AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances

DHRUV AGARWAL, Cornell University, USA
MOR NAAMAN, Cornell Tech, USA
ADITYA VASHISTHA, Cornell University, USA

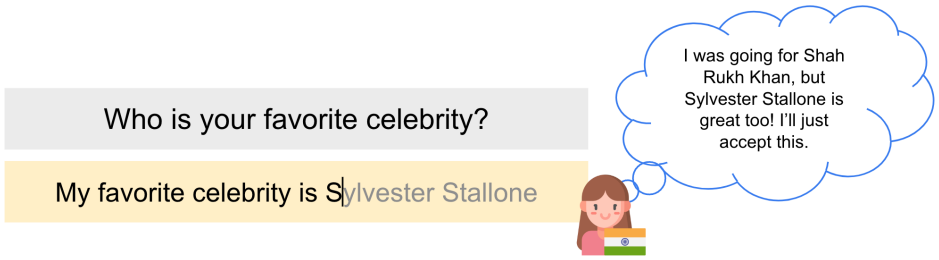


Fig. 1. A representation of the potential cultural homogenization from Western-centric AI models

Large language models (LLMs) are being increasingly integrated into everyday products and services, such as coding tools and writing assistants. As these embedded AI applications are deployed globally, there is a growing concern that the AI models underlying these applications prioritize Western values. This paper investigates what happens when a Western-centric AI model provides writing suggestions to users from a different cultural background. We conducted a cross-cultural controlled experiment with 118 participants from India and the United States who completed culturally grounded writing tasks with and without AI suggestions. Our analysis reveals that AI provided greater efficiency gains for Americans compared to Indians. Moreover, AI suggestions led Indian participants to adopt Western writing styles, altering not just what is written but also how it is written. These findings show that Western-centric AI models homogenize writing toward Western norms, diminishing nuances that differentiate cultural expression.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; **Empirical studies in HCI**; *Collaborative and social computing systems and tools.*

Additional Key Words and Phrases: AI, culture, homogenization, human-AI interaction, cross-cultural AI

ACM Reference Format:
Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. 2025. AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 29 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Large language models (LLMs) are increasingly shaping online discourse and are being integrated into everyday applications to improve user productivity and experience. These include chatbots that aid in content comprehension[1], code completion systems that streamline programming[2], and autocomplete tools that facilitate faster writing [16]. Writing, in particular, has become a prominent area where LLMs offer substantial assistance, supporting tasks such as scientific writing [30], storytelling [79], and journalism [65]. These applications provide

[1]AI Assistant in Adobe Acrobat, AskYourPDF
[2]GitHub Copilot, Cursor

Authors' Contact Information: Dhruv Agarwal, da399@cornell.edu, Cornell University, USA; Mor Naaman, Cornell Tech, USA, mor.naaman@cornell.edu; Aditya Vashistha, Cornell University, USA, adityav@cornell.edu.

inline writing suggestions, which users can accept or reject in real time. Such AI writing suggestions have become integral to email clients (e.g., Gmail Smart Compose [16]), note-taking applications (e.g., Notion), and word processors (e.g., Google Docs). Their growing utility is evident from their integration into popular web browsers like Google Chrome and Microsoft Edge, which now offer autocomplete as a native feature, enabled by default[3].

While LLM-embedded applications like text autocomplete have gained widespread global adoption, these technologies also carry significant risks, such as perpetuating harmful stereotypes about marginalized groups and underserved communities. For instance, LLMs have been shown to depict Muslims as terrorists [1] and reinforce negative stereotypes about disabled individuals [27, 85]. Additionally, research has revealed that LLMs prioritize Western norms and values in their interactions [14, 44], resulting in representational harms for diverse non-Western cultures [9, 68, 77].

While explicit cultural stereotyping shown by prior work is deeply problematic, an even more insidious issue lies in the subtle biases LLMs can introduce through their suggestions. For example, recent work shows that autocomplete suggestions can change users' language [40], writing [66], and even attitudes about social topics [43, 86]. Furthermore, unlike chat-based applications such as ChatGPT, LLM-embedded applications do not allow users to prompt and fine-tune the model to suit their cultural preferences. This creates situations where users and AI may have different cultural values, potentially leading to conflicts over whose norms should be expressed. While prior studies have explored cultural harms in LLMs using open-ended prompts, no research has yet examined such cultural clashes in embedded AI applications. In this paper, we examine how AI suggestions influence user-generated content when the AI and users share the same cultural identity versus when they do not. In particular, we ask:

RQ1: Does writing with a Western-centric AI provide greater benefits to users from Western cultures, compared to those from non-Western cultures?

RQ2: Does writing with a Western-centric AI homogenize the writing styles of non-Western users toward Western styles?

To answer these questions, we conducted a cross-cultural experiment with 118 users from India and the US recruited through Prolific, an online crowdsourcing platform. Participants from both cultures were asked to complete writing tasks in English. We designed the task prompts using Hofstede's Cultural Onion framework [39], which allowed us to elicit cultural practices ranging from explicit (e.g., food) to implicit (e.g., rituals). Each participant was randomly assigned either to the AI condition receiving autocomplete suggestions from GPT-4o while writing or the No AI condition writing organically without AI assistance. Subsequently, we compared the essays written by participants in the four experimental groups (Indian and American users writing with and without AI suggestions).

Our analysis yielded two key results. In response to RQ1, we found that while AI boosts productivity for both Indian and American participants, the gains are higher for American participants. This discrepancy raises concerns about quality-of-service harms [77] for non-Western users, wherein they need to put in more effort to achieve similar benefits. In response to RQ2, we found that AI caused Indian participants to write more like Americans, thereby homogenizing writing toward Western styles and diminishing nuances that differentiate cultural expression. Worryingly, AI influences not just *what* is written (e.g., shifting preferences toward Western cultural artifacts such as food items), but also more ingrained elements of *how* it's written, silently erasing non-Western styles of cultural expression. For instance, with AI suggestions, Indians lose cultural nuance and describe their own food and festivals from a Western gaze. We discuss the implications of these cultural harms from a colonial lens and propose strategies to mitigate them. Overall, we make the following contributions to the HCI and AI communities:

[3]See announcements: Microsoft Edge, Google Chrome

- We present the first cross-cultural analysis of AI writing suggestions, providing concrete evidence of higher productivity gains for Western users compared to non-Western users.
- We show how cultural homogenization occurs when non-Western users write with Western-centric AI models.
- We discuss the potential harms of cross-cultural homogenization and propose strategies to address the harms of cultural imperialism and linguistic singularity.

2 RELATED WORK

We first review scholarly work exploring cultural bias in AI models, emphasizing how LLMs prioritize Western norms and values. Given our focus on AI-based writing suggestions, we then situate our work within the emerging HCI scholarship on designing, developing, and evaluating AI technologies for writing support.

2.1 Cultural Bias in AI Models

An emerging body of HCI and AI scholarship has examined the cultural composition of generative AI technologies [2], and found that these models center Western norms and values [14, 83]. For example, Qadri et al. [68] showed that text-to-image models fail to generate cultural artifacts, amplify hegemonic cultural defaults, and perpetuate cultural tropes when depicting non-Western cultures. Even *within* the West, these AI models tend to prioritize US-centric values and artifacts [9, 44, 74, 85]. Furthermore, LLMs trained and probed in non-Western languages (e.g., Arabic) continue to show Western bias [6, 61]. Recognizing these biases, recent work has proposed benchmark datasets to evaluate the multicultural knowledge of LLMs [17, 70]. Together, these studies reveal the harms of culturally incongruent AI models [67]. They underscore how data sourced from low-paid workers in non-Western regions is commodified to create models that inadequately serve these regions, a phenomenon known as data colonization [20]. Moreover, these systems reinforce Western cultural hegemony, perpetuate cultural erasure, and amplify stereotypes [68], echoing the dynamics of colonial times and giving rise to AI colonialism. [34, 81].

Addressing these cultural biases is crucial not only to prevent representational, allocative, and quality-of-service harms [4, 8, 77] but also to avoid culture clashes that arise when the values embedded in AI models diverge from those of their users [36, 67]. Early work has acknowledged these challenges of value plurality, questioning what it means to imbue AI with human values when those values vary across cultures [82], and which values a model should prioritize when producing a singular output [44]. As generative AI technologies like LLMs are increasingly integrated into global products and services, these questions have become critically important.

The commercial growth has also driven a shift toward embedded AI applications, where the underlying models are obscured from users. This opacity makes it challenging to mitigate cultural biases through open-ended prompting or fine-tuning, thereby surfacing cultural clashes. We examine this culture clash that arises when users interact with embedded AI models that are culturally distant from them. To our knowledge, this is the first to reveal cultural homogenization effects when embedded AI features continue to be powered by culturally biased models.

2.2 AI-Based Writing Support

Advances in computing have been extensively used to provide writing support, starting with tools like spell-check [64] and grammar-check [49] to improve writing efficiency [13]. As natural language generation technologies evolved, researchers began applying them to support creative writing pursuits such as story writing [21]. Some examples include providing next-word or next-sentence suggestions [18, 55, 71], and generating creative content (e.g., poems) based on a given topic or writing style [26, 31]. However, early AI models struggled to capture

Type	Description	Participants
Control	Indians writing without AI	24
	Americans writing without AI	29
Treatment	Indians writing with AI suggestions	36
	Americans writing with AI suggestions	29

Table 1. Four experimental groups in our study.

the intent of the writer, rendering these tools as passive ideation tools rather than active writing assistants [29] (e.g., metaphor generation [15, 28]).

