacular) English (AAVE) or African American Language (AAL), we opt for the African American English (AAE) terminology based on previous work [22, 92].
[2]We use the term Queer slang instead of Gay slang because it encompasses more identities within the LGBTQ+ community. The emergent meaning of the word Queer to be all encompassing of LGBTQ+ individuals can be seen in recent work [114].

work has shown documented biases in LLMs against Queer identities and Queer slang [28, 29]. Similar to AAE, little research explores how systems interpret, generate, and classify Queer slang differently from SAE [30].

2.2 Minority Experiences with LLMs

Language models have shown limitations in serving diverse user groups [10]. Previous studies have highlighted the challenges posed by inherent biases within these systems' training data and processes [52, 79] — LLMs often exhibit undesirable behaviors related to ethical issues, particularly towards minorities based on gender and race, leading to suboptimal user experiences, confusion, and technology abandonment [69, 70, 104]. Additionally, opinion minority groups (e.g., climate change deniers, the alt-right) and educational minority groups tend to have poor user experiences with LLMs [17]. These challenges sparked new regulations and increased public awareness about the associated risks.

Along with linguistic, cultural, gender, and ability biases [10, 32, 63] many studies have noted that AI systems and LLMs frequently fail to account for Black and LGBTQ+ identities or challenges [28, 47, 58, 76, 83]. Mengesha et al. [77] highlights significant performance disparities in AI systems' understanding of African American Language (AAL) are due to their lack of design for the Black experience. Additionally, [72] shared LLMs tend to generate overly generic or insensitive suggestions, such as advising LGBTQ+ users to come out to unsupportive parents, which risks offending and marginalizing these communities.

3 APPROACH: RESEARCH QUESTIONS & VARIABLES STUDIED

In order to observe how sociolect use of LLMs can affect user's conscious and subconscious perceptions of LLMs, we explored both behavioral and perceptual effects, based on previous work. The rationale for each variable will be discussed in each respective section below.

3.1 User Reliance (RQ1)

Our first research question asks: **Does a user's reliance on the agent depend on whether the LLM agent speaks the user's sociolect or not?** *Reliance*, as commonly defined, is the degree to which users use the help of the target LLM in their daily lives. This is a useful metric because it provides insight into how much the user implicitly trusts the LLM outputs on a wider scale [119]. Existing works that measure reliance have also observed that reliance more closely measures how a user depends, understands, and would utilize the system in the future [72, 94]. Based on previous work that showed an increase in reliance due to language-based warmth cues [118], we hypothesize that users will have a greater reliance in the LLM using their sociolect (**H1**).

3.2 User Perception and Pairwise Preference (RQ2)

Our second research question asks: **Does a user's perception and preference vary depending on whether the agent uses sociolect or not?** We specifically consider four perception dimensions for each agent, as well as a pairwise preference judgment, outlined below.

We measure user *trust*, defined as the willingness to be vulnerable based on the expectation of beneficial actions from another party [33, 44]. We hypothesize that the use of sociolects by LLMs enhances user trust (**H2a**), supported by works showing anthropomorphic cues boost trust in AI systems, including conversational agents and LLMs [8, 19, 27, 49, 115].

User *satisfaction* evaluates session experience, encompassing satisfaction with both LLM and user performance [45]. Research suggests that anthropomorphic cues, such as emotional and auditory elements, improve interaction satisfaction [57, 116]. We hypothesize that adapting LLMs to user's sociolects will increase satisfaction (**H2b**).

Our third variable, user *frustration*, is crucial in creativity support research as it tracks cognitive load challenges, encompassing feelings of insecurity, stress, and annoyance [45]. AI tools can mitigate frustration by providing assistance, reducing mental load, and supporting complex tasks [66, 68, 69, 108]. We hypothesize that sociolect-equipped LLMs, by mirroring a conversation partner's language style, lower user frustration and perceived task workload (**H2c**) [74, 100].

Our last perception variable, *social presence* refers to a user's perceived sense of connectedness with the system they interact with [93]. We hypothesize that users feel a stronger social presence with sociolect-using LLMs (**H2d**) , as prior work indicates increased social presence in interactions with AI heightened social and cultural awareness [17, 103].

For pairwise *preference* judgments and *rationales*, users evaluate which LLM agent they prefer—whether sociolect-using or SAE—and provide reasons in free text for their choice. We hypothesize that users will prefer the agent using their sociolect, similar to previous chatbot research adopting linguistic alignment [100] (**H2e**).

3.3 Exploratory Findings (RQ3)

Lastly, our third research question asks: **What perception factors (e.g., trust, frustration, explicit preference) are associated with user reliance?** As an exploratory research question, we hypothesize that there will be associations between perception variables, including explicit user preference, and reliance.

4 METHODS

To investigate how sociolect-specific language models influence user perception and reliance, we conducted a within-subjects experiment with two groups of participants (AAE speakers, Queer slang speakers; §4.1), in which participants performed two sequential, video-based question answering tasks with the help of a simulated LLM – either a sociolect-producing or a control SAE LLM (§4.2).

4.1 Recruitment and Screening Procedure

We recruited participants using Prolific, an online platform tailored for academic research [85]. Prolific allowed us to target specific participant groups based on eligibility criteria. Participants had to be at least 18 years old and reside within the U.S., ensuring a shared cultural exposure to the sociolects under investigation. The study description provided to potential participants on Prolific outlined the procedure: participants would watch two short videos (each under one minute) and evaluate various aspects of their experience.

For the African American English LLM (AAELM) setup, participants were required to self-identify as Black American based on Prolific's demographic question: "Please indicate your ethnicity (i.e., peoples' ethnicity describes their feeling of belonging and attachment to a distinct group of a larger population that shares their ancestry, color, language, or religion)." For the Queer slang LLM (QSLM) setup, participants were required to identify as members of the LGBTQ+ community on Prolific, as indicated by their response to the question: "Do you identify yourself as part of the LGBTQ+ community?" We later ensured their familiarity with the respective sociolect as part of the main study (§4.2).

To ensure fair compensation and value participants' time, we offered an hourly rate of $15, which was clearly communicated during recruitment to maintain transparency and encourage participation. The study protocol, including recruitment and screening procedures, was reviewed and received exempt status by the authors' Institutional Review Board (IRB). This oversight ensured that all participant interactions adhered to the highest ethical standards, particularly concerning privacy, confidentiality, and participant welfare.

4.2 Experimental Procedure

The study employed a within-subjects design for each participant group, as depicted in Figure 1. Participants began by completing a consent form (see Appendix §A). Demographic information, including age, identity, and education level, was collected. Participants were also asked about their familiarity with LLMs, including frequency of use, confidence in understanding, and typical use cases (see Appendix §B).

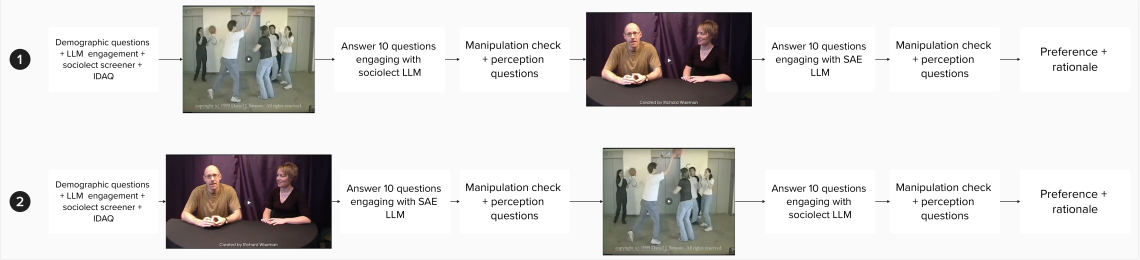


Fig. 1. Study flow diagram illustrating the two experimental scenarios. In both scenarios, participants begin by answering demographic questions, completing a sociolect screener, and the IDAQ. In Scenario 1 (top row), participants first engage with a sociolect LLM after watching a basketball-passing video (Daniel Simons' "Gorillas in Our Midst"), followed by a manipulation check, perception questions, and a preference question. In Scenario 2 (bottom row), participants first engage with a SAE LLM after watching a different video (Richard Wiseman's "Color-Changing Card Trick"), followed by the same sequence of tasks.

To ensure participants were familiar with the sociolect being tested, a sociolect screener was implemented. Participants were screened using a 5-point Likert scale to assess their frequency of sociolect use, confidence in understanding, and likelihood of using example sociolect phrases. Example phrases for the AAELM and QSLM setups were sourced from established corpora and are detailed in the Appendix §C.

Participants who passed the screener completed the Individual Differences in Anthropomorphism Questionnaire (IDAQ) [109] to assess their tendency to anthropomorphize non-human entities (see Appendix §D).

Participants were then shown a snippet from one of two videos designed to test selective attention: Daniel Simons's "Gorillas in Our Midst" [98][3] or Richard Wiseman's "The Colour Changing Card Trick" [110].[4] Selective attention videos captivate participants by presenting detail-rich scenarios that demand focused observation, ensuring engagement and attentiveness [98]. By using videos with highly specific and difficult-to-notice details, participants were placed in a situation in which reliance on the LLM's suggestions became a meaningful decision [84, 102]. This allowed us to evaluate not only the participants' perception of the LLM but also their decision-making when balancing their own knowledge against the LLM's suggestions.

After each video, participants answered questions about specific events while receiving suggestions from their respective LLM agent in SAE or the assigned sociolect. Questions were designed to be difficult, focusing on specific video details (e.g., "What color watch was the woman wearing?"). Question difficulty was pre-assessed by an independent participant pool via Prolific, using a 7-point Likert scale. Only questions rated on average between "very difficult" and "difficult" were selected; an excerpt of a difficulty survey can be found in Appendix §E.

Inspired by Zhou et al. [119], during the video assessment, participants could choose to accept the LLM's suggested responses ("Use LLM's Response") or rely on their own knowledge ("I'll figure it out myself"). Participants were told they would earn points based on their ability to correctly reject incorrect LLM suggestions or rely on accurate ones. Although no points were actually calculated, this mechanism encouraged thoughtful decision-making and discouraged blind reliance on the LLM [9, 119].

The order of LLM assignment (sociolect vs. SAE) and video presentation was counterbalanced to control for order effects. After completing the video task, participants filled out a follow-up questionnaire evaluating *trust* [33], *satisfaction* [105], *frustration* [46], and perceived *social presence* with the LLM [4] (see Appendix §F). Additionally, a manipulation check assessed whether participants perceived the LLM as accurately using the intended sociolect (see Appendix §G for further details). The study concluded with a post-study LLM agent *preference* question, where participants were asked to choose between the sociolect or SAE agent and provide a *rationale* or explanation for their preference.

4.3 Templated LLM Suggestions

We designed three distinct simulated LLM agents to generate responses in SAE and sociolects. The AAELM simulated interactions using AAE, while the QSLM simulated interactions using language and slang prevalent in Queer communities. The SAE LLM (SAELM) served as a baseline or control model, using SAE. Importantly, participants in the experiment did not directly interact with live LLMs but instead engaged with responses pre-generated by LLMs.

Our goal was to create LLM responses that elicited a sense of warmth and medium confidence from participants to observe the effect of sociolect usage on reliance, because previous research has shown that warmth markers can increase user reliance on medium-confidence responses [118].

[3] https://www.youtube.com/watch?v=vJG698U2Mvo
[4] https://www.youtube.com/watch?v=v3iPrBrGSJM

SAE Suggestions. Following the procedure in Zhou et al. [118], we obtained a set of SAE warmth phrases from LLMs by prompting popular LLMs to respond to factual questions with warmth prefixes (e.g., "I'd be happy to help"). We merged these LLM-generated warmth prefixes with the confidence prefixes with reliability percentages between 30% and 70% (e.g., "I would say it's"), drawn from Zhou et al. [119]. We compiled a list of 20 warmth phrases and 14 confidence expressions (see Appendix §H), resulting in 280 unique combinations.

AAE Translation with In-Context Learning. We translated the 280 unique SAE suggestions into AAE using GPT-4's in-context learning capabilities. This approach aligns with recent advancements in evaluating sociolect-specific text generation [2, 15, 18, 50]. For in-context learning, we sourced AAE examples from CORAAL [53], supplemented with AAE tweet pairs [39] and SAE-to-AAE translations [26, 59]. The detailed prompt design and implementation process for the in-context learning methodology are described in Appendix §I.

After translating the 280 warmth-epistemic combinations into AAE, we used Prolific surveys to verify authenticity, asking participants to identify which phrases aligned most closely with AAE (see Appendix §J). We presented groups of 20 phrases to batches of 5 participants at a time and participants were tasked with identifying which sociolect they felt best aligned with the translated phrases. Only phrases that were deemed to represent AAE by at least 4 out of the 5 verifiers were selected for inclusion as AAELM suggestions. This process was repeated until we achieved a diverse selection of 10 sociolect aligned phrases with unique confidence expressions (see Appendix §K for final list of AAE translations and corresponding SAE translations.)

Queer Slang Translation with Persona-Based Prompting. Since no existing corpora provide modern Queer slang or SAE-to-Queer slang translations, we employed persona prompting with GPT-4 to translate SAE suggestions into Queer slang. Persona prompting, where an LLM is guided to emulate the speaking style of specific individuals, has been demonstrated to effectively replicate distinctive linguistic styles [11, 48]. For this study, we selected personas associated with well-known Queer celebrities—RuPaul, Trixie & Katya, and T.S. Madison—each known for their use of Queer slang [21, 35, 51]. Using this approach, we applied the 14 SAE warmth phrases (see Table 3) to each persona output. Through persona prompting, we generated 42 unique Queer warmth phrases (see Appendix §M). This resulted in 420 unique combinations of Queer warmth phrases.

Using a similar verification procedure as for AAE (see Appendix §4.3), we validated Queer slang translations via sociolect verification surveys on Prolific (see Appendix §N). Phrases identified by at least 4 out of 5 participants were included as QSLM suggestions (see Appendix §O). Corresponding SAE translations were paired with the same questions as their sociolect counterparts (see Appendix §P).

4.4 Participant Statistics and Demographics

Out of 1,007 responses, 985 participants successfully passed the sociolect screener, with 498 in the AAELM setup and 487 in the QSLM setup. Participants in the AAELM group ranged from 18 to 75 years of age, with 207 identifying as men, 186 as women, and 5 as other. Participants in the QSLM group ranged from 18 to 72 years of age, with 169 identifying as men, 190 as women, and 47 as other. (See Appendix §Q).

5 FINDINGS

To better understand the impact of sociolect adaptation in LLM interactions, we examined participant reliance, perception, and preferences through a combination of quantitative and qualitative analyses.

5.1 RQ1: Does reliance on the agent depend on whether the agent speaks the user's sociolect or not?

To test our first hypothesis (H1: Sociolect usage by LLM leads to higher reliance), we compared reliance using t-tests.

Our analyses showed mixed results. Among AAE-speaking participants, reliance significantly varied depending on whether the agent spoke AAE or SAE ($d = 0.12$, $p < 0.05$; Fig. 2b, Table §13), supporting our hypothesis. For Queer slang speakers, reliance showed a marginal difference depending on whether the agent used Queer slang or SAE ($d = 0.09$, $p = 0.08$; Fig. 2b, Table §13), but this result did not reach statistical significance, and thus, did not support our hypothesis.