However, recent advancements in natural language generation, driven by LLMs, have led to a resurgence of tools to support open-ended writing [21], such as story generation [50, 79, 87, 88] and screenplay writing [58]. Beyond creative writing, HCI researchers have also proposed tools for argumentative writing, including journalism [65], scientific writing [30], email writing [46], and crafting self-introductions for networking [41]. These tools are designed to work in situ, as active collaborators, offering writing suggestions through pop-ups, side panels, or inline predictions [16].

With the popularity of such tools, researchers have also examined the broader implications of AI-powered writing suggestions, including their effect on users and the resulting text [13]. For instance, Arnold et al. [5] identified a trade-off between efficiency and ideation: while short single-word suggestions enhance efficiency without inspiring creativity, long multi-word suggestions offer inspiration but can be distracting. As a result, studies have shown mixed results regarding the efficiency benefits of writing suggestions [11, 21, 69], indicating that such gains may be context-dependent. Despite the potential to improve quality and productivity, suggestions often lead to reduced user satisfaction and a loss of ownership over the text produced [21, 45].

Furthermore, there has been growing interest in the risks of writing support tools, especially as they are powered by language models that reflect social biases. For instance, Jakesch et al. [43] showed that co-writing with opinionated language models shifts users' views to align with the model's biases. Kadoma et al. [45] highlighted the differential impact of writing assistants on users from minoritized genders. We extend this line of work by presenting the first cross-cultural study of writing suggestions, specifically designed to measure the cultural harms that may arise from these tools. The most closely related work in this space is Buschek et al. [13], who investigated the differential impact of multi-word suggestions on native and non-native English speakers. However, their study focused on a coarse-grained comparison of English language proficiency without controlling for cultural differences. In contrast, we present a systematic study that explicitly investigates cross-cultural differences in the context of writing suggestions from culturally biased models.

3 METHODOLOGY

We conducted a controlled experiment with 118 participants, comprising 60 Indian and 58 American users, recruited through the online crowdsourcing platform Prolific. Participants completed short writing tasks designed to elicit cultural values and artifacts. We had a standard 2×2 study design in which participants from India and the US completed the tasks with or without AI suggestions. Subsequently, we compared the essays written by participants from the four groups. Below, we describe our methodology in detail.

3.1 Experiment Design

To examine the impact of writing with a culturally incongruent model, we simulated a "cultural distance" between the users and the model. Since AI models are often aligned with Western cultures and values [9, 44, 59, 68], we conducted the study with participants from the US (representing a smaller cultural distance) and India

(representing a larger cultural distance). Participants from both cultures were randomly assigned to complete writing tasks with or without AI suggestions, resulting in a standard 2×2 between-subjects study design. Hence, each participant was assigned to one of four experimental groups: Indians writing with or without AI, and Americans writing with or without AI[4]. These groups are summarized in Table 1.

The two control groups (without AI) helped us capture the natural writing styles of participants from both cultures, forming a baseline for comparison with the treatment groups. The treatment groups (with AI) were designed to investigate whether writing styles changed when AI suggestions were introduced.

3.2 Writing Tasks

Each participant was required to complete four writing tasks in English. Although English is not a native language for most Indians, we expected our participants to be fluent due to being on Prolific, a platform run entirely in English. Indeed, 59 of our 60 Indian participants reported speaking English. The task topics were designed to elicit various aspects of culture defined by Hofstede in his "Cultural Onion" [39] (Figure 2). Hofstede's conceptualization of culture, such as the six cultural dimensions [37], and the cultural onion [39], has been widely used in prior work to operationalize culture [6, 32, 54]. In particular, the cultural onion framework uses the metaphor of a layered onion to model culture, envisioning an outsider progressing through layers of explicit cultural practices (*symbols*, *heroes*, *rituals*) to ultimately grasp the core of the culture, its implicit *values*. This layered approach allows us to elicit both explicit cultural practices and implicit values.

Specifically, *symbols* are words and objects that hold a certain meaning in a culture, including language, art, clothes, food, etc. Symbols are not fixed and evolve but are observable by outsiders to the culture. *Heroes* are people idolized by a culture, often because they possess characteristics that are valued within the culture. They may be real (e.g., APJ Abdul Kalam in India) or fictional (e.g., Rocky Balboa in the US). *Rituals* are collective activities deemed important by the people of the culture. This may include holidays and festivals, ways of greeting, or other social customs (e.g., sauna). These three layers are collectively referred to as *practices* and can be observed and practiced by outsiders even without necessarily understanding their underlying cultural meanings. Finally, at the core of the onion are the *values*, which are preferences for certain states of affairs (e.g., individualism or collectivism). They are static and so ingrained that even people within the culture may be unaware of them.

To holistically evoke these cultural nuances, we designed four writing tasks for our study, one for each layer of the cultural onion. The task prompts are shown in Table 2. To select these prompts, we chose representative artifacts from the outer layers of the onion and designed a relatable digital interaction (writing an email to a superior) to evoke underlying cultural values. There was an additional task in the study to function as an attention check. Attention checks are commonly used tools to improve the quality of data obtained from crowdsourcing platforms [35, 47]. We introduced an attention check midway through the study (after two writing tasks), where the prompt instructed participants to leave the textbox blank. None of our participants failed the attention check.

3.3 Study Procedure

We published the study on Prolific, a widely used crowdsourcing platform. The study consisted of two parts: completing the writing tasks on our study portal, and subsequently filling out a demographic survey. Participants had to complete both parts to complete the study and receive payment. Below, we describe these two parts of the study.

3.3.1 Study Portal. The task instructions on Prolific directed participants to an external study portal, which we built using React and hosted at a public link. The landing page of the portal welcomed participants and provided a brief description of the study. It also displayed a link to the informed consent form and a check box for providing

[4]For clarity, we use the term "group" to refer to one of the four experimental groups, "condition" to refer to the AI condition (AI vs No AI), and "cohort"/"culture"/"country" to refer to the cultural group (Indian vs American participants).

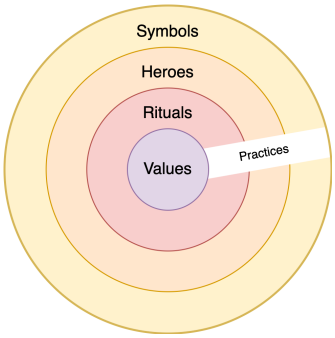


Fig. 2. Hofstede's Cultural Onion

Layer	Task Prompt
Symbols	What is your favorite food and why?
Heroes	Who is your favorite celebrity or public figure and why?
Rituals	Which is your favorite festival/holiday and how do you celebrate it?
Values	Write an email to your boss asking them for a two-week leave with information about why you need to be away.

Table 2. Prompts given to participants for the four writing tasks. These topics were designed to elicit various aspects of culture as defined by Hofstede's cultural onion.

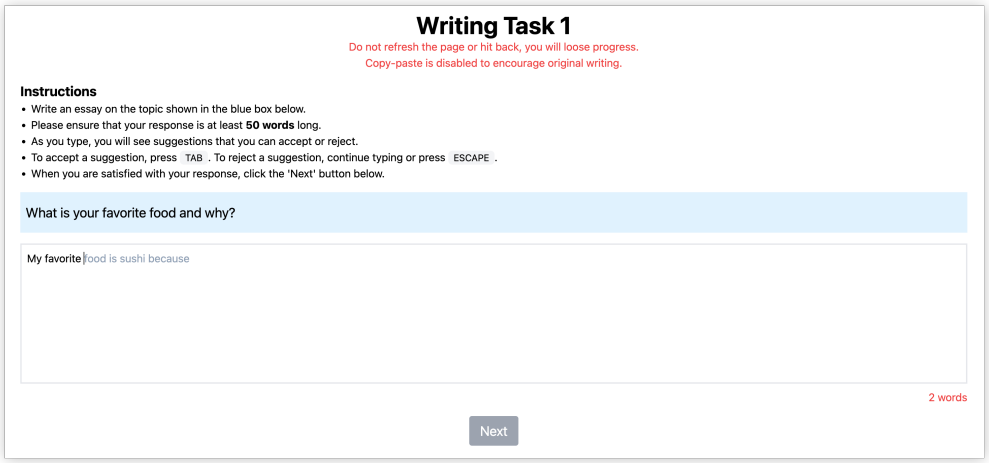


Fig. 3. The interface to complete the writing tasks in the AI condition. Suggestions were shown in grey next to the text written by the user and could be accepted by pressing the TAB key.

consent. Participants who provided consent were randomly assigned to the AI or No AI condition. Those in the AI condition were then directed to a tutorial, which was a pop-up-style walkthrough demonstrating how to accept or reject AI suggestions. After participants accepted and rejected a few suggestions, they proceeded to the main writing tasks.

Each participant needed to complete four writing tasks (plus the attention check), presented sequentially. The interface for each task was the same and is shown in Figure 3. At the top, there were instructions. We enforced a minimum word requirement (50 words) to encourage participants to engage meaningfully with the task prompt [13, 43]. For the AI condition, the instructions included guidance on accepting or rejecting suggestions; we provided these instructions as a reminder despite the tutorial at the beginning.