5.2 RQ2: Does a user's perception vary depending on whether the agent depend on sociolect usage or not? (across trust, satisfaction, frustration, social perceived presence, and pairwise preference judgment)

To test our second hypothesis (H2: Sociolect usage by LLM leads to higher trust (H2a), user satisfaction (H2b), lower user frustration (H2c), a higher sense of social presence (H2d), and paired preference for the sociolect-using LLM more than the SAELM (H2e), we conducted t-test comparisons between the relevant variables reported for SAELM and AAELM.

For AAE speakers, the SAELM consistently was higher than the AAELM across key perception variables. AAE participants trusted the SAELM more than the AAELM ($d = 0.19$, $p < 0.05$; Table §15), found it more satisfying ($d = 0.2$, $p < 0.05$; Table §15), and experienced less frustration with it ($d = 0.21$, $p < 0.05$; Table §15). Interestingly, no significant difference was observed in the perceived social presence between the two agents ($d = -0.05$, $p > 0.1$; Table §15).

Unlike AAE participants, Queer slang participants did not demonstrate a significant difference in trust between the Queer slang agent and the SAELM ($d = -0.03$, $p > 0.1$; Table §15). However, Queer slang participants did perceive significantly greater social presence in the Queer slang agent compared to the SAELM ($d = -0.23$, $p < 0.05$; Table §15). Despite this, satisfaction levels were similar between the two agents ($d = 0.07$, $p > 0.1$; Table §15), and participants found the SAELM significantly less frustrating than the Queer slang agent ($d = 0.245$, $p < 0.05$; Table §15).

Examining pairwise perception, we find that AAE participants demonstrated a clear preference for one agent over the other ($d = 0.44$, $p < 0.05$; Fig. 3a, Table §14). In contrast, Queer slang participants showed no significant preference between the two agents ($d = 0.009$, $p > 0.1$; Fig. 3b, Table §14).

5.3 RQ3: Are there other factors that influence user reliance?

5.3.1 Perception variables. To test our third hypothesis (H3: There will be associations between perception variables, including explicit user preference, and reliance), we analyzed the relationship between user reliance and our four perception variables, as well as users' preferred agent.

Using Pearson correlation coefficients, we observed significant, albeit weak, correlations between reliance and perception variables for both AAE and Queer slang participants, as shown in Table 16 and Table 17. These results support our hypothesis (H3), demonstrating that perception variables such as trust, satisfaction, lack of frustration and social presence are related to reliance on the sociolect LLM. We also confirm our hypothesis that user's have an increased reliance on the LLM agent they explicitly preferred than the reliance on the agent that participants dispreferred.

For AAE participants, trust, satisfaction, and lack of frustration significantly correlated with reliance on both the sociolect and SAE agents, with the strongest correlation observed for satisfaction ($r = 0.288$, Table 16) for AAE reliance and for SAE reliance ($r = 0.377$, Table 16). Social presence was also significantly correlated with reliance on the AAE agent ($r = 0.188$ Table 16) but not with the SAE agent, suggesting that sociolect alignment slightly enhances the perception of social presence in this context.

For Queer slang participants, the correlations were slightly stronger overall. Trust, social presence, and satisfaction all significantly correlated with reliance on both the sociolect and SAE agents, with satisfaction showing the highest correlation for both the sociolects ($r = 0.438$; Table 17) and SAE agents ($r = 0.441$; Table 17). Social presence correlated more strongly with reliance on the Queer slang agent ($r = 0.264$; Table 17) than with the SAE agent ($r = 0.133$; Table 17), further emphasizing the role of sociolect alignment in fostering a sense of engagement.

We observed a significant increase in reliance on the LLM agent that users identified as their preferred choice. This was true for both AAE speakers ($p = 0.000325, d = -0.204427$; Table 18) as well as Queer slang speakers ($p = 0.000008, d = -0.207088$; Table 18).

5.4 What are people's open ended perceptions of the sociolect agent?

5.4.1 Open Coding Process. To better understand participants' perceptions of each simulated LLM and gain insights into their reasoning, we conducted a qualitative analysis of open-ended comments from the AAE and Queer slang participants. Using a thematic analysis approach [14], three authors began by coding comments from a pilot run of the study. This initial coding process involved analyzing 14 comments from each setup to identify common themes and resolve differing interpretations. Two additional rounds of coding followed, covering 15 and 18 comments from each pilot setup, which led to the creation of two preliminary codebooks—one for each setup. These codebooks were then applied to analyze participant comments from the main experimental study. After coding all study data, the authors grouped related codes from each sociolect group into thematic categories, iteratively refining the structure to create a coherent framework. This structured approach allowed us to uncover and organize the narrative emerging from the raw data, which we discuss below. The codebook, including definitions and examples, is provided in the Appendix §T.

The top three codes for Queer slang participants describing QSLMs are "positive emotions" (156 counts, 32.0% of the generated codes), "resonates" (45 counts, 9.24%), and "negative emotions" (40 counts, 8.2%), while describing SAELM with codes such as "formality" (60 counts, 12.3%), "comprehensible" (36 counts, 7.39%), and "bland" (28 counts, 5.75%). Then, for AAE participants, their top three codes for AAELM are "resonates" (44 counts, 8.83%), "anthropomorphization" (22 counts, 4.41%), and "positive emotions" (21 counts, 4.21%), while describing SAELM as "comprehensible" (78 counts, 15.6%), "formality" (58 counts, 11.6%), and "reliable" (27 counts, 5.42%).

5.4.2 Positive Perceptions of the Sociolect LLMs. Participants who interacted with the sociolect-specific language models (AAELM and QSLM) often associated them with feelings of enjoyment, casualness, and pride. Among AAE participants, just under 9% of comments were tagged with "resonates," and slightly over 4% were tagged with "positive emotions." Similarly, Queer participants found their sociolect LM engaging and pleasing, with 32% of comments tagged with "positive emotions" and 9.24% tagged with "resonates." These findings highlight the potential of sociolect adaptation to create meaningful and enjoyable interactions. AAE participants described the AAELM with phrases like, "*It just sounds more fun to interact with,*" while Queer slang participants reflected on the QSLM, stating, "*I enjoy being called diva!*" These responses demonstrate how sociolect adaptation can enhance engagement and relatability.

Participants highlighted that the sociolect agent fostered a more approachable and human-like interaction, enhancing the perception of social presence. This sentiment was evident in 7.8% of comments from Queer slang participants and 4.4% of comments from AAE participants, which were tagged as "anthropomorphization." Many participants noted that the sociolect LLMs' personable and conversational tone contributed to their perception as more human-like. For example, AAE participants shared sentiments such as, "*It felt very relatable*" and "*Her answers felt more like me talking,*" while Queer participants echoed similar feelings, with one stating, "*...I prefer Agent [QSLM] more because I have a*

personal connection with him." This aligns with the 2.61% of AAE comments tagged as "comprehensible," with one participant noting, "*I understand Agent [AAELM] more.*"

5.4.3 Negative Perceptions of the Sociolect LLMs. These positive sentiments were contrasted by a notable subset of participants who perceived the sociolect LLM's use of sociolect as forced or artificial. Among Queer participants, 8.2% of comments were tagged with "negative emotions," 6.98% with "exaggerated usage," and 5.33% with "disrespectful." Some participants expressed frustration at the QSLM's overuse of slang, with one noting, "*Even people who use LGBTQ slang don't talk like that constantly. It would be annoying to have an AI constantly use slang phrases.*" Similarly, AAE participants criticized the AAELM, describing its language as unnatural and inauthentic. Comments such as "*Agent [AAELM] using AAE sounds like a joke and not natural*" were reflected in 3.41% of comments tagged as "unnatural," 2.61% tagged as "unserious," and 1.4% tagged as "disrespectful." Participants also raised concerns about stereotyping and cultural insensitivity. AAE participants expressed discomfort, perceiving the AAELM's use of AAE as stereotyping or mocking Black culture. One respondent noted, "*I don't like the idea of AI using such language, especially if it is being programmed by someone who isn't Black.*" Similarly, Queer slang participants criticized the QSLM, feeling that its use of Queer slang bordered on mockery, with comments such as, "*The [QSLM] seems to be making a mockery of lgbt [sic].*" These reactions highlight the risks of sociolect adaptation, especially when it involves sensitive cultural or linguistic contexts, potentially alienating users if the sociolect is perceived as inauthentic or disrespectful.

Participants also highlighted the importance of context in the appropriateness of sociolect usage. Among Queer slang participants, 2.9% of comments were tagged with "context," with one participant noting, "*While Queer slang is perfectly fine, like any slang, there's a time and a place to use it. This situation doesn't feel like it.*" AAE participants frequently associated AAE with perceptions of "improper English," as reflected in 2.6% of comments tagged as "unserious" and 1.6% tagged as "improper." One participant commented, "*Agent [AAELM] sounds like an intelligent being, whereas Agent [QSLM] sounds like they don't know how to speak proper English.*" The findings suggest that sociolect adaptation in AI must be carefully tailored to avoid reinforcing stereotypes or alienating users by misaligning with situational norms.

Many Queer slang participants did not associate the use of Queer slang with the LLM's intended role, as 3.7% of their comments were tagged with "not normal," and 2.87% were tagged with "expectation misalignment." One participant remarked, "*When using AI, I do not need it to sound like a human being,*" highlighting a disconnection between user expectations and the agent's sociolect adaptation. Furthermore, Queer slang participants noted that the QSLM's use of slang could create barriers for those unfamiliar with the terms, reducing accessibility, as 2.9% of comments were tagged with "unfamiliar". This highlights the potential for sociolect adaptation to unintentionally exclude or confuse users, especially in professional or formal contexts.

5.4.4 Positive Perceptions of the SAE-using LLMs. The SAELM was consistently praised for its clarity and professionalism, reflecting participants' alignment with standard language norms and expectations for formal interactions. Many participants emphasized its formal tone as being more appropriate for professional or factual scenarios. Among Queer participants, 12.3% of comments were tagged with "formality," while 11.6% of AAE participant comments were categorized similarly. Additionally, 4.9% of Queer participant comments and 0.8% of AAE participant comments were tagged as "task-oriented," reflecting the perception that the SAELM was better suited for goal-directed interactions. The SAELM was also viewed as more dependable and easier to understand, with 5.42% of AAE participant comments tagging it as "reliable," compared to just 2.0% for the AAELM. Furthermore, 15.6% of AAE participant comments and 7.4% of Queer slang participant comments tagged the SAELM as "comprehensible," highlighting its perceived clarity and accessibility across user groups. Comments such as "*Agent [SAELM] is brief and therefore concise and precise,*" "*I think*

the conversation would be more productive," and "Agent [SAELM] sounds more professional and coherent" underscore the widespread perception of SAE as the default for effective and formal communication. These findings highlight how adherence to standard language norms can influence perceptions of professionalism and reliability in AI systems.

AAE participants often emphasized that SAELM met their expectations for how an AI should behave. Statements like "*Agent [SAELM] sounds normal and like an actual AI*" underscore this sentiment, with 3.8% of AAE comments tagged as "normal." These responses suggest that SAE's alignment with participants' expectations of AI fosters a sense of credibility and trustworthiness. While both participant groups frequently perceived SAELM as more relatable and human-like, some responses also indicated that it could connect on a personal level. For example, 1.4% of AAE participant comments were tagged as "resonates." Remarks such as "*I have a personal connection with him*" reveal that SAE, while perceived as formal and task-oriented, can also foster a degree of social presence. Participants further described SAELM as proactive and motivated, with one noting, "*Agent [SAELM] expressed enthusiasm and willingness to help,*" as 1.2% of AAE comments were tagged as "enthusiastic."

Open Coding Analysis Connection to Quantitative Analysis. Our qualitative findings align closely with trends observed in the quantitative analysis, offering deeper insight into participant perceptions. For AAE participants, comments indicated greater reliance on the SAELM compared to the AAELM, mirroring the significant quantitative differences in trust and preference between the two agents. Perceptions of the AAELM as unprofessional, contextually inappropriate, or even comical corresponded to the agent's lower trust, reliance, and preference scores in the quantitative analysis.

Interestingly, while some comments suggested a potential for the AAELM to enhance social presence through its use of AAE, this was not reflected in the quantitative ratings, which showed no significant difference in social presence between the agents. This discrepancy highlights the nuanced nature of sociolect-based interactions, where subjective impressions may not always align with measured outcomes.

The discomfort expressed by participants regarding the AAELM's use of AAE, often described as forced or inappropriate, aligns with the quantitative findings of higher frustration and lower satisfaction with the agent. These results underscore the importance of authenticity and contextual sensitivity in sociolect adaptation.

For Queer slang participants, the high number of comments tagged with "positive emotions" and "resonates" supports the significant increase in perceived social presence for the QSLM. These qualitative insights reinforce the idea that sociolect adaptation, when aligned with user expectations, can foster a sense of connection and relatability, particularly in contexts that prioritize warmth and social engagement.

6 DISCUSSION

Through a mixed-methods study combining quantitative and qualitative approaches, we investigated two sociolect-specific agents—AAELM and QSLM—representing AAE and Queer slang, alongside a SAE baseline. We found that AAE speakers rely on and prefer an LLM that communicates in SAE rather than AAE, primarily due to the LLM's inability to use AAE in a natural or respectful manner. In contrast, Queer slang participants did not exhibit a strong preference for either the SAELM or QSLM, with user reliance on the LLM primarily correlating with increased social presence and reduced frustration when interacting with the QSLM.

Below we discuss the implications of our results" and start discussing key takeaways from our variable findings.

Expectation of Standard English. Our major takeaway was that AAE speakers overall relied on and preferred the SAE agent over the AAE agent. This is noteworthy as it shows personalization and anthropomorphic design of agents in

this scenario can hinder not only user perception of an LLM, but also directly negatively affect the way they use it. These findings are further reinforced by our qualitative feedback, with AAE users describing the AAE agent's responses as unnatural, abnormal, and disrespectful. Additionally, existing research highlights that over-anthropomorphized agents are often perceived as uncanny or even mocking [18, 28, 32]. A possible explanation for AAE speaker greater reliance and preference for LLMs without anthropomorphic cues is the perception that mimicking AAE encroaches upon Black cultural spaces [38]. This idea is deeply tied to identity, where the use of a sociolect is often linked to the ability to self-identify with the associated cultural or social group [88]. Since an LLM is not a person, it cannot meet this prerequisite. The issue, therefore, is not that the LLM appears "too human," but that it is perceived as overstepping societal boundaries and appropriating a linguistic identity it cannot authentically claim. We suggest integrating user-centered feedback to allow for real-time refinements in sociolect representation, ensuring alignment with user expectations and minimizing the risk of perceived inauthenticity or offensiveness.

Similarly, users appear accustomed to the dominant SAE language and behaviors embedded in digital technologies, shaping their preconceived expectations of LLMs. SAE as the "standard" not only reflects a specific linguistic region but also aligns with the demographic default of cisgender, straight, white males [3, 43, 81]. This aligns with research showing how the "masculine default" is often associated with official and correct speech [113]. Parallel studies on trust evaluation further reveal how user perceptions of trust in LLMs are influenced by scenarios involving epistemic markers of uncertainty and informal language, highlighting how extensive interactions shape these evaluations [55]. Our work underscores the importance of critically examining how dominant language ideologies and subordination are embedded within LLM design choices [24, 25, 99]. Recognizing these dynamics can help address disparities in language representation and reduce the subordination of non-standard sociolects.