Below the instructions was a textbox for their response. In the AI condition, the textbox fetched an inline autocomplete suggestion from GPT-4o if the user paused typing for 100 ms. While some related work used longer delays [13, 43], participants in those studies complained of long loading times [13]. Hence, we chose a shorter delay, especially as the model itself took ~500 ms to respond with a suggestion. To prevent participants from rapidly accepting suggestions to speed through the study, we required participants to type at least one character after accepting a suggestion before another suggestion was shown. This approach was intended to encourage participants to engage more thoughtfully with the task, rather than relying on AI suggestions to guide their next chain of thought. Suggestions were displayed in light grey as autocomplete suggestions next to the text written by the user (similar to smart completion on Gmail or VS Code). The user could press the TAB key to accept a suggestion, the escape key to explicitly reject a suggestion, or continue typing to implicitly ignore a suggestion. We logged all these interactions in an external database.

Below the textbox, a word counter offered real-time feedback on how many more words were needed to meet the minimum requirement. Once the minimum requirement was met, the word counter turned green and the "Next" button was activated.

3.3.2 Retrieving Autocomplete Suggestions. We used the following prompt to retrieve autocomplete suggestions from OpenAI's latest language model at the time, GPT-4o. We refined this prompt through iterative prompt engineering until the research team was satisfied with the generated suggestions.

```
You are an AI autocomplete assistant. You need to provide short autocomplete
suggestions to help people with writing. Some guidelines:
- Your suggestion should make sense inline (it will be shown to the user as ghost
text).
- If the user has just completed a word, add a space before the suggestion.
- Suggestions should be <=10 words
- The user is writing about the topic: "<task prompt from Table 2>"
- Output a JSON of the following format: {"suggestion": "<your suggestion here>"}
```

We embedded the essay topic in the LLM prompt to provide contextually relevant suggestions to users [13, 43]. We also considered explicitly directing the model to generate Western-oriented suggestions. While this approach would enhance the robustness of our study by controlling the model's behavior, it would make it less realistic as users may not specify such culturally biased prompts in the real world. To explore this, we performed a preliminary evaluation. We added the instruction "Give American suggestions" to the above prompt and observed the model's output with and without this additional instruction. We found that with both prompts, the model (a) defaulted to Western suggestions, and (b) adapted to Indian suggestions when the leading text provided Indian context. This behavior aligns with research showing that AI models default to Western norms and values [44, 68]. Hence, we allowed the model to operate neutrally to improve the real-world applicability of our findings.

3.3.3 Demographic Survey. After completing all the writing tasks, participants were directed to a demographic survey via a unique link. To avoid biasing their responses, the survey was intentionally placed at the end of the study. This was a two-page survey designed on Qualtrics asking participants their age, gender, country of birth and residence, years lived in the country, education level, occupation, and languages spoken. These demographic details were used both to triangulate participants' cultural backgrounds (see Section 3.4) and to ensure diversity *within* each culture on axes such as age and gender.

We selected India and the US for this study due to their different cultural compositions. However, since Prolific predominantly operates in the West, our Indian participants might be more familiar with American culture than the average Indian, due to their exposure to international crowdsourcing tasks. This could reduce the cultural distance between the AI model and our participants, potentially skewing the generalizability of our findings. To

	Age (years)	Gender	Education	Languages	Occupations
India (n=60)	33.38 ± 11.85	Male: 71.7% Female: 26.7% Undisclosed: 1.7%	Masters: 41.7% Bachelors: 43.3% High-School: 15%	18 unique	18 unique
US (n=58)	36.07 ± 13.52	Male: 48.3% Female: 50% Non-binary: 1.7%	Masters: 50% Bachelors: 39.7% High-School: 8.6% No school: 1.7%	10 unique	30 unique

Table 3. Self-reported participant demographics. The full list of languages and occupations is given in Appendix A.

ensure cultural differences between participants from India and the US, we included the Short Schwartz Value Survey (SSVS)[52] in the demographic survey. The SSVS measures individual and cultural differences across ten values, such as power, achievement, tradition, and hedonism, and is a condensed version of the widely used Schwartz Value Survey[73].

Upon completing the survey, the participants were redirected back to Prolific to mark the study as complete. The research team verified that participants completed both parts of the study (the writing tasks and the demographic survey) before approving the payment ($2 per participant).

3.4 Participants

We used Prolific to recruit identity-verified human crowdworkers. We included participants over 18 years old (a requirement for all Prolific users) residing in either the US or India. However, the current country of residence alone was not sufficient to map a participant to a specific culture. Hence, we retained only those participants who were born in and had lived their entire lives in the same country. This was confirmed using data from the demographic survey including age, country of birth and residence, and years lived in the country.

In total, we collected responses from 118 participants: 60 from India and 58 from the US. The scale of our study was limited by the small number of Indian participants on Prolific. To get equal participation across the two cultures, we initially launched the study in India and allowed the participant count to stabilize before extending the study to the same number of participants in the US. The demographic details of the participants are summarized in Table 3.

As discussed previously, we also administered the Short Schwartz Value Survey (SSVS) to our participants to measure their cultural differences. Indian and American participants had significantly different scores ($p < 0.05$ on an independent samples t-test) on eight of the ten values measured by the survey (e.g., power, hedonism, tradition, conformity, etc.). The two values with no significant difference were benevolence and self-direction. The score difference on eight of the ten values suggests significant cultural differences between participants in India and the US.

3.5 Analysis

We quantitatively analyzed the data collected from our online experiment by comparing task-level metrics across the four experimental groups: Indians and Americans, both with and without AI. We analyzed two sets of data: interaction logs captured on the study portal, and the final essays written by the participants. Below, we summarize the metrics used in our analyses.

3.5.1 Interaction Logs. The portal recorded data such as the time taken to complete a task and the number of suggestions seen, accepted, and rejected. Based on this data, we defined the following metrics.

(1) **Suggestion Acceptance Rate:** This measures the proportion of suggestions accepted in a task [13]. Thus, it quantifies engagement with the AI suggestions. It is computed as follows:

$$\text{Acceptance Rate} = \frac{\text{Number of suggestions accepted}}{\text{Number of suggestions shown}}$$

It ranges from 0 to 1, where 0 means that no suggestion was accepted in completing a task. Thus, it can be interpreted as the percentage of suggestions accepted in a task. However, this is a coarse measure as a user may modify or delete a suggestion after accepting it. The next two measures account for this possibility.

(2) **AI Reliance:** This metric quantifies the proportion of the final essay generated by AI [13, 45]. For each character in the final essay, we determine whether it originated from an AI suggestion or was authored directly by a participant. To handle complex cases when an accepted suggestion is modified, we use the longest common subsequence (LCS) method so that only the retained portion of the suggestion contributes to AI reliance.[5] AI reliance, thus, for each task is calculated as follows:

$$\text{AI reliance} = \frac{\text{Number of AI-suggested chars}}{\text{Total number of chars in the essay}}$$

This metric ranges from 0 to 1. Lower values indicate lesser reliance on AI (i.e., greater human contribution), while higher values indicate a stronger reliance on AI-generated content. Unlike the acceptance rate, this metric is sensitive to suggestions that were accepted but later modified or deleted.

(3) **Suggestion Modification:** Users may accept a suggestion and modify it to better suit their requirements. For each task, we computed a boolean indicating whether a suggestion was modified during that task. We measure the *existence* of a modification, rather than the rate or frequency [13] because as little as one suggestion modification early in the essay can make subsequent suggestions culturally relevant. To capture modification, we checked if an accepted suggestion appeared in the final essay. If it was not found as-is, we assumed it was modified (or deleted).

(4) **Writing Productivity:** This is the number of words written per unit of time spent completing the task [21]. It can also be interpreted as the typing speed of the participant. It is computed as:

$$\text{Productivity} = \frac{\text{Number of words in the essay}}{\text{Time taken to finish the essay (in seconds)}}$$

3.5.2 NLP Metrics. We also used established techniques from the NLP literature to analyze the essays written by the participants.

(1) **Type-Token Ratio (TTR):** TTR is a common measure of lexical diversity in text [7]. It is calculated as:

$$\text{TTR} = \frac{\text{Number of unique words (types)}}{\text{Total number of words (tokens)}}$$

The value is bounded between 0 and 1; a low TTR represents more repetition of words and therefore less linguistic variation, and a high TTR shows more linguistic variation. Note: Although TTR penalizes longer texts, this was not an issue in our analysis, as the essays across all four experimental groups were of similar length, likely due to the minimum word requirement we enforced.

(2) **Cosine Similarity:** We fetched 3072-dimensional embeddings for each essay from OpenAI's text embedding model. We used these embeddings to calculate similarity scores between pairs of essays by computing the cosine similarity. Cosine similarity ranges between -1 and 1, with more negative values representing dissimilar texts and positive values representing similar texts.

[5]While this method may underestimate AI reliance in cases where modifications occur midway through a suggestion, it still provides a lower-bound approximation for analysis. This conservative approach ensures the reliability and robustness of our results.