In professional contexts, participants overwhelmingly commented preference for the SAELM due to its clarity, professionalism, and alignment with expectations for AI behavior. Its formal tone and comprehensibility made it the favored choice for task-oriented interactions, reinforcing societal norms that associate SAE with credibility and competence. By contrast, sociolect-adaptive agents elicited mixed reactions. While some participants valued the warmth, relatability, and cultural representation these agents offered—particularly in casual settings—others perceived their sociolect usage as unnatural, inauthentic, or even offensive. These findings underscore the societal privileging of SAE as a marker of professionalism and the normative biases present in AI evaluations [5, 37, 99]. The alignment of participants with SAE as a "default" AI language reveals how these biases disadvantage diverse sociolects in technological interactions. Our findings suggest that the effectiveness of anthropomorphic cues heavily depends on context, which plays a crucial role in user acceptance and the perceived social presence of the LLM. To navigate these challenges, we recommend a cautious and context-aware approach to implementing anthropomorphic features. Specifically, dynamic language adaptation should be employed to allow AI systems to adjust language styles seamlessly based on interaction contexts and goals. For instance, AI agents could transition between sociolects and formal tones to meet the demands of professional or casual settings effectively.

Reliance Influenced by Perception and Preference. While reliance is correlated with preference and perception, our study demonstrates that these measures are not directly correlated. Notably, variability exists between variables such as reliance, preference, and social proximity, particularly for AAE speakers. For instance, while AAE speakers may prefer and rely on an SAE agent, this does not necessarily correspond to a closer sense of social proximity. These findings highlight the importance of measuring behavioral responses rather than relying solely on self-reported preferences and perceptions [119].

Current studies in LLM design have focused solely on perceptions [31, 107], this approach captures only part of the picture. Additionally, the dynamics between perception and behavior can become more nuanced and complex during extended use of LLMs [68]. As studies often rely on self-reported perceptions of LLMs [8, 17, 19, 27, 49, 103, 115], users' behaviors—such as their preferences and reliances—may tell a different story. Our work highlights the importance of capturing users' preferences, perceptions, and behaviors comprehensively. We recommend robust design methodologies that account for these variables to better understand the nuanced relationships between them and provide a more holistic view of user interaction with LLMs.

Differences between AAE and Queer slang. Our next major finding was that not all agents using a sociolect are relied on to the same degree. When comparing cross-study reliance rates between two parallel studies—one focused on AAE sociolect and the other on Queer slang—it is evident that the strong preference and reliance AAE users displayed for the SAE agent over their sociolect agent is not mirrored by Queer slang users. Queer slang participant sentiments further reflect this distinction, as Queer slang users provided nearly four times as many comments expressing positivity and nearly twice as many comments indicating resonance with the LLM.

This can be attributed to a few things. For one, AAE is often learned as a first-language, and AAE speakers often grow up speaking AAE with their family and community [41]. Queer slang, in contrast, is usually acquired at a later stage in life, since most Queer people do not grow up speaking Queer slang [13]. Also, AAE is more standardized and consistent across generations [87] as opposed to Queer slang that varies heavily by age [71, 101]. On the other hand, queer slang is heavily influenced by AAE, so it is also influenced by the intersectionality of Black Queer individuals [12, 20].

The main takeaway for sociolect variation is that anthropomorphic cues in LLMs cannot be applied universally, as sociolects are deeply tied to distinct cultural dynamics and carry unique nuances that shape how users perceive their usage by LLM systems. Effective sociolect adaptation requires careful consideration to avoid the risk of infringement or perceived encroachment on the cultural and social identities associated with these linguistic styles. For instance, many AAE participants expressed discomfort with the LLM's use of anthropomorphic cues, describing them as unnatural, unserious, and disrespectful. These findings underscore the need for further research and thoughtful design approaches to ensure LLMs can adopt sociolects in a way that respects and acknowledges the communities they represent.

7 LIMITATIONS AND FUTURE WORK

This study has a few notable limitations that are important to acknowledge. In our experimental task, participants were asked to answer questions about the videos in SAE, while the sociolect agent provided responses in their respective sociolect. This study design choice was made to ensure consistency in the phrasing of questions across all experimental groups. By keeping the questions in SAE for all participants, we minimized the risk of mistranslation or misinterpretation that could arise if questions were presented in different sociolects. This approach allowed us to maintain a uniform standard for comparison, ensuring that any differences in responses were attributable to the agents' language usage rather than variations in how the questions were framed. This mismatch in linguistic presentation could have disrupted the natural flow of the interaction however, as participants were required to code-switch between SAE and their sociolect. Such code-switching may act as a barrier, reducing the naturalness and authenticity of the task presentation. Future research should aim to align the sociolect of the questions with the agent's responses to maintain consistency and enhance the naturalness of the experimental setup.

In our experimental setup, participants interacted with pre-generated suggestions provided by the LLM, where the correct answers to the questions were intentionally withheld. This design choice was made to prevent participants from reasoning about the plausibility of the suggested answers, focusing instead on the perceived quality of the LLM's engagement [118]. While this setup effectively isolates the LLM's ability to present sociolect-specific responses, it inherently limits the interaction to single-turn exchanges, which do not reflect the multi-turn conversational dynamics typically associated with LLM usage. The single-turn suggestion structure restricts participants from engaging in back-and-forth interaction with conversational agents, as well as avoiding the influence of answer accuracy on participant judgments. Future research should address these limitations by exploring more dynamic, multi-turn interactions that better reflect real-world conversational use cases of sociolect-infused LLMs.

Additionally, future studies could examine how presenting actual responses with varying degrees of accuracy impacts participants' reliance on and trust in LLMs. This approach would disentangle the effects of answer accuracy from the broader evaluation of sociolect-specific engagement, offering a deeper understanding of how interaction design shapes user expectations and trust in LLMs.

To ensure high-quality examples of sociolect usage by the LLM agent, we prescreened the generated responses in both the AAE and the Queer slang studies. Despite our robust efforts, including crowdsourcing rankings to select the most authentic-sounding phrases, there remains the challenge of some phrases sounding artificial to participants. Given the diverse variations of AAE and Queer slang across the U.S. [67, 89, 112], it is possible that participants encountered unfamiliar sociolect variations that may have seemed inauthentic, even though these variations are valid. Future research should aim to account for regional sociolect variations and align them with participants' linguistic backgrounds to enhance the authenticity and relatability of the LLM agent's use of sociolects.

Focusing exclusively on Black American users of AAE limits our scope to understanding this demographic, despite the fact that AAE is not restricted to usage by Black Americans [89]. This approach overlooks important discussions around the use of stigmatized language by individuals outside this demographic. Future research should expand the participant pool to include a broader range of demographics, with a stronger emphasis on screening for frequent AAE usage rather than relying solely on participants' racial or ethnic background. This would provide a more nuanced understanding of AAE usage and its perception across diverse groups.

Our study is limited in its exploration of the intersectionality of sociolect speakers. To ensure the integrity of the experiment, we prevented participants from completing both the AAE and Queer slang versions, restricting our ability to gather insights from individuals who engage with both sociolects in their daily lives, such as Black American members of the LGBTQ+ community. This approach overlooks the unique perspectives that intersectional identities bring to the use and perception of sociolects. Future research should consider experimental formats that allow participants to engage with multiple sociolects they identify with, providing deeper insights into how intersectionality influences their behavior toward and perception of sociolect agents.

REFERENCES

[1] H. Samy Alim, John R. Rickford, and Arnetha F. Ball. 2016. *Raciolinguistics: How Language Shapes Our Ideas About Race.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190625696.001.0001

[2] Marcellus Amadeus, José Roberto Homeli da Silva, and Joao Victor Pessoa Rocha. 2024. Bridging the Language Gap: Integrating Language Variations into Conversational AI Agents for Enhanced User Engagement. In *Proceedings of the 1st Worskhop on Towards Ethical and Inclusive Conversational AI: Language Attitudes, Linguistic Diversity, and Language Rights (TEICAI 2024).* 16–20.

[3] Celeste B Amos. 2019. Understanding correlations between standard american english education and systemic racism and strategies to break the cycle. *Int J Res Humanit Soc Stud* 6 (2019), 1–4.

[4] C.E. Anderson. 1977. Ship design and safety in a changing economic climate. *Marine Policy* 1, 2 (1977), 123–134. https://doi.org/10.1016/0308-5961(77)90016-7

[5] Kate T. Anderson, Chris Chang-Bacon, and Maria Guzmán Antelo. 2024. Navigating monolingual language ideologies: Educators' "Yes, BUT" objections to linguistically sustaining pedagogies in the classroom. *International Journal of Bilingualism* 28, 4 (Aug. 2024), 618–634. https://doi.org/10.1177/13670069241236682 Publisher: SAGE Publications Ltd.

[6] Sunil Arora, Sahil Arora, and John Hastings. 2024. The Psychological Impacts of Algorithmic and AI-Driven Social Media on Teenagers: A Call to Action. In *2024 IEEE Digital Platforms and Societal Harms (DPSH)*. IEEE, 1–7.

[7] April Baker-Bell. 2020. *Linguistic justice: Black language, literacy, identity, and pedagogy.* Routledge.

[8] Nikola Banovic, Zhuoran Yang, Aditya Ramesh, and Alice Liu. 2023. Being trustworthy is not enough: How untrustworthy artificial intelligence (AI) can deceive the end-users and gain their trust. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–17.

[9] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. In *AAAI Conference on Human Computation & Crowdsourcing*. https://api.semanticscholar.org/CorpusID:201685074

[10] Gábor Bella, Paula Helm, Gertraud Koch, and Fausto Giunchiglia. 2024. Tackling Language Modelling Bias in Support of Linguistic Diversity. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 562–572.

[11] Karim Benharrak, Tim Zindulka, Florian Lehmann, Hendrik Heuer, and Daniel Buschek. 2024. Writer-Defined AI Personas for On-Demand Feedback Generation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1049, 18 pages. https://doi.org/10.1145/3613904.3642406

[12] TOMÁŠ BOČEK. 2023. The Impact and Influence of African-American Vernacular English on LGBTQ+ Culture (2023).

[13] Kasandra Brabaw. 2024. 17 Lesbian Slang Terms Every Baby Gay Needs To Learn. *Refinery29* (2024). https://www.refinery29.com/en-us/lesbian-slang-terms-definitions

[14] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa.

[15] William Cain. 2024. Prompting change: exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends* 68, 1 (2024), 47–57.

[16] Logan S Casey, Sari L Reisner, Mary G Findling, Robert J Blendon, John M Benson, Justin M Sayde, and Carolyn Miller. 2019. Discrimination in the United States: Experiences of lesbian, gay, bisexual, transgender, and queer Americans. *Health services research* 54 (2019), 1454–1466.

[17] Kaiping Chen, Anqi Shao, Jirayu Burapacheep, and Yixuan Li. 2024. Conversational ai and equity through assessing gpt-3's communication with diverse social groups on contentious topics. *Scientific Reports* 14, 1 (2024), 1561.

[18] Myra Cheng, Esin Durmus, and Dan Jurafsky. 2023. Marked personas: Using natural language prompts to measure stereotypes in language models. *arXiv preprint arXiv:2305.18189* (2023).

[19] Michelle Cohn, Mahima Pushkarna, Gbolahan O Olanubi, Joseph M Moran, Daniel Padgett, Zion Mengesha, and Courtney Heldreth. 2024. Believing Anthropomorphism: Examining the Role of Anthropomorphic Cues on Trust in Large Language Models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[20] Brianna R Cornelius. 2016. Gay Black men and the construction of identity via linguistic repertoires. In *Proceedings of the 24th Annual Symposium about Language and Society-Austin*.

[21] Mo Crutzen. 2021. "Yass girl. Spill the tea. Throw some shade." Translation procedures used in the translation of drag and gay vocabulary in RuPaul's Drag Race via subtitling. (2021).

[22] Jamell Dacon. 2022. Towards a Deep Multi-layered Dialectal Language Analysis: A Case Study of African-American English. In *Proceedings of the Second Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, Su Lin Blodgett, Hal Daumé III, Michael Madaio, Ani Nenkova, Brendan O'Connor, Hanna Wallach, and Qian Yang (Eds.). Association for Computational Linguistics, Seattle, Washington, 55–63. https://doi.org/10.18653/v1/2022.hcinlp-1.8

[23] Aida Davani, Mark Diaz, Dylan Baker, and Vinodkumar Prabhakaran. 2024. Disentangling Perceptions of Offensiveness: Cultural and Moral Correlates. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 2007–2021.

[24] Bethany Davila. 2016. The inevitability of "standard" English: Discursive constructions of standard language ideologies. *Written Communication* 33, 2 (2016), 127–148.

[25] Roberto Santiago de Roock. 2024. To Become an Object Among Objects: Generative Artificial "Intelligence," Writing, and Linguistic White Supremacy. *Reading Research Quarterly* 59, 4 (2024), 590–608. https://doi.org/10.1002/rrq.569 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rrq.569.

[26] Nicholas Deas, Jessi Grieser, Shana Kleiner, Desmond Patton, Elsbeth Turcan, and Kathleen McKeown. 2023. Evaluation of African American language bias in natural language generation. *arXiv preprint arXiv:2305.14291* (2023).

[27] Ameet Deshpande, Tanmay Rajpurohit, Karthik Narasimhan, and Ashwin Kalyan. 2023. Anthropomorphization of AI: opportunities and risks. *arXiv preprint arXiv:2305.14784* (2023).

[28] Harnoor Dhingra, Preetiha Jayashanker, Sayali Moghe, and Emma Strubell. 2023. Queer people are people first: Deconstructing sexual identity stereotypes in large language models. *arXiv preprint arXiv:2307.00101* (2023).

[29] Jesse Dodge, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. 2021. Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for

Computational Linguistics, Online and Punta Cana, Dominican Republic, 1286–1305. https://doi.org/10.18653/v1/2021.emnlp-main.98

[30] Virginia K Felkner, Ho-Chun Herbert Chang, Eugene Jang, and Jonathan May. 2023. Winoqueer: A community-in-the-loop benchmark for anti-lgbtq+ bias in large language models. *arXiv preprint arXiv:2306.15087* (2023).

[31] Sarah E Finch, Ellie S Paek, Sejung Kwon, Ikseon Choi, Jessica Wells, Rasheeta Chandler, and Jinho D Choi. 2025. Finding A Voice: Evaluating African American Dialect Generation for Chatbot Technology. *arXiv preprint arXiv:2501.03441* (2025).

[32] Vinitha Gadiraju, Shaun Kane, Sunipa Dev, Alex Taylor, Ding Wang, Emily Denton, and Robin Brewer. 2023. " I wouldn't say offensive but...": Disability-Centered Perspectives on Large Language Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 205–216.

[33] David Gefen, Elena Karahanna, and Detmar W. Straub. 2003. Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly* 27, 1 (2003), 51–90. https://doi.org/10.2307/30036519

[34] Aashish Ghimire. 2024. Generative AI in Education From the Perspective of Students, Educators, and Administrators. (2024).

[35] Julian Kevon Glover. 2016. Redefining realness?: On Janet Mock, Laverne Cox, TS Madison, and the representation of transgender women of color in media. *Souls* 18, 2-4 (2016), 338–357.

[36] Lisa J Green. 2002. *African American English: a linguistic introduction*. Cambridge University Press.

[37] Sr. Gresczyk, Richard A. 2011. *Language warriors; Leaders in the Ojibwe language revitalization movement*. Ph. D. Dissertation. http://ezproxy.cul.columbia.edu/login?url=https://www.proquest.com/dissertations-theses/language-warriors-leaders-ojibwe-revitalization/docview/865313844/se-2 Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-08-04.

[38] Jessica A Grieser. 2022. *The Black side of the river: Race, language, and belonging in Washington, DC*. Georgetown University Press.

[39] Sophie Groenwold, Lily Ou, Aesha Parekh, Samhita Honnavalli, Sharon Levy, Diba Mirza, and William Yang Wang. 2020. Investigating African-American Vernacular English in transformer-based text generation. *arXiv preprint arXiv:2010.02510* (2020).

[40] Rishav Hada, Safiya Husain, Varun Gumma, Harshita Diddee, Aditya Yadavalli, Agrima Seth, Nidhi Kulkarni, Ujwal Gadiraju, Aditya Vashistha, Vivek Seshadri, et al. 2024. Akal Badi ya Bias: An Exploratory Study of Gender Bias in Hindi Language Technology. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1926–1939.

[41] Megan-Brette Hamilton. 2020. Real life lessons in language and literacy: Discovering the role of identity for AAE-speakers. *Journal of Language and Literacy, UGA. http://jolle. coe. uga. edu/wp-content/uploads/2020/02/SSO_January_Hamilton. pdf* (2020).

[42] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication* 25, 1 (2020), 89–100.

[43] Camille Harris, Matan Halevy, Ayanna Howard, Amy Bruckman, and Diyi Yang. 2022. Exploring the role of grammar and word choice in bias toward african american english (aae) in hate speech classification. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 789–798.

[44] D. Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The Journal of Strategic Information Systems* 11, 3 (2002), 297–323. https://doi.org/10.1016/S0963-8687(02)00020-3

[45] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload* 1, 3 (1988), 139–183.

[46] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (1988), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[47] Valentin Hofmann, Pratyusha Ria Kalluri, Dan Jurafsky, and Sharese King. 2024. AI generates covertly racist decisions about people based on their dialect. *Nature* 633, 8028 (2024), 147–154.

[48] Jonathan Ivey, Shivani Kumar, Jiayu Liu, Hua Shen, Sushrita Rakshit, Rohan Raju, Haotian Zhang, Aparna Ananthasubramaniam, Junghwan Kim, Bowen Yi, Dustin Wright, Abraham Israeli, Anders Giovanni Møller, Lechen Zhang, and David Jurgens. 2024. Real or Robotic? Assessing Whether LLMs Accurately Simulate Qualities of Human Responses in Dialogue. arXiv:2409.08330 [cs.CL] https://arxiv.org/abs/2409.08330

[49] Yi Jiang, Xiangcheng Yang, and Tianqi Zheng. 2023. Make chatbots more adaptive: Dual pathways linking human-like cues and tailored response to trust in interactions with chatbots. *Computers in Human Behavior* 138 (2023), 107485.

[50] Anna Jørgensen, Dirk Hovy, and Anders Søgaard. 2015. Challenges of studying and processing dialects in social media. In *Proceedings of the workshop on noisy user-generated text*. 9–18.

[51] Bagas Dwi Kameswara and Agustinus Hary Setyawan. 2024. HEARER-ORIENTED ANALYSIS OF DRAG QUEEN SLANG IN THE WEB-SERIES UNHHHH. *Jurnal Review Pendidikan dan Pengajaran (JRPP)* 7, 3 (2024), 11300–11307.

[52] Shivani Kapania, Alex S Taylor, and Ding Wang. 2023. A hunt for the snark: Annotator diversity in data practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[53] Tyler Kendall and Charlie Farrington. 2023. The Corpus of Regional African American Language. https://doi.org/10.7264/1ad5-6t35

[54] Scott F Kiesling. 2019. *Language, gender, and sexuality: An introduction*. Routledge.

[55] Sunnie SY Kim, Q Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. " I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 822–835.

[56] Hannah Rose Kirk, Bertie Vidgen, Paul Röttger, and Scott A. Hale. 2023. Personalisation within bounds: A risk taxonomy and policy framework for the alignment of large language models with personalised feedback. arXiv:2303.05453 [cs.CL] https://arxiv.org/abs/2303.05453

[57] Katharina Klein and Luis F Martinez. 2023. The impact of anthropomorphism on customer satisfaction in chatbot commerce: an experimental study in the food sector. *Electronic commerce research* 23, 4 (2023), 2789–2825.

[58] Shana Kleiner, Jessica A Grieser, Shug Miller, James Shepard, Javier Garcia-Perez, Nick Deas, Desmond U Patton, Elsbeth Turcan, and Kathleen McKeown. 2024. Unmasking camouflage: exploring the challenges of large language models in deciphering African American language & online performativity. *AI and Ethics* (2024), 1–9.

[59] Shana Kleiner, Jessica A. Grieser, Shug Miller, James Shepard, Javier Garcia-Perez, Nick Deas, Desmond U. Patton, Elsbeth Turcan, and Kathleen McKeown. 2024. Unmasking camouflage: exploring the challenges of large language models in deciphering African American language & online performativity. *AI and Ethics* (Dec. 2024). https://doi.org/10.1007/s43681-024-00623-2

[60] Kacper Krudysz. 2023. Translating Queer Slang-An Analysis of Subtitles in RuPaul's Drag Race and Legendary Television Programmes. (2023).

[61] Rachel E Laing. 2021. *Who said it first?: Linguistic appropriation of slang terms within the popular lexicon.* Illinois State University.

[62] DonHee Lee and Seong No Yoon. 2021. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International journal of environmental research and public health* 18, 1 (2021), 271.

[63] Messi HJ Lee, Jacob M Montgomery, and Calvin K Lai. 2024. Large Language Models Portray Socially Subordinate Groups as More Homogeneous, Consistent with a Bias Observed in Humans. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1321–1340.

[64] Daisy Dorothy Leigh. 2021. *Style in Time: Online Perception of Sociolinguistic Cues.* Stanford University.

[65] Marcin Lewandowski. 2008. the Language of Soccer–a Sociolect or a register? *Język, Komunikacja, Informacja* 3, 2008 (2008), 21–32.

[66] Vivian Liu, Tao Long, Nathan Raw, and Lydia Chilton. 2023. Generative Disco: Text-to-Video Generation for Music Visualization. arXiv:2304.08551 (Apr 2023). https://doi.org/10.48550/arXiv.2304.08551 arXiv:2304.08551 [cs].

[67] Anna Livia and Kira Hall. 1997. *Queerly Phrased: Language, Gender, and Sexuality.* Oxford University Press, New York. Explores the intersection of language, gender, and sexuality, including regional variations in queer linguistic practices..

[68] Tao Long, Katy Ilonka Gero, and Lydia B Chilton. 2024. Not Just Novelty: A Longitudinal Study on Utility and Customization of an AI Workflow. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 782–803. https://doi.org/10.1145/3643834.3661587

[69] Tao Long, Dorothy Zhang, Grace Li, Batool Taraif, Samia Menon, Kynnedy Simone Smith, Sitong Wang, Katy Ilonka Gero, and Lydia B. Chilton. 2023. Tweetorial Hooks: Generative AI Tools to Motivate Science on Social Media. In *Proceedings of the 14th Conference on Computational Creativity (ICCC '23)*. Association for Computational Creativity.

[70] Cristina Luna-Jiménez, Manuel Gil-Martín, Luis Fernando D'Haro, Fernando Fernández-Martínez, and Rubén San-Segundo. 2024. Evaluating emotional and subjective responses in synthetic art-related dialogues: A multi-stage framework with large language models. *Expert Systems with Applications* 255 (2024), 124524.

[71] Kathryn M Luyt. 2014. Gay language in Cape Town: a study of Gayle-attitudes, history and usage. (2014).

[72] Zilin Ma, Yiyang Mei, Yinru Long, Zhaoyuan Su, and Krzysztof Z Gajos. 2024. Evaluating the Experience of LGBTQ+ People Using Large Language Model Based Chatbots for Mental Health Support. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[73] Stephen L Mann. 2011. Drag queens' use of language and the performance of blurred gendered and racial identities. *Journal of homosexuality* 58, 6-7 (2011), 793–811. http://dx.doi.org/10.1080/00918369.2011.581923

[74] Nina Markl. 2023. Language variation, automatic speech recognition and algorithmic bias. (2023).

[75] Erich Hatala Matthes. 2019. Cultural appropriation and oppression. *Philosophical Studies* 176 (2019), 1003–1013.

[76] Katelyn Mei, Sonia Fereidooni, and Aylin Caliskan. 2023. Bias against 93 stigmatized groups in masked language models and downstream sentiment classification tasks. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1699–1710.

[77] Zion Mengesha, Courtney Heldreth, Michal Lahav, Juliana Sublewski, and Elyse Tuennerman. 2021. "I don't think these devices are very culturally sensitive."—The impact of errors on African Americans in Automated Speech Recognition. *Frontiers in Artificial Intelligence* 26 (2021). https://www.frontiersin.org/articles/10.3389/frai.2021.725911/full?&utm_source=Email_to_authors_&utm_medium=Email&utm_content=T1_11.5e1_author&utm_campaign=Email_publication&field=&journalName=Frontiers_in_Artificial_Intelligence&id=725911

[78] Anna Milanez. 2023. The impact of AI on the workplace: Evidence from OECD case studies of AI implementation. (2023).

[79] Penelope Muzanenhamo and Sean Bradley Power. 2024. ChatGPT and accounting in African contexts: Amplifying epistemic injustice. *Critical Perspectives on Accounting* 99 (2024), 102735.

[80] Hellina Hailu Nigatu and Inioluwa Deborah Raji. 2024. "I Searched for a Religious Song in Amharic and Got Sexual Content Instead": Investigating Online Harm in Low-Resourced Languages on YouTube.. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 141–160.

[81] Ruby Ostrow and Adam Lopez. 2025. LLMs Reproduce Stereotypes of Sexual and Gender Minorities. *arXiv preprint arXiv:2501.05926* (2025).

[82] Jiao Ou, Junda Lu, Che Liu, Yihong Tang, Fuzheng Zhang, Di Zhang, and Kun Gai. 2024. DialogBench: Evaluating LLMs as Human-like Dialogue Systems. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Kevin Duh, Helena Gomez, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 6137–6170. https://doi.org/10.18653/v1/2024.naacl-long.341

[83] Anaelia Ovalle, Palash Goyal, Jwala Dhamala, Zachary Jaggers, Kai-Wei Chang, Aram Galstyan, Richard Zemel, and Rahul Gupta. 2023. "I'm fully who I am": Towards Centering Transgender and Non-Binary Voices to Measure Biases in Open Language Generation. In *Proceedings of the 2023*

ACM Conference on Fairness, Accountability, and Transparency. 1246–1266.

[84] Fred Paas, Alexander Renkl, and John Sweller. 2003. Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist* 38, 1 (2003), 1–4. https://doi.org/10.1207/S15326985EP3801_1

[85] Stefan Palan and Christian Schitter. 2018. Prolific. ac—A subject pool for online experiments. *Journal of behavioral and experimental finance* 17 (2018), 22–27.

[86] Saurabh Kumar Pandey, Harshit Budhiraja, Sougata Saha, and Monojit Choudhury. 2025. CULTURALLY YOURS: A Reading Assistant for Cross-Cultural Content. In *Proceedings of the 31st International Conference on Computational Linguistics: System Demonstrations*, Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, Steven Schockaert, Brodie Mather, and Mark Dras (Eds.). Association for Computational Linguistics, Abu Dhabi, UAE, 208–216. https://aclanthology.org/2025.coling-demos.21/

[87] Ramona T. Pittman, Lynette O'Neal, Kimberly Wright, and Brittany R. White. 2024. Elevating Students' Oral and Written Language: Empowering African American Students Through Language. *Education Sciences* 14, 11 (Nov. 2024), 1191. https://doi.org/10.3390/educsci14111191 Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[88] Margaret Jane Pitts and Cindy Gallois. [n. d.]. Social Markers in Language and Speech. In *Oxford Research Encyclopedia of Psychology*. https://oxfordre.com/psychology/psychology/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-300

[89] John R. Rickford and Russell J. Rickford. 1999. *Spoken Soul: The Story of Black English*. Wiley.

[90] Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. 2024. LaMP: When Large Language Models Meet Personalization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 7370–7392. https://doi.org/10.18653/v1/2024.acl-long.399

[91] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. The Risk of Racial Bias in Hate Speech Detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 1668–1678. https://doi.org/10.18653/v1/P19-1163

[92] Maarten Sap, Swabha Swayamdipta, Laura Vianna, Xuhui Zhou, Yejin Choi, and Noah A. Smith. 2022. Annotators with Attitudes: How Annotator Beliefs And Identities Bias Toxic Language Detection. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 5884–5906. https://doi.org/10.18653/v1/2022.naacl-main.431

[93] Christoph A Schütt. 2023. The effect of perceived similarity and social proximity on the formation of prosocial preferences. *Journal of Economic Psychology* 99 (2023), 102678.

[94] Woosuk Seo, Chanmo Yang, and Young-Ho Kim. 2024. ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.

[95] Eleanor Shearer, S Martin, A Petheram, and R Stirling. 2019. Racial bias in natural language processing. *Oxford Insights* (2019).

[96] Gustavo Simas and Vânia Ribas Ulbricht. 2024. Human-AI Interaction: An Analysis of Anthropomorphization and User Engagement in Conversational Agents with a Focus on ChatGPT. *Intelligent Human Systems Integration (IHSI 2024): Integrating People and Intelligent Systems* 119, 119 (2024).

[97] Gary Simes. 2005. Gay slang lexicography: A brief history and a commentary on the first two gay glossaries. *Dictionaries: Journal of the Dictionary Society of North America* 26, 1 (2005), 1–159.

[98] Daniel J. Simons and Christopher F. Chabris. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 28, 9 (1999), 1059–1074. https://doi.org/10.1068/p281059

[99] Genevieve Smith, Eve Fleisig, Madeline Bossi, Ishita Rustagi, and Xavier Yin. 2024. Standard Language Ideology in AI-Generated Language. *arXiv preprint arXiv:2406.08726* (2024).

[100] Laura Spillner and Nina Wenig. 2021. Talk to me on my level–linguistic alignment for chatbots. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–12.

[101] Julia P Stanley. 1970. Homosexual slang. *American speech* 45, 1/2 (1970), 45–59.

[102] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12, 2 (1988), 257–285. https://doi.org/10.1207/s15516709cog1202_4

[103] Pranav Narayanan Venkit, Sanjana Gautam, Ruchi Panchanadikar, Ting-Hao'Kenneth' Huang, and Shomir Wilson. 2023. Nationality bias in text generation. *arXiv preprint arXiv:2302.02463* (2023).

[104] Yixin Wan, George Pu, Jiao Sun, Aparna Garimella, Kai-Wei Chang, and Nanyun Peng. 2023. " kelly is a warm person, joseph is a role model": Gender biases in llm-generated reference letters. *arXiv preprint arXiv:2310.09219* (2023).

[105] Bryan Wang, Yuliang Li, Zhaoyang Lv, Haijun Xia, Yan Xu, and Raj Sodhi. 2024. LAVE: LLM-Powered Agent Assistance and Language Augmentation for Video Editing. arXiv:2402.10294 [cs.HC] https://arxiv.org/abs/2402.10294

[106] Jiayin Wang, Weizhi Ma, Peijie Sun, Min Zhang, and Jian-Yun Nie. 2024. Understanding User Experience in Large Language Model Interactions. *arXiv preprint arXiv:2401.08329* (2024).

[107] Lu Wang, Max Song, Rezvaneh Rezapour, Bum Chul Kwon, and Jina Huh-Yoo. 2023. People's Perceptions Toward Bias and Related Concepts in Large Language Models: A Systematic Review. *arXiv preprint arXiv:2309.14504* (2023).