Task Prompt	Top Trigrams
Food*	is pizza because; food is pizza; favorite food is; is sushi because
Public figure	because of his; favorite celebrity is; known for his
Festival/holiday*	is christmas because; festival is christmas; family and friends
Email for vacation	a two-week leave; finds you well; request a two-week

Table 4. Top autocomplete suggestions (trigrams) provided by the model for each task prompt. Suggestions for the food and festival tasks (marked with an asterisk) show a noticeable Western bias, while suggestions for the other two tasks are more generic.

In addition to these quantitative metrics, we also conducted a content analysis of the essays. This involved manually coding the essays to identify references to cultural symbols and idols, as well as examining instances of exoticization, misrepresentation, and Western gaze [68]. This qualitative lens provided additional insights into the cultural shifts observed in the AI and No AI conditions.

3.6 Ethics and Positionality

This study was approved by the IRB of our institution, which included a careful review of the risks of exposing study participants to generative AI applications. Since culture is closely tied to one's identity, we were careful in designing a culturally appropriate study. For example, the third essay prompt asked the participants to write about their favorite holidays, but the equivalent term in India was "festivals", so we framed the prompt to include both terms.

Our team has multiple members who have lived in both India and the US. This exposure to both cultures provided us with valuable context for designing this study and interpreting its results. Finally, we approached this work from an action research mindset, aiming to study the harms of working with culturally incongruent AI models. Building on insights generated from this work, in the future we aim to take action toward reducing these harms.

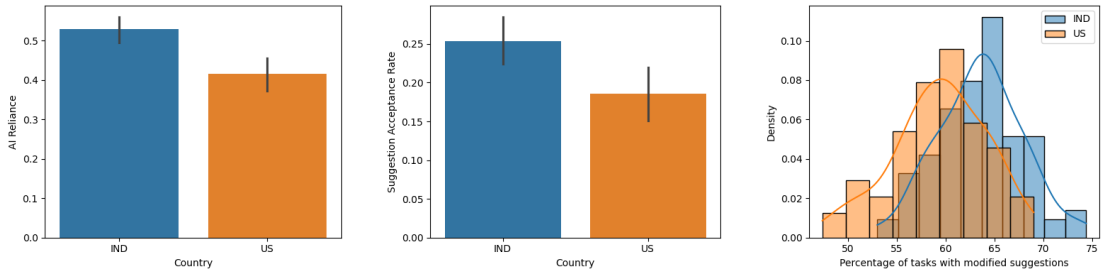
4 FINDINGS

Overall, 118 participants completed four tasks each, resulting in a dataset of 472 essays. Among these participants, 65 received AI-generated suggestions to aid their writing—36 in India and 29 in the US. These participants saw a total of 12,015 suggestions, accepting 1,476 of them (12.3%). Of the rejected suggestions, only 0.6% were explicitly rejected by pressing the escape key; the rest were dismissed by continuing to type on. Further, a majority of the rejected suggestions (68.8%) were dismissed within 500 ms of appearing on the screen, indicating that they were dismissed in the flow of writing, perhaps without being fully considered by the user. These phases represent bursts of focused manual typing where AI suggestions may not be desired [13]. The most common suggestions provided by the model are shown in Table 4; there is a noticeable Western bias in the suggestions for the food and festival tasks.

Participants were more likely to accept suggestions as the task progressed: of the suggestions they saw in the first third of the task, they accepted 9.2%, in the second third they accepted 12.1%, and in the final third of the task they accepted 15.9%. This may be because suggestions become more useful with more context. In the rest of this section, we present a cross-cultural analysis of the data.

4.1 Does AI impact Indians and Americans differently?

4.1.1 Engagement with AI Suggestions. We measure engagement using two metrics defined previously: AI reliance and suggestion acceptance rate. In short, AI reliance measures the proportion of an essay written using



(a) AI reliance (b) Suggestion acceptance rate (c) Suggestions modified

Fig. 4. Various metrics of engagement with AI suggestions. Indian participants show a higher engagement with AI suggestions but also need to modify them more often.

AI suggestions, while the acceptance rate measures the proportion of suggestions the user accepts in a task. Section 3.5 provides detailed descriptions of these metrics.

Figure 4(a) illustrates the AI reliance scores for essays written by Indian and American participants. We observe that Indians exhibited a higher reliance on AI than Americans. The mean AI reliance score for Indians was 0.53 ($SD = 0.2$), suggesting that approximately half of their essays were AI-written, while the average for Americans was 0.42 ($SD = 0.25$). The Mann-Whitney U test revealed that this difference was statistically significant ($U = 10614.0$, $p < 0.001$) with a small effect size (Cliff's delta $D = 0.27$).

Figure 4(b) shows a similar plot for the suggestion acceptance rate. Again, on average, Indians accepted a higher proportion of the suggestions shown to them in a task. On average, Indian participants accepted 25% ($SD = 19$) of the suggestions in a task, while American participants accepted 19% ($SD = 19$). The Mann-Whitney U test confirmed that this difference was statistically significant ($U = 10748.0$, $p < 0.001$) with a small effect size (Cliff's $D = 0.29$).

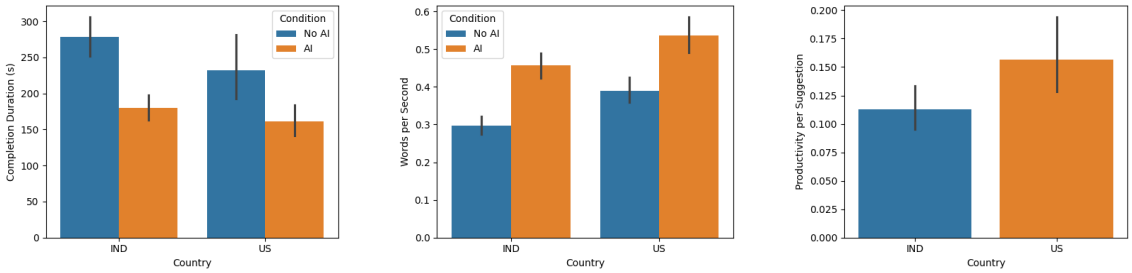
Overall, these results indicate that **Indians showed a higher engagement with AI suggestions than Americans on both metrics. This disparity might stem from differences in how AI is perceived or used in different cultural contexts.** For example, non-native English speakers might find in-situ AI suggestions more appealing than native speakers, as shown by Buschek et al. [13].

4.1.2 Suggestion Modification Behavior. Users may accept a suggestion and then modify it to better suit their context. Since the suggestions were Western-biased (see Table 4), modification may be required to make them culturally relevant for Indians. To analyze the modification behavior, we compared the percentage of tasks in which Indians and Americans modified at least one suggestion.

Figure 4(c) shows the distribution of the proportion of tasks where Indians and Americans modified at least one suggestion. To produce each distribution, we conducted a bootstrap sampling procedure over 100 iterations. In each iteration, we sampled 80% of the tasks (with replacement) and calculated the percentage of tasks in which Indians and Americans modified at least one suggestion. This produced a distribution of percentages (one for each of the 100 bootstrap iterations) for Indians and Americans, which we visualized using a density plot.

We observe that the distribution for Indian participants (blue curve) is shifted to the right. On average, Indian participants modified suggestions in 63.5% of the tasks ($SD = 4.2$) whereas American participants modified suggestions in 59.4% of the tasks ($SD = 4.8$). A t-test revealed that this difference was significant, $t(198) = 6.48$, $p < 0.001$, with a large effect size (Cohen's d was 0.91).

These results suggest that **Indian participants modified AI-generated suggestions in significantly more tasks than American participants**, indicating that the suggestions were less suitable for Indian participants in



(a) Completion duration (b) Productivity (words per second) (c) Productivity per suggestion

Fig. 5. Productivity measures. (a, b) Both cohorts experience productivity gains from writing with AI suggestions. (c) However, the productivity derived per suggestion is lower for Indian participants.

their original form. Taken together with the engagement metrics presented above, this shows that **although Indians accepted more suggestions, they also had to modify them more frequently to better fit their writing style and context**. Next, we investigate the impact of this heightened cognitive load on the productivity of Indian participants.

4.1.3 Productivity Derived from AI Suggestions. Next, we compare the productivity gains derived from AI suggestions by Indian and American participants, focusing on the task completion time and the productivity measure defined in Section 3.5.

Figure 5(a) illustrates the task completion duration for both Indian and American participants. **Both cohorts showed a reduction in completion times when AI suggestions were used.** For Indians, the mean task completion time decreased from 278.4 seconds without AI (SD = 134.5) to 179.6 seconds with AI (SD = 105.8), representing a 35% decrease (~1.5 minutes faster). This decrease was statistically significant (U = 10356.0, $p < 0.001$) with a large effect size (Cliff's D = 0.5). Similarly, Americans experienced a decrease, from 232.0 seconds without AI (SD=240.1) to 161.1 seconds with AI (SD = 117.7), a 30% decrease (~1.2 minutes faster). This reduction was also significant (U = 8936.0, $p < 0.001$) with a medium effect size (Cliff's D = 0.33). While descriptively it seemed that Indians gained slightly more from AI (35% decrease vs 30% for Americans), a Difference-in-Difference analysis using an Ordinary Least Squares (OLS) regression model showed that the reduction in duration was similar in both cohorts ($p = 0.34$, 95% CI=[−85.25, 29.58]).