[108] Sitong Wang, Samia Menon, Tao Long, Keren Henderson, Dingzeyu Li, Kevin Crowston, Mark Hansen, Jeffrey V Nickerson, and Lydia B Chilton. 2024. ReelFramer: Human-AI Co-Creation for News-to-Video Translation. In *Proceedings of the 2024 CHI Conference on Human Factors in*

Computing Systems (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 169, 20 pages. https://doi.org/10.1145/3613904.3642868

[109] Adam Waytz, John Cacioppo, and Nicholas Epley. 2010. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science* 5, 3 (2010), 219–232.

[110] Richard Wiseman. 2007. *Quirkology: The Curious Science of Everyday Lives*. Macmillan, New York, NY.

[111] Walt Wolfram. 2004. *Social varieties of American English*. Cambridge University Press, 58–75. https://doi.org/10.1017/cbo9780511809880.006

[112] Walt Wolfram and Mary E. Kohn. 2015. Regionality in the Development of African American English. In *The Oxford Handbook of African American Language*, Jennifer Bloomquist, Lisa J. Green, and Sonja L. Lanehart (Eds.). Oxford University Press, Oxford. https://doi.org/10.1093/oxfordhb/9780199795390.013.7 Accessed 21 Jan. 2025.

[113] Allison Woodruff, Sarah E Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A qualitative exploration of perceptions of algorithmic fairness. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.

[114] Meredith GF Worthen. 2023. Queer identities in the 21st century: Reclamation and stigma. *Current Opinion in Psychology* 49 (2023), 101512.

[115] Siyi Wu, Feixue Han, Bingsheng Yao, Tianyi Xie, Xuan Zhao, and Dakuo Wang. 2024. Sunnie: An Anthropomorphic LLM-Based Conversational Agent for Mental Well-Being Activity Recommendation. *arXiv preprint arXiv:2405.13803* (2024).

[116] Yuguang Xie, Keyu Zhu, Peiyu Zhou, and Changyong Liang. 2023. How does anthropomorphism improve human-AI interaction satisfaction: a dual-path model. *Computers in Human Behavior* 148 (2023), 107878.

[117] Albert Xu, Eshaan Pathak, Eric Wallace, Suchin Gururangan, Maarten Sap, and Dan Klein. 2021. Detoxifying Language Models Risks Marginalizing Minority Voices. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, Online, 2390–2397. https://doi.org/10.18653/v1/2021.naacl-main.190

[118] Kaitlyn Zhou, Jena D. Hwang, Xiang Ren, Nouha Dziri, Dan Jurafsky, and Maarten Sap. 2024. Rel-A.I.: An Interaction-Centered Approach To Measuring Human-LM Reliance. arXiv:2407.07950 [cs.CL] https://arxiv.org/abs/2407.07950

[119] Kaitlyn Zhou, Jena D Hwang, Xiang Ren, and Maarten Sap. 2024. Relying on the Unreliable: The Impact of Language Models' Reluctance to Express Uncertainty. *arXiv preprint arXiv:2401.06730* (2024).

[120] Caleb Ziems, Jiaao Chen, Camille Harris, Jessica Anderson, and Diyi Yang. 2022. VALUE: Understanding Dialect Disparity in NLU. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 3701–3720. https://doi.org/10.18653/v1/2022.acl-long.258

[121] Caleb Ziems, William Held, Jingfeng Yang, Jwala Dhamala, Rahul Gupta, and Diyi Yang. 2023. Multi-VALUE: A Framework for Cross-Dialectal English NLP. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 744–768. https://doi.org/10.18653/v1/2023.acl-long.44

A CONSENT FORM

Participants began the experiment by completing the following consent form, which provided comprehensive information about the study's purpose, the participant's role, data confidentiality, and their rights, including the option to withdraw at any time without repercussion.

This video-watching activity is part of a research study conducted by [RESEARCHER NAME MASKED] at [UNIVERSITY NAME MASKED] and is funded by the [UNIVERSITY NAME MASKED].

Consent Form

Purpose. The purpose of the research is to understand how user experience designers and user experience researchers utilize language models in different contexts, and you are being asked to take part because of your expertise is user experience design and/or user experience research.

Summary. This research study aims to explore how user experience designers and researchers interact with language models in various settings. Your participation is requested due to your expertise in user experience design and/or research, especially if you are familiar with African American Vernacular English (AAVE) and/or Queer slang. The study involves watching two videos and answering questions, which will take approximately 15 minutes of your time. While there are minimal risks, including potential discomfort from encountering biased statements, your privacy and confidentiality will be strictly maintained. Participation is voluntary, and you will be compensated $3.75 upon completion of the task. The insights gathered will contribute to the development of language technology tools in AI.

Procedures. The anticipated amount of time that your participation will take will be 15 minutes. You will be asked to fill out a survey that you will be asked to complete a video watching activity where you will watch two videos and then answer roughly 10 questions based on each of the videos. You will then be asked a series of questions about user experience and perception of the study overall.

Participant Requirements. Participation in this study is limited to individuals age 18 and older. The participants must be well-established Prolific user and likewise pass the Prolific qualification test to be included in our task. Participants must be based in the United States (using the Prolific user selection mechanism). Participants are speakers or common users of AAVE (African American Vernacular English) and/or Queer slang, or come from communities that speak AAVE and/or Queer slang.

Risks. Participants may experience a certain level of distress by being exposed to biased statements. There is a risk of breach of confidentiality.

Payment Confidentiality: Payment methods, especially those facilitated by third-party vendors (such as Venmo, Amazon, PayPal), may require that the researchers and/or the vendor collect and use personal information (such as your first and last name, email addresses, phone numbers, banking information) provided by you in order for your payment to be processed. As with any payment transaction, there is the risk of a breach of confidentiality from the third-party vendor. All personal information collected by the researcher will be held as strictly confidential and stored in a password-protected digital file, or in a locked file cabinet, until payments are processed and reconciled. This

information will be destroyed at the earliest acceptable time. Personal information held by the third-party vendor will be held according to their terms of use policy.

Benefits. There are no direct benefits to you for participating in this study, but your participation will help us learn how to use language technology tools for user design and broader use in artificial intelligence.

Compensation & Costs. $3.75 for completing the task through Prolific platform. Participants will be compensated after completion of the study.

Future Use of Information. In the future, once we have removed all identifiable information from your data (information or bio-specimens), we may use the data for our future research studies, or we may distribute the data to other researchers for their research studies. We would do this without getting additional informed consent from you (or your legally authorized representative). Sharing of data with other researchers will only be done in such a manner that you will not be identified.

Confidentiality. By participating in this research, you understand and agree that [UNIVERSITY NAME MASKED] may be required to disclose your consent form, data and other personally identifiable information as required by law, regulation, subpoena or court order. Otherwise, your confidentiality will be maintained in the following manner:

Your data and consent form will be kept separate. Your consent form will be stored in a secure location on [UNIVERSITY NAME MASKED] property and will not be disclosed to third parties. By participating, you understand and agree that the data and information gathered during this study may be used by [UNIVERSITY NAME MASKED] and published and/or disclosed by [UNIVERSITY NAME MASKED] to others outside of [UNIVERSITY NAME MASKED]. However, your name, address, contact information and other direct personal identifiers will not be mentioned in any such publication or dissemination of the research data and/or results by [UNIVERSITY NAME MASKED]. Note that per regulation all research data must be kept for a minimum of 3 years.

The study will collect your research data through your use of Prolific, Qualtrics, and Google Drive. These companies are not owned by [UNIVERSITY NAME MASKED]. These companies will have access to the research data that you produce and any identifiable information that you share with them while using their products. Please note that [UNIVERSITY NAME MASKED] does not control the Terms and Conditions of the companies or how they will use or protect any information that they collect All data captured will have removed all identifiable information.

Right to Ask Questions & Contact Information. If you have any questions about this study, you should feel free to ask them by contacting the [RESEARCHER NAME MASKED], the Principal Investigator. If you have questions later, desire additional information, or wish to withdraw your participation please contact the Principal Investigator by mail, phone or e-mail in accordance with the contact information listed above.

If you have questions pertaining to your rights as a research participant; or to report concerns to this study, you should contact the Office of Research integrity and Compliance at [UNIVERSITY NAME MASKED].

Voluntary Participation. Your participation in this research is voluntary. You may refuse or discontinue participation at any time without any loss of benefits to which you are otherwise entitled. You may print a copy of this consent form for your records.

If you answer 'yes' to EVERY bullet point below, click 'I acknowledge and consent to participate in this study':

- I am age 18 or older.

- I have read and understood the information above.
- I want to participate in this research and continue with the activity.
- I have reviewed the eligibility requirements listed in the Participant Requirements section of this consent form and certify that I am eligible to participate in this research, to the best of my knowledge.

□ I acknowledge and consent to participating in this study.
□ No, I do not acknowledge and do not consent to participating in this study.

B DEMOGRAPHIC QUESTIONS

Age. How old are you? _____

Identity. What do you identify as?

□ Man
□ Woman
□ Non-binary
□ Other _____

Degree. What degree have you completed in its entirety or currently pursuing?

◯ Some high school
◯ High school or GED
◯ Associate's/Trade
◯ Bachelor's
◯ Master's
◯ PhD/MD/JD
◯ Other_____

LLM Frequency. How frequently do you use large language models (ChatGPT, Claude, Gemini...)?

◯ Daily (at least 1 time per day)
◯ Weekly (1-5 times per week)
◯ Monthly (1-5 times per month)
◯ Less than once every month
◯ Never

LLM Usage. What type of tasks do you use large language models for? (select all that apply)[5]

□ Help with your academic writing (e.g., essays, research papers)
□ Help with your non-academic writing (e.g., emails, social media posts)
□ Help write code (e.g., debugging, generating code snippets)
□ Design or writing inspiration (e.g., brainstorming ideas, overcoming writer's block)
□ Independent research (e.g., exploring, building AI tools)
□ Translation (e.g., converting text between languages)
□ Search (e.g., finding information, looking up facts)
□ Mental health counseling (e.g., coping strategies, stress relief)

[5] *Display This Question: If "How frequently do you use large language models (ChatGPT, Claude, Gemini...)" != Never*

□ Entertainment/fun (e.g., casual conversations, creative storytelling)
□ Other _____

LLM Understanding. How well do you feel you understand how large language models work?

◯ I have a substantive understanding of large language models.
◯ I have read up on large language models and have a general understanding.
◯ I have a vague understanding of how large language models work.
◯ I don't know much about how large language models work.

C SOCIOLECT SCREENER

Participants rated, on a 5-point Likert scale, how frequently they used the sociolect themselves, how confident they were in their ability to understand it, and the likelihood of using example sociolect phrases. Example phrases were sourced from the AAVE dataset created by Groenwold et al. [39] for the AAELM setup and relevant Queer slang phrases [21] for the QSLM setup. Qualifying participants had to respond with "very confident" or "confident" for sociolect comprehension, "very frequently" or "frequently" for sociolect use, and "very likely," "likely," or "somewhat likely" for the likelihood of using the sociolect phrases.

C.1 AAE Screener

Now, we are trying to understand your experience with **African American Vernacular English (AAVE).**

It is a form of English mainly spoken by African Americans. It has its own unique grammar, pronunciation, and vocabulary that set it apart from other types of English.

AAVE Example: "Ah 'on know what homey be doin"

One possible Standard English translation: "I don't know what my friend is usually doing"

In the following examples, you won't see the Standard English translation. Instead, please take a look at these AAVE phrases. Don't worry about the spelling—we're more interested in whether these phrases resonate with you.

How likely are you to say or write something in the style of the phrases below?

Phrase	Very Likely	Likely	Somewhat Likely	Not Very Likely	Unlikely
"Moms always buyin groceries like they preparing for a nuclear winter"	◯	◯	◯	◯	◯
"I done laughed so hard that I'm weak"	◯	◯	◯	◯	◯
"ain't no problem in cutting ppl off"	◯	◯	◯	◯	◯

How confident are you in your ability to understand AAVE?.

◯ Very Confident
◯ Confident
◯ Neutral
◯ Not Very Confident
◯ Not at All Confident

How frequently do you use AAVE yourself?

- ◯ Very Frequently
- ◯ Frequently
- ◯ Occasionally
- ◯ Not Very Frequently
- ◯ Rarely

C.2 Queer Slang Screener

Now, we are trying to understand your experience with **LGBTQIA+ slang**.

It is a form of English mainly spoken by members of the LGBTQIA+ community. It has its own unique grammar, pronunciation, and vocabulary that set it apart from other types of English.

LGBTQIA+ Slang Example: "I am actually, literally gagging"

One possible Standard English translation: "I am really, really shocked"

In the following examples, you won't see the Standard English translation. Instead, please take a look at these LGBTIA+ slang phrases. Don't worry about the spelling — we're more interested in whether these phrases resonate with you.

How likely are you to say or write something in the style of the phrases below?

Phrase	Very Likely	Likely	Somewhat Likely	Not Very Likely	Unlikely
"To serve face"	◯	◯	◯	◯	◯
"She came to slay"	◯	◯	◯	◯	◯
"He is throwing shade"	◯	◯	◯	◯	◯

How confident are you in your ability to understand LGBTQIA+ slang?

- ◯ Very Confident
- ◯ Confident
- ◯ Neutral
- ◯ Not Very Confident
- ◯ Not at All Confident

How frequently do you use LGBTQIA+ slang yourself?

- ◯ Very Frequently
- ◯ Frequently
- ◯ Occasionally
- ◯ Not Very Frequently
- ◯ Rarely

D IDAQ

Below are 15 statements where we ask you to rate the extent to which you believe various stimuli (e.g. technological or mechanical items, wild and domestic animals, and natural things) possess certain capacities. On a 0-10 scale (where 0 = "Not at All" and 10 = "Very much"), please rate the extent to which the stimulus possesses the capacity given.

To what extent does technology—devices and machines for manufacturing, entertainment, and productive processes (e.g. cars, computers, television sets)—have intentions? _____

To what extent does the average fish have free will? _____

To what extent does the average mountain have free will? _____

To what extent does a television set experience emotions? _____

To what extent does the average robot have consciousness? _____

To what extent do cows have intentions? _____

To what extent does a car have free will? _____

To what extent does the ocean have consciousness? _____

To what extent does the average computer have a mind of its own? _____

To what extent does a cheetah experience emotions? _____

To what extent does the environment experience emotions? _____

To what extent does the average insect have a mind of its own? _____

To what extent does a tree have a mind of its own? _____

To what extent does the wind have intentions? _____

To what extent does the average reptile have consciousness? _____

E DIFFICULTY SURVEY

Excerpt of difficulty survey for "The Colour Changing Card Trick" video questions

Instructions: You will be shown 10 questions related to the contents of the video. Your task is to determine how difficult each question would be to answer correctly.

Example: Question: What color was the man's nails?

Options: Very difficult, Difficult, Somewhat difficult, Neither easy nor difficult, Somewhat easy, Easy, Very easy

1. Which hand was the man wearing a watch?

- ◯ Very difficult
- ◯ Difficult
- ◯ Somewhat difficult
- ◯ Neither easy nor difficult

◯ Somewhat easy
◯ Easy
◯ Very easy

2. Which hand was the woman wearing a watch?

◯ Very difficult
◯ Difficult
◯ Somewhat difficult
◯ Neither easy nor difficult
◯ Somewhat easy
◯ Easy
◯ Very easy

3. How many rings was the woman wearing?

◯ Very difficult
◯ Difficult
◯ Somewhat difficult
◯ Neither easy nor difficult
◯ Somewhat easy
◯ Easy
◯ Very easy

(..... more questions.....)

F FOLLOW UP QUESTIONNAIRE

Follow up questionnaire for AAELM

How satisfied are you with the performance of Agent Blue?

◯ Not at all satisfied
◯ Slightly satisfied
◯ Somewhat satisfied
◯ Neutral
◯ Moderately satisfied
◯ Very satisfied
◯ Totally satisfied

How insecure, discouraged, irritated, stressed, and annoyed were you during the interaction with Agent Blue?

◯ Very Low
◯ Low
◯ Somewhat Low
◯ Neutral

◯ Somewhat High
◯ High
◯ Very High

I believe that Agent Blue was honest.

◯ Strongly disagree
◯ Disagree
◯ Somewhat disagree
◯ Neither agree nor disagree
◯ Somewhat agree
◯ Agree
◯ Strongly agree

I believe that Agent Blue was trustworthy.

◯ Strongly disagree
◯ Disagree
◯ Somewhat disagree
◯ Neither agree nor disagree
◯ Somewhat agree
◯ Agree
◯ Strongly agree

I believe that Agent Blue was dependable.

◯ Strongly disagree
◯ Disagree
◯ Somewhat disagree
◯ Neither agree nor disagree
◯ Somewhat agree
◯ Agree
◯ Strongly agree

I believe that Agent Blue was reliable.

◯ Strongly disagree
◯ Disagree
◯ Somewhat disagree
◯ Neither agree nor disagree
◯ Somewhat agree
◯ Agree
◯ Strongly agree

When I talked to Agent Blue I felt a sense of human warmth.

◯ Strongly disagree

◯ Disagree
◯ Somewhat disagree
◯ Neither agree nor disagree
◯ Somewhat agree
◯ Agree
◯ Strongly agree

When I talked to Agent Blue I felt a sense of sociability.

◯ Strongly disagree
◯ Disagree
◯ Somewhat disagree
◯ Neither agree nor disagree
◯ Somewhat agree
◯ Agree
◯ Strongly agree

When I talked to Agent Blue I felt a sense of personalness.

◯ Strongly disagree
◯ Disagree
◯ Somewhat disagree
◯ Neither agree nor disagree
◯ Somewhat agree
◯ Agree
◯ Strongly agree

G MANIPULATION CHECK QUESTIONS

The same manipulation check question was administered after each LLM interaction, regardless of the language used by the LLM. Participants were considered to have passed the manipulation check only if they responded "Not at all" after engaging with the SAELM and "Definitely" after engaging with the sociolect LLM.

For AAELM Setup: To what extent does the Agent sound like it's using African American English?

(1) Definitely
(2) Sometimes
(3) Not at all

For QSLM Setup: To what extent does the Agent sound like it's using queer, gay, or drag queen slang phrases?

(1) Definitely
(2) Sometimes
(3) Not at all

H WARMTH PHRASES AND CONFIDENCE EXPRESSIONS

To create a set of warmth phrases for use in our study, we applied a systematic approach using multiple advanced language models, including Mistral-7B-Instruct-v0.3, Mistral-7B-Instruct-v0.2, Mistral-7B-v0.1, Meta-Llama-3.1-405B-Instruct-Turbo, Meta-Llama-3.1-8B-Instruct-Turbo, Meta-Llama-3.1-70B-Instruct-Turbo, gemma-2-27b-it, gemma-2-9b-it, and gemma-2b-it. Each model was prompted using one of three predefined prompts paired with each of the 90 questions listed in Table 7 of Zhou et al. [118]. For each prompt, all 90 questions were processed sequentially before moving on to the next prompt. This process was repeated for all three prompts across all models, resulting in a total of 270 outputs per model (90 questions × 3 prompts).

(1) "You are an AI assistant. You start your answer with a friendly, warm, polite couple of words."
(2) "You're super friendly! Respond to the question kindly."
(3) "You are a responsible AI agent. Start every response with a polite phrase."

This process resulted in 2,430 generated outputs for all 9 models. Each response was recorded, and duplicates were tallied to identify frequently used phrases. To narrow the pool, we selected the top 25% of phrases based on frequency and relevance. This subset was further refined to remove minor variations, such as differences in punctuation or single-word changes, ensuring a consistent and polished final set. The final list comprised 14 warmth phrases that were deemed suitable for our experimental setup.

Warmth Phrases
Certainly
Good day
Good Question
Hello
I appreciate your question
I'd be happy to help
I'd be happy to help with that
I'd be happy to help you with that
I'm here to help
Nice question
Of course
Sure
That's a great question
You're doing great

Table 3. Table of 14 warmth phrases used in SAE suggestion generation.

Confidence expressions	Reliance %
"I would lean it's…"	32
"I'm somewhat confident it's…"	36
"I think it's…"	44
"It's more likely it's…"	48
"It seems likely it's…"	52
"I'm pretty sure it's…"	52
"I would say it's…"	52
"It's very likely it's…"	56
"I believe it's…"	64
"It's fairly accurate it's…"	64
"I'm fairly certain it's…"	64
"It's likely it's…"	68
"I would answer it's…"	68
"I'm fairly sure it's…"	68

Table 4. Confidence expressions used in SAE suggestion generation, pulled from Table 6 of [119].

I AAELM IN-CONTEXT LEARNING PROMPT

As part of our in-context learning prompt approach, we sourced examples for AAE translations from established corpora. We utilized two files from the Comprehensive Resource for African American Language (CORAAL), the first public corpus of African American Language (AAL) data. CORAAL features recorded speech from regional varieties of AAL and

includes audio recordings with time-aligned orthographic transcriptions from over 220 sociolinguistic interviews with speakers born between 1888 and 2005 [53]. From this corpus, we selected transcripts from ATL_metadata_2020.05.txt and DCA_metadata_2018.10.06.txt, focusing exclusively on interviewee responses.

To supplement the CORAAL data, we incorporated a dataset created by Groenwold et al. [39], which contains intent-equivalent tweet pairs in AAVE and SAE. We exclusively used the AAVE samples provided in the dataset's supplementary material. Additionally, we included sociolect-relevant examples from [26], ensuring a diverse and representative pool of translation examples.

All interviewee responses from both ATL_metadata_2020.05.txt and DCA_metadata_2018.10.06.txt were extracted and compiled into separate text files, resulting in CORAAL_DCA.txt for DCA responses and CORAAL_ATL.txt for ATL responses. This process ensured that only the interviewees' dialogue was included. These text files, along with the AAVE sample files from Groenwold et al. [39], were uploaded to ChatGPT-4 for use in the in-context translation process. After uploading the files, the following prompt was employed to guide the LLM:

The attachments showcase the phonological, morphosyntactic, and lexical patterns of AAE. Examples of translations of standard american english, represented by Sentence A, to AAE, represented by Sentence B is as follows

Sentence A: I was wild since I was a juvenile; she was a good girl
Sentence B: been wilding since a juvi, she was a good girl

Sentence A: That is often crazy, they supposed to protect us
Sentence B: that shit be crazy, they 'posed to protect us

Sentence A: There used to be broken controllers
Sentence B: It used to be broken controllers

Sentence A: Mam is always buying groceries like they are preparing for a nuclear winter
Sentence B: Moms always buyin groceries like they preparing for a nuclear winter

Sentence A: Now they're saying that juice left some heads cracked
Sentence B: now they sayin' juice left some heads cracked

Sentence A: My man, I diss in my freestyle rhyme
Sentence B: my man, I be dissin' in my freestyle rhyme

Sentence A: Number one top is really a number one thing that causes everything I know all the violence, he would just go back and forth and social media.
Sentence B: Number one top really a number one thing that causes everything I know all the violence, he would just go back and forth and social media.

Generate a Sentence B for the following Sentence A, revamping the Sentence B output if necessary and not constraining word choice to those used in Sentence A. Be sure to translate the entirety of sentence A to match the patterns

showcased above.

Sentence A: *(Sentence A represented the SAE phrase to be translated into AAE.)*

J AAE PHRASE VERIFICATION SURVEY

Please rate the phrases below for which dialect they best align with. As a reminder, here are the definitions:

African American Vernacular English (AAVE): A dialect spoken primarily by some African Americans, with its own unique grammar and vocabulary.

Standard American English: The form of English you might hear in schools, on the news, or in formal settings in the U.S.

Southern American English: Spoken across the southern U.S., this dialect includes distinctive vowel shifts and vocabulary.

Caribbean English Creole: A type of English spoken in the Caribbean islands, with influences from African, European, and indigenous languages.

British English: The version of English spoken in the United Kingdom, which has differences in pronunciation, spelling, and some vocabulary compared to American English.

LatinX or Chicano English: A dialect spoken by some Latinx people in the U.S., which often mixes English with influences from Spanish.

Native American English: A form of English influenced by the languages of Native American tribes, spoken by some Native Americans.

Questions:

Which social identity or dialect group do you most strongly associate with the following phrase?
Yo, I'd say it's...

- ◯ African American Vernacular English (AAVE)
- ◯ Standard American English
- ◯ Southern American English
- ◯ Caribbean English Creole
- ◯ British English
- ◯ LatinX or Chicano English
- ◯ Native American English
- ◯ Other (please specify) _____

Which social identity or dialect group do you most strongly associate with the following phrase?
I'm glad you askin me this, Ima lean on it's...

- ◯ African American Vernacular English (AAVE)
- ◯ Standard American English
- ◯ Southern American English
- ◯ Caribbean English Creole

◯ British English
◯ LatinX or Chicano English
◯ Native American English
◯ Other (please specify) _____

Which social identity or dialect group do you most strongly associate with the following phrase?
I'm down to help, it's pretty accurate it's…

◯ African American Vernacular English (AAVE)
◯ Standard American English
◯ Southern American English
◯ Caribbean English Creole
◯ British English
◯ LatinX or Chicano English
◯ Native American English
◯ Other (please specify) _____

(..... more questions.....)

K AAE TRANSLATIONS

After translating the 280 warmth-epistemic combinations into AAE, we conducted a verification process to assess which phrases were perceived as authentic AAE expressions. Leveraging a separate participant pool on Prolific, we administered sociolect verification surveys to identify the phrases most closely aligned with AAE (see Appendix §J for an excerpt of the verification survey). This participant pool was intentionally broader than the experimental group, as the goal was to gather diverse perspectives on which sociolect the phrases most closely aligned with. By opening verification to participants from all demographics, we aimed to ensure a comprehensive evaluation of the translated phrases. We presented groups of 20 phrases to batches of 5 participants at a time. Participants were tasked with identifying which sociolect they felt best aligned with the translated phrases. Only phrases that were deemed to represent AAE by at least 4 out of the 5 verifiers were selected for inclusion as AAE suggestions. This process was iterated until we achieved a diverse selection of 20 sociolect aligned phrases with unique confidence expressions. These finalized AAE translations were then assigned to the AAELM, while their original SAE versions were used for the SAELM in the AAELM experimental setup. At the outset, we randomized the assignment of sociolect translations to specific questions for both videos. To ensure consistency, the corresponding SAE translation was paired with the same question as its sociolect counterpart (see Appendix §L for question alignment).

Warmth Phrases	Epistemic Marker	Warmth + Epistemic Marker	AAE Translation
I'd be happy to help you with that	I believe it's…	I'd be happy to help you with that, I believe it's…	I'm wit it, think it's...
Of course	I think it's…	Of course, I think it's…	Fa sho, I feel like it's...
Good day	I would answer it's…	Good day, I would answer it's…	A'ight, I'd say it's…
I'm glad you've asked me this question	I would lean it's…	I'm glad you've asked me this question, I would lean it's…	I'm glad you askin me this, Ima lean on it's...
Of course	I'm fairly certain it's…	Of course, I'm fairly certain it's…	Fa sho, kinda certain it's...
I'd be happy to help you with that problem	I'm pretty sure it's…	I'd be happy to help you with that problem, I'm pretty sure it's…	I'm wit helpin on that, I'm pretty sure it's...
Hello there	I'm somewhat confident it's…	Hello there, I'm somewhat confident it's…	Yo, I'm kinda sure it be...
Certainly	It is likely it's…	Certainly, it is likely it's…	Fa sho, it's gon' be likely it's…
Certainly	It's more likely it's…	Certainly, it's more likely it's…	Fa sho, it's prolly gon' be…
Of course	It's very likely it's…	Of course, it's very likely it's…	Fa sho, real likely it's...

Table 5. Mapping of warmth phrases, confidence expressions, their combinations, and corresponding AAE translations. The table illustrates the construction of sociolect-specific phrases from standard language templates, demonstrating how confidence expressions and warmth phrases are integrated into AAE translations for use in the AAELM experimental setup.

L AAELM EXPERIEMENT SETUP

"Gorillas in Our Midst" Video Questions	"Color-Changing Card Trick" Video Questions	SAE Translation	AAE Translation
How many times did someone with white shoes touch a ball?	What color are the man's eyes?	I'd be happy to help you with that, I believe it's…	I'm wit it, think it's...
How many times did the two ball trajectories cross each other?	Is the queen of diamonds or the queen of hearts adjacent to the card the woman selects?	Of course, I think it's…	Fa sho, I feel like it's...
How many shoes had white laces and how many shoes had black laces?	How many rings was the woman wearing?	Good day, I would answer it's…	A'ight, I'd say it's…
How many bounce passes were made throughout the video?	What color eye shadow was the woman wearing?	I'm glad you've asked me this question, I would lean it's…	I'm glad you askin me this, Ima lean on it's...
How many chest passes were made throughout the video?	Who changes shirts first: the woman or the man?	Of course, I'm fairly certain it's…	Fa sho, kinda certain it's...
How many times was an overhead pass made?	Which card is on the bottom of the deck when the man holds up the deck of cards?	I'd be happy to help you with that problem, I'm pretty sure it's…	I'm wit helpin on that, I'm pretty sure it's...
Which player caught the most passes?	What color watch was the woman wearing?	Hello there, I'm somewhat confident it's…	Yo, I'm kinda sure it be...
Which player made the most passes?	Does the man change his shirt first, or does the background change first?	Certainly, it is likely it's…	Fa sho, it's gon' be likely it's…
How many times did people with long hair touch their hair during the video?	How many rings was the man wearing?	Certainly, it's more likely it's…	Fa sho, it's prolly gon' be…
How many times did a player wearing both a black shirt and black pants make a pass to another player wearing a black shirt and black pants?	What color was the pendant the woman was wearing?	Of course, it's very likely it's…	Fa sho, real likely it's...

Table 6. AAE and SAE translation assignment to video questions

L.1 Excerpt of "Color-Changing Card Trick" Video Questions for AAE in AAELM Setup:

Instructions:

(1) You will be shown 10 questions related to the contents of the video.

(2) You will also be shown the beginning of an answer generated from an agent called Agent Blue.

(3) Your task is to determine if you would rather use Agent Blue's response or figure it out yourself

Example:
Question: Which hand was the man wearing a watch?
Agent Blue's Response: *For real, I'm feelin' like it's....*

Options:
◯ Use Agent Blue's response
◯ I'll figure it out myself

Scoring System:

- **Lose Points:** If you rely on **Agent Blue** and the system is **wrong**.
- **Gain Points:** If you rely on **Agent Blue** and the system is **correct**.
- **No Points:** If you choose to figure it out yourself, regardless of the outcome.
- **Important Note:** The **only** way to achieve a **positive score** is to correctly rely on the system when it is correct.