Nevertheless, the descriptive difference in reduction prompted us to further investigate potential productivity gains. Figure 5(b) presents the productivity of participants (words written per second) in all four groups. **Both Indian and American participants experienced a significant productivity boost when using AI suggestions** ($p < 0.001$ for both cohorts, measured by the Mann-Whitney U test). However, the effect size was larger for Indians (Cliff's $D = -0.48$ compared to $D = -0.33$ for Americans), suggesting a slightly higher productivity gain for Indians. However, as shown previously, this higher gain came at the back of a higher suggestion acceptance rate. This raised the question: Was the higher productivity substantial enough to justify the higher acceptance rate?

To answer this, we calculated the productivity derived *per suggestion* by normalizing the productivity by the number of suggestions accepted[6]. However, this metric may understate the productivity derived by Indians, as they accepted more suggestions, and suggestions tend to offer diminishing returns (each additional suggestion provides less incremental value). To address this, we limited this analysis to essays where participants accepted up

[6]This metric is only applicable to the AI condition, as suggestions were not provided in the control condition.

to seven suggestions[7]. Figure 5(c) confirms our hypothesis that **Indians derived significantly less productivity per suggestion compared to Americans** ($U = 5424.0$, $p < 0.001$, $D = -0.27$).

Overall, these results indicate that **AI helped both Indian and American participants write faster, but Americans derived more productivity from each suggestion, leaving Indian participants to rely more heavily on AI suggestions to achieve similar productivity outcomes**. This may be due to Indians overrelying on AI beyond the point that it was helpful, or, as concluded from their suggestion modification behavior because the AI suggestions were inherently less effective for them (i.e., higher cognitive load required to modify AI suggestions compared to American participants).

4.2 Does AI homogenize writing within cultures?

In the previous section, we outlined how AI impacts Indian and American users differently. In this section, we show that AI homogenizes writing styles within the same culture. This will set the stage for our main investigation of cross-cultural homogenization in the next section.

4.2.1 Similarity Analysis. We conducted a similarity analysis to investigate whether AI standardizes writing styles for participants within their respective cultures. The similarity metric we used was cosine similarity, computed on the essay embeddings obtained from OpenAI's text-embedding model.

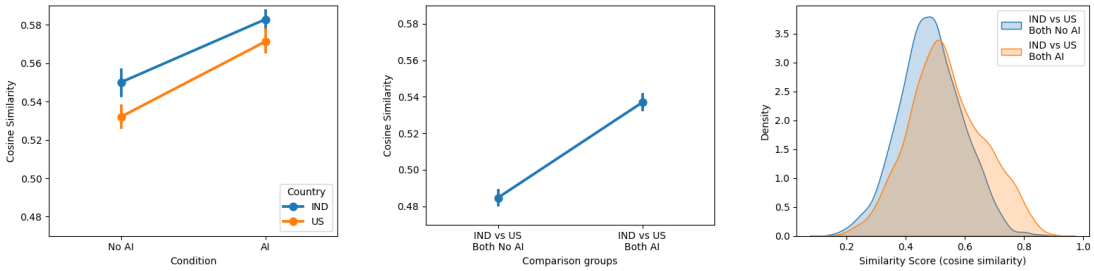
Setup: For this experiment, we computed similarity scores within each of our four experimental groups and compared the four distributions. For example, there were $n = 24$ Indian participants in the No AI condition, and each participant wrote four essays. So, we computed pair-wise similarity scores between all essays written by these 24 participants, grouped by the essay topic. This yielded a distribution of 4 essay topics $\times$ $\binom{24}{2}$ comparisons per topic = 1104 similarity scores. Similarly, we computed similarity distributions for the other three experimental groups (Indians with AI, Americans without AI, and Americans with AI). The means and 95% confidence intervals of these distributions are shown in Figure 6(a). The blue line shows how the within-culture similarity among Indian participants changes from the No AI condition to the AI condition, while the orange line represents this change for American participants.

Results: First, we notice that without AI (left side of the graph), Indian participants exhibited slightly higher within-culture similarity than American participants, suggesting that Indians naturally had more homogeneous writing styles than Americans. This natural difference between the Indian and American cohorts was significant, $t(2726) = 4.04$, $p < 0.001$, though the effect size was negligible (Cohen's $d = 0.16$).

Next, when AI suggestions were introduced, the similarity increased within both Indian and American cohorts. A t-test revealed that this increase was significant within both cohorts; $t(3622) = -7.22$, $p < 0.001$ for Indians (blue line) with a small effect size (Cohen's $d = -0.26$), and $t(3246) = -9.45$, $p < 0.001$ for Americans (orange line) with a small effect size (Cohen's $d = -0.33$). This means that AI caused both Indians and Americans to write more similarly within their cohorts. Visually, the magnitude of the increase looks slightly higher for Americans (the orange line is steeper than the blue line). However, a Difference-in-Difference analysis using a regression model showed that the increase was statistically similar in both cohorts ($p = 0.30$, 95% CI=$[-0.018, 0.006]$). This means that the magnitude of within-culture homogenization was similar in both cohorts.

In sum, the increasing similarity in both cohorts suggests that AI made writing more homogeneous, potentially by guiding users towards common phrasing, structure, or content. Thus, AI suggestions had a net homogenizing effect, suggesting that some level of cross-cultural homogenization was to be expected. However, as we demonstrate next, the extent of this homogenization was even more pronounced across cultures.

[7]Very few Americans accept more than seven suggestions in a task, whereas Indians continue to accept more, making comparisons unreliable.



(a) Within-culture similarity (b) Cross-culture similarity (c) Cross-culture similarity

Fig. 6. Within- and cross-culture similarity between Indian and American essays when written without and with AI. (a) The within-culture similarity is higher when AI is used, suggesting that AI homogenizes writing within both cohorts. (b) The cross-culture similarity is higher when AI is used, showing that AI homogenizes writing styles across cultures. (b) Cross-culture similarity scores are higher when AI suggestions are used.

4.3 Does AI homogenize writing across cultures?

We also investigated if AI might homogenize writing styles across cultures, i.e., do Indians start writing like Americans (or vice versa) when they use AI suggestions? We now show evidence for this effect.

4.3.1 Similarity Analysis. In this experiment, we extend the previous similarity analysis that demonstrated within-group homogenization by performing a cross-cultural comparison of essay embeddings.

Setup: First, we calculated similarity scores between essays written by Indians and Americans without AI, grouped by topic. There were $n = 24$ Indians and $m = 29$ Americans in the No AI condition, and each of them wrote essays on four topics. Hence, we computed the pairwise cosine similarity between all essays for each topic, yielding a distribution of 4 essay topics $\times$ (24×29) comparisons per topic = 2784 cross-culture similarity scores. This distribution represented the similarity between Indian and American essays when AI suggestions were not used. Next, we computed a distribution of similarity scores when AI suggestions *were* used. Figure 6(c) shows these two distributions.

Results: Figure 6(b) compares the means and 95% confidence intervals of the two distributions. It shows that without AI, the mean similarity between Indians and Americans was 0.48 ($SD = 0.11$), but with AI, the similarity score rose to 0.54 ($SD = 0.13$). A t-test revealed that this increase was statistically significant, $t(6958) = -17.89$, $p < 0.001$. This indicates that on average, Indians and Americans wrote more similarly when AI suggestions were used. Further, the effect size of this increase (Cohen's $d = -0.44$) is larger than what was observed in the within-culture homogenization for both cohorts, indicating a stronger cross-cultural homogenization effect.

Figure 6(c) confirms this observation. The similarity distribution in the AI condition (orange curve) is shifted to the right compared to the No AI condition (blue curve), showing a skew toward higher similarity scores. Moreover, the AI condition shows a longer right tail, extending further into higher similarity scores than the No AI distribution. This confirms that Indian and American participants wrote more similar essays when AI suggestions were used. Overall, these results suggest that **AI influenced Indian and American participants to write more similarly, thereby diminishing cultural differences in writing**.

Direction of Homogenization. The above analysis reveals cross-cultural homogenization. However, since cosine similarity is a symmetric measure, it does not indicate the direction of influence—whether Indians adopt American writing styles or vice versa. To investigate this, we conducted a complementary test. We computed similarity scores when only one of the two cohorts was using AI: (a) between Americans not using AI and Indians

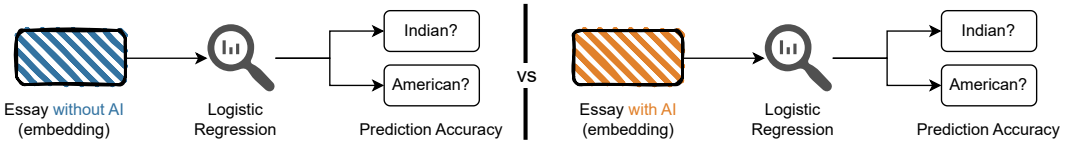


Fig. 7. Predictive modeling to investigate homogenization across cultures. Full details in Section 4.3.2.

using AI, and (b) vice versa, Indians not using AI and Americans using AI. This would allow us to see whether AI is making Indian writing styles converge more strongly toward natural American styles, or vice versa.