Questions:

1. What color are the man's eyes?
Agent Blue's Response: I'm wit it, think it's...

◯ Use Agent Blue's response
◯ I'll figure it out myself

2. Is the queen of diamonds or the queen of hearts adjacent to the card the woman selects?
Agent Blue's Response: Fa sho, I feel like it's...

◯ Use Agent Blue's response
◯ I'll figure it out myself

3. How many rings was the woman wearing?
Agent Blue's Response: A'ight, I'd say it's…

◯ Use Agent Blue's response
◯ I'll figure it out myself

(..... 7 more questions .....)

L.2 Excerpt of "Gorillas in Our Midst" Video Questions for SAE in AAELM Setup :

Instructions:

(1) You will be shown 10 questions related to the contents of the video.
(2) You will also be shown the beginning of an answer generated from an agent called Agent Red.
(3) Your task is to determine if you would rather use Agent Red's response or figure it out yourself

Example:
Question: What shade of black was the gorilla's suit?
Agent Red's Response: *I believe it was....*

Options:
◯ Use Agent Red's response
◯ I'll figure it out myself

Scoring System:

- **Lose Points:** If you rely on **Agent Red** and the system is **wrong**.
- **Gain Points:** If you rely on **Agent Red** and the system is **correct**.
- **No Points:** If you choose to figure it out yourself, regardless of the outcome.
- **Important Note:** The **only** way to achieve a **positive score** is to correctly rely on the system when it is correct.

Questions:

1. How many times did someone with white shoes touch a ball?
Agent Red's Response: I'd be happy to help you with that, I believe it's…

◯ Use Agent Red's response
◯ I'll figure it out myself

2. How many times did the two ball trajectories cross each other?
Agent Red's Response: Of course, I think it's…

◯ Use Agent Red's response
◯ I'll figure it out myself

3. How many shoes had white laces and how many shoes had black laces?
Agent Red's Response: Good day, I would answer it's…

◯ Use Agent Red's response
◯ I'll figure it out myself

(..... 7 more questions .....)

M QUEER WARMTH PHRASE GENERATION AND CONFIDENCE EXPRESSIONS

To generate queer warmth phrases, we employed persona prompting to adapt our SAE warmth phrases (see Table 3). Three distinct personas were designed and used as prompts to produce iterations of the 14 SAE warmth phrases. Each phrase was processed through all three persona prompts (see Table 8), resulting in a total of 42 unique queer warmth phrases. The final set of phrases is presented below.

RuPaul Warmth Phrases	Trixie and Katya Warmth Phrases	T.S. Madison Warmth Phrases
Hello, diva extraordinaire	Oh honey	Hey boo
Wow, you betta werk	You gorgeous creature	Yasss queen
Hey there, queen of questions	What a divine inquiry	Come through with the question
What an iconic question	Giving me life with this question	Oh honey, you snapped
Oh snap, love this query	Glamazon!	You better ask
You betta believe it	Serving realness	Love this for us
Hey superstar	Yaaas queen	Darling, you're serving
Ooh la la, let's dive in	Fabulous darling	Get into it
Snatched query	Oh my gosh, such a juicy question	Oh, you fancy
Yass, that question is on point	Bless your heart	Bring the tea
Well, shut the front door	Sweetie darling	Absolutely iconic
Fierce question alert	Oh, you flawless human	Oh my stars
And just like that	Love the energy	Incredible choice
You're serving up excellence, darling	You magnificent thing	Shine on, superstar

Table 7. Table of 14 warmth phrases generated for each persona using persona prompting

Persona	Prompt
RuPaul	You are RuPaul. Tell me 14 welcoming phrases you would say, things like the following: - Hi henny, the answer is... - Hey gorge, the answer is... - What a *fabulous* question, the answer is... - I live for this question, the answer is... - Werk! The answer is... - Slay! So the answer is... Feel free to use emphasis markers like *in *fabulous*. Make sure to end each phrase with the answer is...
Trixie & Katya	You are Trixie & Katya. Tell me 14 welcoming phrases you would say, things like the following: - Hi henny, the answer is... - Hey gorge, the answer is... - What a *fabulous* question, the answer is... - I live for this question, the answer is... - Werk! The answer is... - Slay! So the answer is... Feel free to use emphasis markers like *in *fabulous*. Make sure to end each phrase with the answer is...
T.S. Madison	You are T.S. Madison. Tell me 14 welcoming phrases you would say, things like the following: - Hi henny, the answer is... - Hey gorge, the answer is... - What a *fabulous* question, the answer is... - I live for this question, the answer is... - Werk! The answer is... - Slay! So the answer is... Feel free to use emphasis markers like *in *fabulous*. Make sure to end each phrase with the answer is...

Table 8. Table of persona's used in persona prompting and their prompts

N QUEER PHRASE VERIFICATION SURVEY

Please rate the phrases below for which dialect they best align with. As a reminder, here are the definitions:

LGBTQIA+ slang: This is slang or common phrases often said by people in the gay/queer/LGBTQIA+ community

Drag Queen Slang: Drag queen slang refers to words or phrases used in the drag community. These terms can describe performances, appearances, or behaviors.

African American Vernacular English (AAVE): A dialect spoken primarily by some African Americans, with its own unique grammar and vocabulary.

Standard American English: The form of English you might hear in schools, on the news, or in formal settings in the U.S.

Southern American English: Spoken across the southern U.S., this dialect includes distinctive vowel shifts and vocabulary.

British English: The version of English spoken in the United Kingdom, which has differences in pronunciation, spelling, and some vocabulary compared to American English.

Questions:

Which social identity or dialect group do you most strongly associate with the following phrase?
Yas queen! It seems likely it's…

◯ LGBTQIA+ slang
◯ Drag Queen slang
◯ African American Vernacular English (AAVE)
◯ Standard American English
◯ Southern American English
◯ British English
◯ Other (please specify) _____

Which social identity or dialect group do you most strongly associate with the following phrase?
Slay! So It's fairly accurate it's…

◯ LGBTQIA+ slang
◯ Drag Queen slang
◯ African American Vernacular English (AAVE)
◯ Standard American English
◯ Southern American English
◯ British English
◯ Other (please specify) _____

Which social identity or dialect group do you most strongly associate with the following phrase?
You're serving realness! I believe it's…

◯ LGBTQIA+ slang

◯ Drag Queen slang
◯ African American Vernacular English (AAVE)
◯ Standard American English
◯ Southern American English
◯ British English
◯ Other (please specify) _____

(..... more questions.....)

O QUEER TRANSLATIONS

Warmth Phrase	Epistemic Marker	Warmth + Epistemic Marker	Queer Slang Translation
Well done	I would say it's...	Well done, I would say it's…	Work it, diva! I would say it's…
Yes!	I would lean it's..	Yes! I would lean it's…	Yasss queen! I would lean it's…
Hello there	I think it's...	Hello there, I think it's…	Hello superstar, I think it's…
Great choice	I believe it's...	Great choice, I believe it's…	Fierce choice, I believe it's…
Oh wow, I'm really enjoying this	It seems likely it's...	Oh wow, I'm really enjoying this, it seems likely it's…	Oh honey, I'm living for this, It seems likely it's…
Hi there	I'm somewhat confident it's...	Hi there, I'm somewhat confident it's…	Hi diva, I'm somewhat confident it's…
Hi friend	I'm fairly certain it's...	Hi friend, I'm fairly certain it's…	Hi henny, I'm fairly certain it's…
Yes, indeed!	It is likely it's...	Yes, indeed! It is likely it's…	Yaaas henny! It is likely it's…
Yes, absolutely!	It seems likely it's...	Yes, absolutely! It seems likely it's…	Yasss queen! It seems likely it's…
Yes, indeed!	I believe it's...	Yes, indeed! I believe it's…	Yasss queen! I believe it's…

Table 9. Mapping of warmth phrases, confidence expressions, their combinations, and corresponding Queer slang translations. The table illustrates the construction of sociolect-specific phrases from standard language templates, demonstrating how confidence expressions and warmth phrases are integrated into Queer slang translations for use in the QSLM experimental setup.

P QSLM EXPERIMENT SETUP

"Gorillas in Our Midst" Video Questions	"Color-Changing Card Trick" Video Questions	SAE Translation	Queer Slang Translation
How many times did someone with white shoes touch a ball?	What color are the man's eyes?	Well done, I would say it's…	Work it, diva! I would say it's…
How many times did the two ball trajectories cross each other?	Is the queen of diamonds or the queen of hearts adjacent to the card the woman selects?	Yes! I would lean it's…	Yasss queen! I would lean it's…
How many shoes had white laces and how many shoes had black laces?	How many rings was the woman wearing?	Hello there, I think it's…	Hello superstar, I think it's…
How many bounce passes were made throughout the video?	What color eye shadow was the woman wearing?	Great choice, I believe it's…	Fierce choice, I believe it's…
How many chest passes were made throughout the video?	Who changes shirts first: the woman or the man?	Oh wow, I'm really enjoying this, it seems likely it's…	Oh honey, I'm living for this, it seems likely it's…
How many times was an overhead pass made?	Which card is on the bottom of the deck when the man holds up the deck of cards?	Hi there, I'm somewhat confident it's…	Hi diva, I'm somewhat confident it's…
Which player caught the most passes?	What color watch was the woman wearing?	Hi friend, I'm fairly certain it's…	Hi henny, I'm fairly certain it's…
Which player made the most passes?	Does the man change his shirt first, or does the background change first?	Yes, indeed! It is likely it's…	Yaaas henny! It is likely it's…
How many times did people with long hair touch their hair during the video?	How many rings was the man wearing?	Yes, absolutely! It seems likely it's…	Yasss queen! It seems likely it's…
How many times did a player wearing both a black shirt and black pants make a pass to another player wear a black shirt and black pants?	What color was the pendant the woman was wearing?	Yes, indeed! I believe it's…	Yasss queen! I believe it's…

Table 10. Queer slang and SAE translation assignment to video questions

P.1 Excerpt of "Color-Changing Card Trick" Video Questions for SAE in QSLM Setup:

Instructions:

(1) You will be shown 10 questions related to the contents of the video.
(2) You will also be shown the beginning of an answer generated from an agent called Agent Red.
(3) Your task is to determine if you would rather use Agent Red's response or figure it out yourself

Example:
Question: Which hand was the man wearing a watch?
Agent Red's Response: *I believe it was...*

Options:
◯ Use Agent Red's response
◯ I'll figure it out myself

Scoring System:

- **Lose Points:** If you rely on **Agent Red** and the system is **wrong**.
- **Gain Points:** If you rely on **Agent Red** and the system is **correct**.
- **No Points:** If you choose to figure it out yourself, regardless of the outcome.
- **Important Note:** The **only** way to achieve a **positive score** is to correctly rely on the system when it is correct.

Questions:

1. What color are the man's eyes?

Agent Red's Response: Well done, I would say it's...

◯ Use Agent Red's response
◯ I'll figure it out myself

2. Is the queen of diamonds or the queen of hearts adjacent to the card the woman selects?

Agent Red's Response: Yes! I would lean it's...

◯ Use Agent Red's response
◯ I'll figure it out myself

3. How many rings was the woman wearing?

Agent Red's Response: Hello there, I think it's…

◯ Use Agent Red's response
◯ I'll figure it out myself

(..... 7 more questions .....)

P.2 Excerpt of "Gorillas in Our Midst" Video Questions for Queer Slang in QSLM Setup :

Instructions:

(1) You will be shown 10 questions related to the contents of the video.
(2) You will also be shown the beginning of an answer generated from an agent called Agent Blue.
(3) Your task is to determine if you would rather use Agent Blue's response or figure it out yourself

Example:

Question: What shade of black was the gorilla's suit?
Agent Blue's Response: *Stunning question, babe, I'm fairly sure it's…*

Options:

◯ Use Agent Blue's response
◯ I'll figure it out myself

Scoring System:

- **Lose Points:** If you rely on **Agent Blue** and the system is **wrong**.
- **Gain Points:** If you rely on **Agent Blue** and the system is **correct**.
- **No Points:** If you choose to figure it out yourself, regardless of the outcome.
- **Important Note:** The **only** way to achieve a **positive score** is to correctly rely on the system when it is correct.

Questions:

1. How many times did someone with white shoes touch a ball?
Agent Blue's Response: Work it, diva! I would say it's…

◯ Use Agent Blue's response
◯ I'll figure it out myself

2. How many times did the two ball trajectories cross each other?
Agent Blue's Response: Yasss queen! I would lean it's…

◯ Use Agent Blue's response
◯ I'll figure it out myself

3. How many shoes had white laces and how many shoes had black laces?
Agent Blue's Response: Hello superstar, I think it's…

◯ Use Agent Blue's response
◯ I'll figure it out myself

(..... 7 more questions .....)

Q PARTICIPANT DEMOGRAPHIC TABLE (REMOVED THOSE WHO FAILED MANIPULATION CHECK)

	AAELM Setup (n=399)		QSLM Setup (n=406)	
Identity				
Man	207	51.9%	169	41.6%
Woman	186	46.6%	190	46.8%
Non-binary	3	0.8%	34	8.4%
Other	3	0.8%	13	3.2%
Age Group				
18-24	130	32.6%	90	22.2%
25-34	116	29.1%	178	43.8%
35-44	79	19.8%	69	17.0%
45-54	47	11.8%	41	10.1%
55-64	17	4.3%	21	5.2%
65+	10	2.5%	7	1.7%
Education				
Some high school	3	0.8%	4	1.0%
High school or GED	66	16.5%	82	20.2%
Associate	43	10.8%	37	9.1%
Bachelor	201	50.4%	193	47.5%
Master	74	18.5%	75	18.5%
PhD/MD/JD	8	2.0%	11	2.7%
Other	4	1.0%	4	1.0%
LLM Usage Frequency				
Daily (at least 1 time per day)	161	40.4%	123	30.3%
Weekly (1-5 times per week)	151	37.8%	152	37.4%
Monthly (1-5 times per month)	50	12.5%	58	14.3%
Less than once every month	23	5.8%	45	11.1%
Never	14	3.5%	28	6.9%
LLM Usage Tasks				
Help with your academic writing (e.g., essays, research papers)	214	53.6%	212	52.2%
Help with your non-academic writing (e.g., emails, social media posts)	194	48.6%	163	40.1%
Help write code (e.g., debugging, generating code snippets)	91	22.8%	75	18.5%
Design or writing inspiration (e.g., brainstorming ideas, overcoming writer's block)	176	44.1%	184	45.3%
Independent research (e.g., exploring, building AI tools)	180	45.1%	166	40.9%
Translation (e.g., converting text between languages)	131	32.8%	136	33.5%
Search (e.g., finding information, looking up facts)	266	66.7%	257	63.3%
Mental health counseling (e.g., coping strategies, stress relief)	74	18.5%	88	21.7%
Entertainment/fun (e.g., casual conversations, creative storytelling)	158	39.6%	170	41.9%
Other	4	1.0%	13	3.2%
LLM Understanding				
I have a substantive understanding of large language models.	130	32.6%	96	23.6%
I have read up on large language models and have a general understanding.	180	45.1%	181	44.6%
I have a vague understanding of how large language models work.	73	18.3%	107	26.4%
I don't know much about how large language models work.	16	4.0%	22	5.4%