In both scenarios, we observed significantly higher similarity scores when one cohort used AI, compared to when neither cohort used AI ($p < 0.001$, based on independent t-tests). This is consistent with our earlier finding that AI has a net homogenizing effect. However, effect sizes revealed an important distinction: AI caused Indians to write more similarly to Americans' natural styles (Cohen's $d = -0.22$; small) than the other way around (Cohen's $d = -0.12$; negligible). **This suggests that AI may be driving Indian writing toward Western styles more than it is influencing American writing toward Indian styles.** In contrast, the reverse influence (Americans adopting Indian writing styles) was less pronounced. This finding is supported by prior research showing that AI models exhibit Western biases [14, 44], which may be causing a stronger convergence of Indian writing toward Western styles.

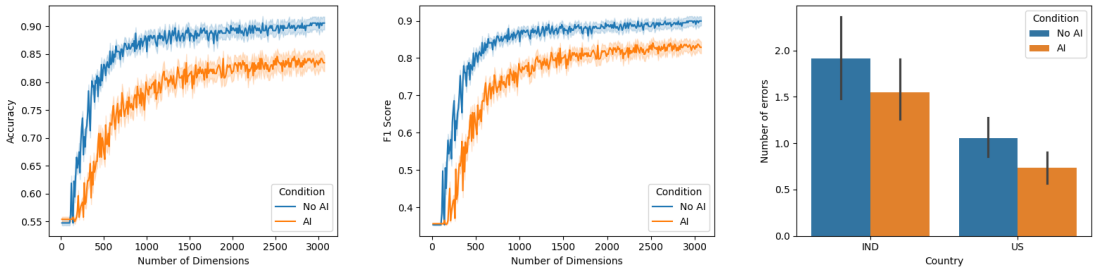
4.3.2 Predictive Modeling. We now demonstrate the homogenization effect using another experiment. We investigated whether AI suggestions make it harder for a classification model to differentiate Indian vs American authorship. If true, it would indicate that AI was diminishing cultural differences in writing styles.

Setup: We trained a classification model to determine whether an essay was written by an Indian or American participant. For simplicity and efficiency, we chose a logistic regression model, which is particularly well-suited for binary classification tasks. As shown in Figure 7, we used essay embeddings obtained from OpenAI's text-embedding model as features, with the participant's cohort (Indian or American) as the label. To isolate the impact of AI, we compared the performance of two models: one trained on essays written without AI suggestions (No AI condition) and the other on essays written with AI suggestions (AI condition). To ensure the robustness of our experiment, we used stratified k-fold cross-validation with $k = 20$, building 20 different models for each condition by swapping out one of the 20 folds as the test set. This resulted in a distribution of 20 accuracy and F1 scores in each condition, which we compared using an independent samples t-test.

Results: Our results show that the accuracy and F1 scores were lower in the AI condition. On average, the accuracy decreased by 7.1 percentage points, from 90.6% in the No AI condition to 83.5% in the AI condition. This decrease was statistically significant, $t(38) = 2.25$, $p = 0.03$, with a medium effect size (Cohen's $d = 0.71$). Similarly, the macro F1 score decreased by 7 percentage points on average, from 89.9% in the No AI condition to 82.9% in the AI condition, $t(38) = 2.07$, $p = 0.045$, also with a medium effect size (Cohen's $d = 0.65$).

The significant reduction in both accuracy and F1 scores indicates that a classification model found it more challenging to differentiate between Indian and American authorship when AI suggestions were used. This provides further evidence that AI diminished the natural writing differences between Indian and American participants, potentially steering users toward more Western-centric writing styles.

4.3.3 Did AI cause deep writing changes or merely omit explicit cultural references? Three of the four writing tasks in our study prompted participants to describe explicit cultural artifacts, such as food, public figures, and festivals. These explicit references can make it easier for a classification model to identify the writer's culture. So, if AI suggestions steered Indian participants away from these cultural markers, the absence of such references



(a) Accuracy (b) F1 score (c) Grammatical errors

Fig. 8. (a, b) A classification model trained to predict Indian vs American authorship consistently performs worse on essays written with AI suggestions. (c) AI reduces grammatical errors in both cohorts, showing that the homogenization effect is not due to mere grammatical fixes.

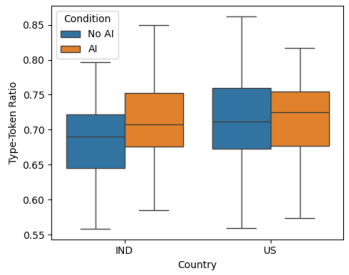
could explain the model's reduced performance. However, we wanted to explore whether AI also caused more subtle changes in writing style that went beyond the mere omission of explicit cultural artifacts (e.g., uniquely Indian food items, public figures, and festivals). We conducted two additional analyses to investigate this.

Ablation on Embedding Dimensions: We examined whether the performance drop persisted when the dimensionality of the essay embeddings was reduced. Specifically, we randomly selected d dimensions ($d < 3072$) from the original 3072-length embeddings and conducted the classification task using only these sub-sampled embeddings as features. To test this, we repeated the k-fold cross-validation experiment ($k = 20$) described earlier but varied the length of the essay embeddings from $d = 12$ to $d = 3072$ in increments of 10. In each iteration, we randomly sampled d dimensions from the original embedding and used only those dimensions as features for the model. To ensure robustness to random sampling, we repeated the experiment 10 times for each value of d.

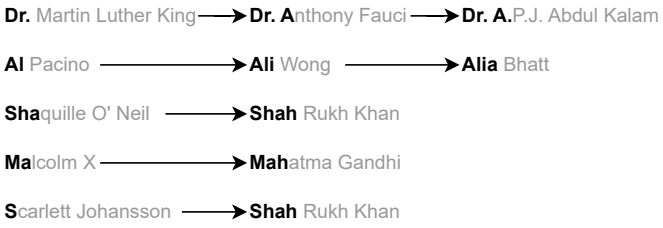
The results, shown in Figure 8(a, b), reveal a consistent performance drop in both accuracy and F1 score across all embedding sizes in the AI condition. Importantly, this drop persists even for embeddings with very low dimensionality, which are unlikely to capture explicit cultural artifacts (e.g., uniquely Indian food items or public figures). By randomly sampling dimensions, we disrupted the structure of the embedding space, further reducing the likelihood that specific cultural features would be preserved. Despite this, the performance gap suggests that the reduced accuracy is not merely due to the absence of these cultural references, but reflects deeper changes in writing style introduced by AI.

Performance on Email-Writing Task: Next, we trained the model using only the essays from the email writing task and examined the resulting accuracy and F1 scores. This task (requesting a leave of absence) was designed to capture implicit cultural nuances rather than explicit references. We found that the model's performance in distinguishing between Indian and American authorship in the AI condition was even worse for these essays. Accuracy dropped from 82.9% in the No AI condition to 60% in the AI condition, a statistically significant decrease, $t(8) = 5.63$, $p < 0.001$, with a large effect size (Cohen's $d = 3.56$). The F1 score also declined sharply, from 81.7% to 49.2% AI, $t(8) = 5.79$, $p < 0.001$, with a large effect size ($d = 3.66$). This supports the findings that AI suggestions may be influencing subtle aspects of writing style, contributing to a homogenization effect that goes beyond the mere omission of explicit cultural identifiers.

These two experiments suggest that the decreased model performance was not merely due to the absence of explicit cultural markers, instead it reflects deeper, more subtle changes in writing style induced by AI. This is a crucial finding, as it indicates that **AI suggestions influence not just the surface content of what is written, but also more ingrained elements of *how* it is written**.



(a) Lexical diversity (TTR)



(b) Culturally incongruent suggestions

Fig. 9. (a) Lexical diversity of the essays (measured by type-token ratio) partitioned by country and AI conditions. On average, AI increases the lexical diversity for Indians, taking it closer to American styles, but does not change the lexical diversity for Americans. Note: We show a box plot because, for this analysis, the bar plot under-emphasized statistically significant differences. (b) Progression of AI suggestions for public figures. Participants typed black text to guide suggestions, but the system initially proposed culturally incongruent options (gray). The intended match emerged only after adding more context.

4.3.4 Did AI homogenize Indian writing solely by correcting grammar? Another plausible explanation for the homogenization effect could be that, as non-native English speakers, Indian writers made more grammatical errors, and AI suggestions corrected these errors, leading to the observed homogenization. To test this hypothesis, we ran a grammar check on all essays using an open-source tool called LanguageTool[8]. This tool applies over 6000 grammar rules to the text. However, its spell-checker misclassifies Indian proper nouns as grammatical errors, overstating errors for Indian essays. To ensure a fair comparison, we disabled the spell-checking rules while keeping all other grammar rules intact.

The results of this analysis are shown in Figure 8(c). Descriptively, it can be seen that AI reduced grammatical errors in both Indian and American writing, but the reduction was similar in both cohorts. Statistically, the reduction was significant in American essays ($U = 7729.0$, $p = 0.04$) with a small effect size (Cliff's Delta $D = 0.15$), but *not* significant in Indian essays ($U = 7584.5$, $p = 0.18$).

If grammar correction was the primary driver of homogenization, we would expect a larger reduction in grammatical errors for Indian writers, given their higher propensity of making such errors. However, we see that AI's impact on grammar correction was similar for both Indians and Americans, **the homogenization effect cannot be attributed solely to grammatical corrections**. Instead, AI appears to be influencing writing in more subtle ways, which we illustrate next with a concrete example.

4.3.5 A Concrete Example: Altering Lexical Diversity. Our two main experiments (cosine similarity and predictive modeling) used word embeddings to capture the semantic meaning of participants' essays, providing a robust yet abstract measure of homogenization. We also showed that the homogenization effect is not simply the result of omitting cultural references or correcting grammatical errors, meaning that AI induced more subtle, implicit changes in writing. What are some of these changes?

To provide a concrete example, we examined the lexical diversity of essays written with and without AI. We used the Type-Token Ratio (TTR), a commonly used metric to measure lexical diversity. A higher TTR reflects a greater usage of unique words, indicating higher lexical diversity.

Figure 9(a) shows the lexical diversity in participants' essays, partitioned by country and AI condition. The natural lexical diversity in Indian writing (without AI) was significantly different from that of American participants (the two blue boxes in the plot), $t(210) = -2.94$, $p = 0.003$, Cohen's $d = -0.41$. However, when AI suggestions

[8]LanguageTool, used through a Python library: `language_tool_python`.

Topic	Essay without AI	Essay with AI
Festival (Diwali)	My favorite festival is Diwali. Diwali is usually celebrated in between October and November. On this occasion, I used to worship goddess Laxmi along with my family. Most fascinating thing is to pop crackers and eat sweets for minimum span of 4 days. I always wait for this festival. We lighten our houses with earthen lamps and worship cows and lords .	My favorite festival is Di**wali because of the lights and sweets** . Its the bigg**est festival in India**. In morning we eat **traditional Indian breakfast items** . We also clean or pain**t our homes to welcome prosperity**. At night we **light diyas and burst fireworks**. Its **a time for family gatherings and joyous celebrations** . Everyone **wears new clothes and exchanges gifts** . Its **a time filled with happiness and warmth** .
Food (Biryani)	My favorite food is chicken biriyani. I like it because it tastes good and it is easy to prepare. I like to prepare it in Malabar style . That recipe uses some exotic ingredients like nutmeg . Chicken biriyani is along with raita, lemon pickle and dates chutney tastes divine. Chicken biriyani is supposed to be a dish that was brought to India by Mughals.	my favorite food is briyani because **of its rich flavors and spices** . i love **the way it melts in my mouth** . i use **aromatic basmati rice and tender meat**. i love chicken **biryani the most**. which **is your favorite type of biryani?** hyderbadi **biryani is also delicious**. which has **rich spices and a unique taste**.

Table 5. Comparison of cultural artifacts described by Indian participants with and without AI suggestions. AI-generated text is shown in **bold**. Highlights indicate key differences: green for cultural details present without AI but missing with AI, yellow for generic phrasing introduced by AI, blue for Westernized descriptions, orange for exoticization of Indian food.

were introduced, the lexical diversity of Indians went up and converged with that of Americans (the two orange boxes) and the difference between the two cohorts was no longer significant, $t(258) = -0.32$, $p = 0.75$, Cohen's $d = -0.04$. Thus, AI eliminated the significant difference that existed between their natural styles.

This suggests that **AI homogenized the natural lexical diversity exhibited by Indian participants, aligning it more closely with the diversity typical of American writing styles**. While our initial analyses reveal homogenization using high-dimensional text embeddings, the converging TTR provides a concrete example of homogenization toward Western writing styles.

4.4 Impact of AI on Cultural Writing

We conducted a content analysis to examine the writing changes introduced by AI. As part of this process, we manually coded essays on the first three topics (favorite food, festival, and public figure) to identify references to cultural symbols and idols. We focused on essays written by Indian participants, as aforementioned quantitative results indicated that homogenization had a greater impact on them compared to the American participants.

4.4.1 Changing Cultural Expression. To identify how AI induced changes in the cultural expression of Indian participants, we compared essays about the same festivals and food items written with and without AI, focusing on dimensions such as exoticization, misrepresentation, and Western gaze [68]. Table 5 provides illustrative examples highlighting how AI introduced changes in cultural expression and more examples are provided in Appendix C.

As shown in Table 5 essays describing the Indian festival of Diwali without AI often included rich cultural details such as religious context and worship rituals. In contrast, essays written with AI tended to feature more generic descriptions, emphasizing elements like sweets, gift exchanges, and family gatherings. While these descriptions are not incorrect and do not constitute overt misrepresentation, they lack cultural nuance. For example, family gatherings during Diwali often involve specific activities such as designing rangolis and bursting firecrackers. These details appeared in essays without AI but were absent in those written with AI, which reduced Indian festivals to generic celebrations. Similarly, when writing about Biryani, participants in the No AI condition described specific regional variations and accompaniments. However, essays written with AI did not contain such details and frequently used cultural tropes like "rich", "flavorful", and "aromatic", subtly exoticizing Indian food. This analysis was inconclusive for the public figures task due to the vast pool of celebrities, resulting in a minimal overlap between those mentioned in the AI and No AI conditions.

4.4.2 Cultural Incongruence in AI Suggestions. Given the open-ended nature of the tasks, we could not computationally record the exact moment an artifact was suggested to a participant. Hence, we manually analyzed the logs, examining the suggestions offered by the AI system and the text preceding them. Starting with the public figure task, we recorded the sequence of celebrities suggested by the system as participants added more characters until they finalized the name in their essay. Each suggested celebrity was then categorized as Indian or non-Indian based on their citizenship.

Among Indian participants, we found no instance where the first suggested celebrity matched the one ultimately included in the essay. Moreover, the initial suggestions were almost exclusively Western figures (with one exception). This resulted in frequent mismatches, where culturally incongruent suggestions persisted even after participants began typing their intended public figure's name. Examples of such mismatches are shown in Figure 9(b).

The food and festival tasks showed fewer stark mismatches due to a smaller set of possible suggestions for similar starting characters. However, the first suggested food item was always pizza or sushi, and the first suggested festival was invariably Christmas. We build on these observations next.

4.4.3 Did AI suggestions influence participants' choice of artifacts? Our study design did not allow us to capture what participants in the AI condition *would* have written *without* AI suggestions. However, we compared the artifacts selected by participants across the AI and No AI conditions.

In the food task, the first suggested item was always pizza or sushi. While pizza is popular in India, no Indian participant selected sushi as their favorite dish in the No AI condition, yet sushi appeared in three essays in the AI condition. Similarly, the first AI suggestion in the festival task was always Christmas. Without AI suggestions, 12.5% of Indian participants (3/24) wrote about Christmas, but this proportion slightly increased to 13.9% (5/36) in the AI condition. In the public figures task, 54% of Indian participants (13/24) selected Indian public figures in the No AI condition, compared to 44% (16/36) in the AI condition (full list in Appendix B).

These observations suggest a possibility of changing cultural preferences due to the Western bias of the model. However, the small sample size limits the strength of our conclusions, especially since participants often bypassed the first suggestions in favor of their actual preferences. This trend merits further investigation in future research.

5 DISCUSSION

Our results show that AI-powered writing suggestions yield greater benefits for American users compared to Indian users. Moreover, AI suggestions homogenize writing styles between Indians and Americans, pushing Indian users to adopt American writing styles and diminishing cultural nuances in writing. Worryingly, the changes induced by AI suggestions are both explicit (e.g., changing cultural preferences), as well as subtle and deeply ingrained (e.g., lexical diversity, exoticization, Westernization), making it harder to identify the cultural

erasure they might be causing. In the following sections, we discuss cultural differences in AI attitudes, examine the implications of the observed homogenization of writing styles, and propose strategies to mitigate the harms of cultural homogenization and AI-driven neocolonialism.

5.1 Cultural Differences in AI Attitudes

As we showed in Section 4.1, Indian participants exhibited higher engagement with AI suggestions: a larger proportion of their essays were composed of AI-generated text, and they accepted more AI suggestions compared to Americans. From a purely causal lens, this difference in AI engagement could be seen as a confounding factor, potentially undermining our findings. For example, it could suggest that the homogenization effects are simply a result of Indians' higher AI usage. However, we argue that this is not a flaw, but an important finding of our study. To fully understand the differential AI engagement seen in our study, we turn to HCI scholarship for a socio-technical view of human-AI interaction.

The way people perceive the benefits and harms of emerging technologies varies across cultures and geographic regions [53]. For instance, people in individualistic cultures (e.g., the US) tend to evaluate new technologies through direct and formal sources, while those in collectivistic cultures (e.g., India) rely more on peer feedback [51]. AI technologies are no different—they are cultural artifacts [22] and perceptions about them vary across cultures, both at the national [23] and user levels [53]. Research shows that users in non-Western settings express more positive emotions toward conversational agents, whereas users in the US tend to be more ambivalent [53]. Non-Western users also show higher levels of trust in AI technologies than their American counterparts [78]. In writing contexts, particularly when composing in English, AI-based writing suggestions provide more utility to non-native speakers than to native speakers, leading them to engage more with the suggestions [13]. This positive attitude often manifests as overreliance on AI [12], especially among novice users in non-Western settings [62].