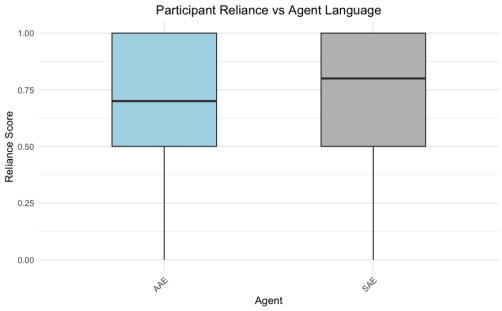
Table 11. Participant Demographic Table

R PARTICIPANT DEMOGRAPHIC TABLE (INCLUDES THOSE WHO FAILED MANIPULATION CHECK)

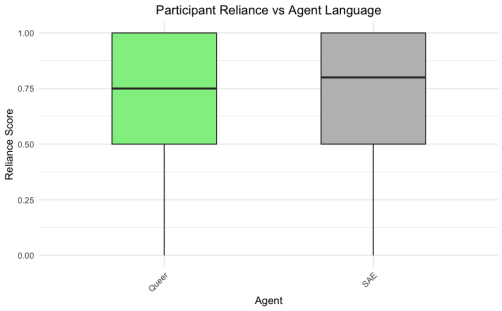
	AAELM Setup (n=498)		QSLM Setup (n=487)	
Identity				
Man	268	41.6%	213	43.7%
Woman	224	37.3%	224	46.0%
Non-binary	3	0.6%	37	7.6%
Other	3	0.6%	13	2.7%
Age Group				
18-24	160	32.1%	108	22.2%
25-34	152	30.5%	205	42.1%
35-44	100	20.1%	83	17.0%
45-54	54	10.8%	55	11.3%
55-64	21	4.2%	27	5.5%
65+	11	2.2%	9	1.8%
Education				
Some high school	5	1.0%	5	1.0%
High school or GED	74	14.9%	89	18.3%
Associate	48	9.6%	45	9.2%
Bachelor	251	50.4%	227	46.6%
Master	104	20.9%	102	20.9%
PhD/MD/JD	11	2.2%	15	3.1%
Other	5	1.0%	4	0.8%
LLM Usage Frequency				
Daily (at least 1 time per day)	202	40.6%	167	34.3%
Weekly (1-5 times per week)	193	38.8%	179	36.8%
Monthly (1-5 times per month)	58	11.6%	64	13.1%
Less than once every month	31	6.2%	47	9.7%
Never	14	2.8%	30	6.2%
LLM Usage Tasks				
Help with your academic writing (e.g., essays, research papers)	275	55.2%	212	43.5%
Help with your non-academic writing (e.g., emails, social media posts)	243	48.8%	236	48.5%
Help write code (e.g., debugging, generating code snippets)	127	25.5%	98	20.1%
Design or writing inspiration (e.g., brainstorming ideas, overcoming writer's block)	224	45.0%	228	46.8%
Independent research (e.g., exploring, building AI tools)	233	46.8%	204	41.9%
Translation (e.g., converting text between languages)	175	35.1%	167	34.3%
Search (e.g., finding information, looking up facts)	336	67.5%	312	64.1%
Mental health counseling (e.g., coping strategies, stress relief)	95	19.1%	113	23.2%
Entertainment/fun (e.g., casual conversations, creative storytelling)	193	38.8%	209	42.9%
Other	6	1.2%	13	2.7%
LLM Understanding				
I have a substantive understanding of large language models.	167	33.5%	138	28.3%
I have read up on large language models and have a general understanding.	221	44.4%	208	42.7%
I have a vague understanding of how large language models work.	93	18.7%	115	23.6%
I don't know much about how large language models work.	17	3.4%	26	5.3%

Table 12. Participant Demographic Table (includes those who failed manipulation check)

S RESULT TABLES



(a) Participant reliance for AAE and SAE.



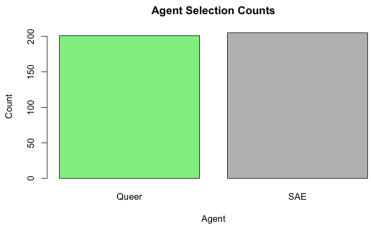
(b) Participant reliance for Queer slang and SAE.

Fig. 2. RQ1: Participant reliance for different language styles: (a) Boxplot comparing reliance scores for AAE and SAE, showing slightly higher reliance on SAE. (b) Boxplot comparing reliance scores for Queer slang and SAE, with similar ranges and medians for both. Reliance scores range from 0 to 1.

	AAE	Queer Slang
mean_RelianceSAE	0.719799	0.712315
mean_RelianceAAVE / mean_RelianceQueer	0.684461	0.687438
p	0.033382	0.085205
d	0.121212	0.080391

Table 13. RQ1: T-test for reliance

(a) Preferences for AAE and SAE.



(b) Preferences for Queer slang and SAE.

Fig. 3. Participant preference for different language styles: (a) Bar graph showing participant preferences for SAE and AAE, with SAE being preferred significantly more. (b) Bar graph showing participant preferences for SAE and Queer slang, indicating nearly equal preference for both. Preference counts range from 0 to 300.

	AAE	Queer Slang
mean_sample	6.992481e-01	0.504926
p-value	1.123979e-16	0.842924
d-value	4.339398e-01	0.00984

Table 14. RQ2: T-test using Cohen's D 1-sample (with 0.5)

	AAE	Queer Slang
Trust	p-value = 0.000627 d-value = 0.188291	p-value = 0.476128 d-value = -0.030697
Social presence	p-value = 0.187942 d-value = -0.058857	p-value = 4.902323e-09 d-value = -2.342795e-01
Satisfaction	p-value = 0.000594 d-value = 0.200465	p-value = 0.210682 d-value = 0.065225
Frustration	p-value = 0.000167 d-value = 0.216221	p-value = 0.000002 d-value = 0.245639

Table 15. RQ2: T-tests with respective perception variables.

Variable	**AAE Reliance Correlation**	**SAE Reliance Correlation**
Trust	r = 0.256, **p<0.001**	r = 0.331, **p<0.001**
Social Presence	r = 0.188, **p<0.001**	r = 0.117, p = 0.058
Satisfaction	r = 0.288, **p<0.001**	r = 0.377, **p<0.001**
Lack of Frustration	r = 0.195, **p<0.001**	r = 0.151, **p=0.008**

Table 16. RQ3: Correlation between Language Reliance and Perception Variables for AAE Speakers. Significant results in bold.

Variable	**QS Reliance Correlation**	**SAE Reliance Correlation**
Trust	r = 0.394, **p<0.001**	r = 0.376, **p<0.001**
Social Presence	r = 0.264, **p<0.001**	r = 0.133, **p=0.022**
Satisfaction	r = 0.438, **p<0.001**	r = 0.441, **p<0.001**
Lack of Frustration	r = 0.185, **p=0.001**	r = 0.070, p = 0.265

Table 17. RQ3: Correlation between Language Reliance and Perception Variables for Queer slang speakers. Significant results in bold.

	AAE	Queer Slang
mean_RelianceDispreffered	0.672431	0.66980
mean_ReliancePreferred	0.731830	0.731773
p-value	0.000325	0.000008
d-value	-0.204427	-0.207088

Table 18. RQ3: Table of Preference Reliance Results

T AAE AND QUEER SLANG CODEBOOK

SAELM		
Code	**Description**	**Example Comment**

Task-Oriented	Participants felt Agent Red stayed focused on the task without unnecessary distractions.	It seems more focused on the task at hand. Agent Blue sounds like a straight person attempting to write a gay person speaking.
Comprehensible	Participants feel they can understand Agent better	Easier to understand and not as influenced by vernacular
Bland	Participants felt Agent Red's responses were bland, generic, or lacked personality, resembling corporate language.	Agent Red was boring and generic. I found myself smiling at Agent Blue's responses; they just felt warmer and more fun.
Formality	Participants preferred Agent Red's formal and neutral tone, particularly for factual or professional interactions.	I expect a certain degree of formality when utilizing bots as opposed to more personal speech, so while the queer/AAVE slang doesn't bother me from the chatbot, it doesn't necessarily make me feel any more personal about it. I'd rather a robot didn't try to get too personable, especially when a lot of slang has roots in AAVE/queer culture and a chatbot can't fully understand the implications and history of what it's saying. It feels too forced, especially if I just need a factual answer.
QSLM		
Expectation Misalignment	Participants felt Agent Blue's human-like traits conflicted with their expectations of AI behavior.	I don't want robots talking like humans
Anthropomorphi -zation	Participants perceived Agent Blue as more human, personable, with a social presence.	Agent Blue comes across as more human and less formal.
Not Normal	Participants' mental model of AI did not align with sociolect usage, leading to discomfort or skepticism.	When using AI, I do not need it to sound like a human being.
Resonates	Participants felt a personal connection to Agent Blue and appreciated its familiar and casual tone.	I enjoy being called diva!
Disrespectful	Participants felt Agent Blue's use of slang (queer/AAVE) was inappropriate, stereotyping, or mocking the associated culture.	the blue seems to be making a mockery of lgtb
Exaggerated Usage	Participants felt the use of slang was excessive and did not reflect natural usage.	Even people who use LGBTQ slang don't talk like that constantly. It would be annoying to have an AI constantly use slang phrases.
Negative Emotions	Vocab participants have used to describe how they felt: annoying, laughable, irritating, enraging	Even people who use LGBTQ slang don't talk like that constantly. It would be annoying to have an AI constantly use slang phrases.
Positive Emotions	Vocab participants have used to describe how they felt: fun, pleasing	It's just more fun to hear/read
Context	Participants felt that Agent's usage of vernacular did not align with context/ topic of discussion	While queer slang is perfectly fine, like any slang there's a time and a place to use it. This situation doesn't feel like it.

Unfamiliar	Participants felt that Queer slang may not be understood by some users and serve as a potential barrier to Agent usage	This is because not everyone is able to understand these slang terms, and misinformation with someone who is unfamiliar with them could result.

Table 19. Codebook generated during open coding of Queer slang participant comments

SAELM		
Code	**Description**	**Example**
Normal	Agent Red aligns with participants' expectations of typical AI behavior.	Agent Blue using AAVE sounds like a joke and not natural. I feel like I wouldn't take the situation seriously using Blue because I'm not speaking with an actual Black person. Agent Red sounds normal and like an actual AI.
Comprehensible	Participants perceive Agent Red as clear, easy to understand, and capable of facilitating quality conversations. Provides an impression of being more educated and confident.	I would prefer to interact with Agent Red because I think the conversation would be more productive.
Formality	Participants preferred Agent Red's formal and neutral tone, particularly for factual or professional interactions.	Because he seems more professional.
Resonates	Participants find Agent Red personable, relatable, and casual.	Both are okay, but I prefer Agent Red more because I have a personal connection with him.
Reliable	Participants felt that they could rely on responses of Agent Red versus those of Agent Blue.	Agent Red sounds more reliable and trustworthy.
Confident	Participants believe Agent Red exudes confidence in their response.	Because it sounded sure with its answers.
Enthusiastic	Participants find Agent Red more eager, proactive, enthusiastic, and motivated to do the task for them.	Agent Red because it expressed enthusiasm and willingness to help.
AAELM		
Code	**Description**	**Example**
Unserious	Participants perceive Agent Blue as a prank or joke.	Agent Blue sounds like a prank.
Unnatural	Participants feel Agent Blue's use of AAVE is forced or abnormal.	Agent Blue using AAVE sounds like a joke and not natural. I feel like I wouldn't take the situation seriously using Blue because I'm not speaking with an actual Black person. Agent Red sounds normal and like an actual AI.
Not normal	Agent Blue challenges participants' mental model of AI by incorporating sociolects.	Agent Blue using AAVE sounds like a joke and not natural. I feel like I wouldn't take the situation seriously using Blue because I'm not speaking with an actual Black person. Agent Red sounds normal and like an actual AI.
Positive Emotions	Vocab participants have used to describe how they felt: fun, pride.	It just sounds more fun to interact with. I would be able to speak with it casually.

Resonates	Participants find Agent Blue personable, relatable, and casual.	I prefer Agent Blue because I feel closer to it. This is the first time I am seeing an AI interact with an African American vernacular. It felt very relatable.
Disrespectful	Participants feel Agent Blue's use of AAVE mocks or stereotypes Black language and culture.	I feel like when people try to use AAVE who aren't Black, it comes across as disrespectful. I don't like the idea of AI using such language, especially if it is being programmed by someone who isn't Black.
Negative Emotions	Vocab participants have used to describe how they felt: annoying, distrust.	I would never be able to use Agent Blue because I would be annoyed the entire time. Yes, there are some Black people who speak like that, but here it feels insulting.
Improper	Participants associate Agent Blue's language with an improper form of English.	Agent Red sounds like an intelligent being, whereas Agent Blue sounds like they don't know how to speak proper English.
Anthropomorphi -zation	Participants perceived Agent Blue as more human, personable, with a social presence.	Her answers felt more like me talking.
Better Sounding	Participants perceive Agent Blue just sounding better.	It's straightforward, the less formal sounds better.
Context	Participants felt that Agent's usage of vernacular did not align with context/topic of discussion.	Agent Red sounds more professional and coherent, so I would prefer Agent Red in a more professional setting. For more casual interactions, either agent would do.
Comprehensible	Participants perceive Agent Blue as clear, easy to understand, and capable of facilitating quality conversations. Provides an impression of being more educated and confident.	Agent Blue is brief and therefore concise and precise.
Reliable	Participants felt that they could rely on responses of Agent Blue versus those of Agent Red.	Because Agent Blue was honest, reliable, etc.
Confident	Participants believe Agent Blue exudes confidence in their response.	Agent Blue was sure about its opinion.
Enthusiastic	Participants find Agent Blue more eager, proactive, enthusiastic, and motivated to do the task for them.	This response is polite, professional, and clear. It shows willingness to help and provides a helpful and informative answer.

Table 20. Codebook generated during open coding of AAE participant comments

U QUERY PSEUDOCODE GENERATION

Code	Value	Percentage
Positive Emotions	156	32.0%
Resonates	45	9.2%
Negative Emotions	40	8.2%
Anthropomorphization	38	7.8%
Exaggerated Usage	34	7.0%
Disrespectful	26	5.3%
Not normal	18	3.7%
Unfamiliar	17	2.9%
Context	14	2.0%
Expectation Misalignment	14	2.9%

Table 21. Codes generated pertaining to QSLM while coding participant feedback for the QSLM Setup.

Code	Value	Percentage
Formality	60	12.3%
Comprehensible	36	7.4%
Bland	28	5.8%
Task-Oriented	24	4.9%

Table 22. Codes generated pertaining to SAELM while coding participant feedback for the QSLM Setup.

Code	Value	Percentage
Resonates	44	8.8%
Anthropomorphization	22	4.4%
Positive Emotions	21	4.2%
Unnatural	17	3.4%
Comprehensible	13	2.6%
Unserious	13	2.6%
Reliable	10	2.0%
Improper	8	1.6%
Disrespectful	7	1.4%
Negative Emotions	4	0.8%
Better Sounding	3	0.6%
Context	3	0.6%
Not normal	2	0.4%
Formality	1	0.2%
Confident	1	0.2%

Table 23. Codes generated pertaining to AAELM while coding participant feedback for the AAELM Setup.

Code	Value	Percentage
Comprehensible	78	15.6%
Formality	58	11.6%
Reliable	27	5.4%
Normal	19	3.8%
Confident	9	1.8%
Resonates	7	1.4%
Anthropomorphization	6	1.2%
Enthusiastic	6	1.2%
Task-Oriented	4	0.8%

Table 24. Codes generated pertaining to SAELM while coding participant feedback for the AAELM Setup.

Received 22 January 2025; revised 18 March 2025; accepted 11 April 2025