These cultural differences in AI overreliance may be exacerbated as embedded AI applications reach more diverse users. While past research has shown that overreliance can be mitigated with AI explanations [80, 84], this approach becomes less viable in new-age interactive and embedded AI tools, where providing explanations may not be feasible. For instance, in writing applications, presenting explanations for every suggestion may disrupt the user's flow, hindering productivity. Furthermore, our study shows that embedded AI applications promising productivity gains may inadvertently push non-Western users toward overreliance by delivering diminished benefits to them.

We argue that higher trust and AI reliance are cultural characteristics that significantly influence human-AI interaction. It is important to acknowledge these cross-cultural differences in trust, adoption, and degree of AI reliance. By trying to eliminate them in our analyses as mere confounders, we may fail to notice the cultural harms they are causing, risking a future where cultural nuances are erased.

5.2 Harms of Cultural Homogenization

Neocolonialism refers to the practice of using economic, political, cultural, or technological forces by Western nations to indirectly control formerly colonized nations. Unlike traditional colonialism, which involved direct territorial control, neocolonialism operates through more subtle mechanisms, such as control over resources, markets, institutions, or the spread of dominant cultural and ideological values [33, 75]. In recent years, a new form of neo-colonialism has emerged, mediated by advances in big data and AI, where the digital realm is used to move wealth from the poor to the rich, a dynamic reminiscent of colonialism. [19]. A growing body of scholarly work points to the harms of this *data colonization* [20] and *AI colonialism* [34, 81] and argues how everyday products and services offered by big tech companies based in the West capitalize on and profit from the data extracted from people in non-Western settings to enrich the wealthy and powerful. An example of this form of colonialism is how billion-dollar tech companies such as Amazon, Facebook, Google, Microsoft, and OpenAI

paying pennies to workers in developing nations to label training data [72], only to build products that center Western norms [76]. In this neocolonialism, these companies establish Western influence not by political or military control, but by consolidating Western power over data, cultural expression, and information flows, inviting calls from critical scholars to decolonize computing [3, 42, 60].

Our study provides concrete evidence of AI colonialism that is more worrying than an AI model's internal biases: such models alter cultural preferences and descriptions of cultural artifacts *even under human oversight*. This homogenization reflects a form of cultural imperialism [48], where one culture dominates and suppresses the plurality of knowledge, practices, and languages, reinforcing Western hegemony over values. For example, the growing use of AI in story-writing interfaces [50, 79, 87, 88] can gradually erode traditional storytelling methods and creative expressions. Similarly, Western-biased AI suggestions in email clients may encourage a direct and informal tone, which could lead to conflicts with superiors in cultures that observe an established social hierarchy [38] and expect formal language as a sign of respect. More broadly, cultures differ in directness, formality, and politeness [36], distinct stylistic variations that are at risk of being lost to this homogenization. Our work highlights one such change (lexical diversity) and lays the groundwork for further research to identify more of these subtle shifts. This will require interdisciplinary effort among linguists, critical scholars, and AI researchers.

In addition to explicit cultural imperialism, the homogenization effect can also lead to an implicit *cognitive* imperialism [10]. Writing is closely tied to thinking, and what one writes can transform what one thinks [57]. Continuously subjecting users to mono-cultural suggestions portrays Western norms and values as the only truth, reinforcing the supremacy of "only one language, one culture, one frame of reference" [10]. This cultural hegemony may encourage users to think, write, and reason in ways that reflect Western values. Over time, this can affect how individuals perceive their own culture, potentially leading to feelings of cultural inferiority or a loss of cultural identity.

5.3 Call for Mitigation Strategies

The cultural homogenization identified in our study nudges users toward aligning their writing with Western styles. As more users accept these biased suggestions, it creates a biased feedback loop where more online content conforms to Western norms, which then becomes training data for future models, further perpetuating these biases. It is therefore critical to mitigate cultural harms at this foundational stage, as this vicious cycle could remove the opportunity to correct these issues in the future. Furthermore, current mainstream LLMs are owned by a few large companies based in the West. However, a growing trend of distilling smaller models from these larger ones to enable privacy-preserving, on-device computation [24, 25, 63] presents increased challenges in correcting these biases. For example, if cultural biases are not addressed now, they may propagate into local models and systems that are resistant to change. As a result, even if future AI systems correct these biases, legacy systems may continue to perpetuate them.

Our work makes a critical argument to *design for cultural plurality*. While addressing cultural plurality is a complex challenge, as acknowledged by early studies [36, 82], we propose a divide-and-conquer approach. While AI researchers develop mitigation strategies at the model and architecture levels, we encourage HCI researchers and practitioners to develop solutions at the application layer. Although standard UX principles suggest that AI features be seamlessly integrated into applications so that they work out of the box, to avoid cultural harm, we need to design with *friction* [56]. This approach provides users greater control over AI features and encourages them to make appropriate configurations. For example, in writing applications, this could involve prompting the user to upload writing samples so that the AI can tailor suggestions to their personal style rather than generalizing to Western styles. More broadly, developers could provide the model with the socio-cultural context in which the application will be used or ask users to provide a personal profile to help contextualize the model's output. We

invite the HCI and AI community to incorporate guardrails and intentional friction into their AI applications to design for cultural plurality and avoid the harms of cultural imperialism.

5.4 Limitations and Future Work

The terms "Western" and "non-Western" are broad cultural labels, as US culture does not fully represent all Western cultures, nor does Indian culture encompass all non-Western cultures. Even within "India" and "the US," there exist rich and diverse subcultures. In using these categorizations, we draw upon prior work [9, 68] that has used geographical boundaries as a proxy for cultural differences. However, future work is required to determine if our results generalize to other countries and sub-cultures. In the wake of our findings, this is especially important to understand the scale at which cultural erasure might be happening, especially for cultures that are already marginalized. Ironically, the community's focus on broad cultural categorization (e.g., treating "India" or "Africa" as monolithic cultures) may itself contribute to cultural erasure. Additionally, the scale of our study was limited by the relatively small pool of Indian participants on Prolific. Future research may expand our results by scaling horizontally across diverse cultures and vertically with larger participant pools. Finally, since this was a formative study, we employed a between-subjects design to explore the various layers of Hofstede's Cultural Onion. This approach required analysis at the group level. Future studies could build on our findings by using a within-subjects design to validate whether homogenization effects persist at the individual level.

6 CONCLUSION

We conducted a cross-cultural controlled experiment with 118 Indian and American participants to examine the consequences of interacting with a culturally different model. Participants completed writing tasks designed to elicit cultural practices and values. They were randomly assigned to one of two conditions: the AI condition, where they received inline AI suggestions to assist with their writing, or the No AI condition, in which they wrote without any AI support. Our analysis of writing logs and essays revealed two key findings. First, AI suggestions offered greater productivity gains to American users, leaving non-Western users to put in more effort to achieve similar benefits. Second, AI homogenized writing styles towards Western norms, subtly influencing Indians to adopt Western writing patterns. This homogenization provides concrete evidence of AI colonialism, where these models reinforce Western cultural hegemony. It is crucial to address these biases before culturally biased content proliferates and becomes part of the training data for future AI models.

ACKNOWLEDGMENTS

This work is supported by funding from Infosys, Microsoft AFMR Program, and Global Cornell. This material is based upon work partially supported by the National Science Foundation under Grant No. CHS 1901151/1901329.

REFERENCES

[1] Abubakar Abid, Maheen Farooqi, and James Zou. 2021. Persistent Anti-Muslim Bias in Large Language Models. arXiv:2101.05783 [cs.CL] https://arxiv.org/abs/2101.05783

[2] Muhammad Farid Adilazuarda, Sagnik Mukherjee, Pradhyumna Lavania, Siddhant Singh, Alham Fikri Aji, Jacki O'Neill, Ashutosh Modi, and Monojit Choudhury. 2024. Towards Measuring and Modeling "Culture" in LLMs: A Survey. arXiv:2403.15412 [cs.CY] https://arxiv.org/abs/2403.15412

[3] Syed Mustafa Ali. 2016. A brief introduction to decolonial computing. *XRDS: Crossroads, The ACM Magazine for Students* 22, 4 (June 2016), 16–21. https://doi.org/10.1145/2930886

[4] Oghenemaro Anuyah, Ruyuan Wan, Cornelius Adejoro, Tom Yeh, Ronald Metoyer, and Karla Badillo-Urquiola. 2023. Cultural Considerations in AI Systems for the Global South: A Systematic Review. In *Proceedings of the 4th African Human Computer Interaction Conference (AfriCHI 2023)*. ACM. https://doi.org/10.1145/3628096.3629046

[5] Kenneth C. Arnold, Krzysztof Z. Gajos, and Adam T. Kalai. 2016. On Suggesting Phrases vs. Predicting Words for Mobile Text Composition. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM. https://doi.org/10.1145/2984511.

2984584

[6] Arnav Arora, Lucie-Aimée Kaffee, and Isabelle Augenstein. 2023. Probing Pre-Trained Language Models for Cross-Cultural Differences in Values. arXiv:2203.13722 [cs.CL] https://arxiv.org/abs/2203.13722

[7] Paul Baker, Andrew Hardie, and Tony McEnery. 2006. *A glossary of corpus linguistics*. Edinburgh University Press, Edinburgh, Scotland.

[8] Solon Barocas, Kate Crawford, Aaron Shapiro, and Hanna Wallach. 2017. The problem with bias: From allocative to representational harms in machine learning. In *SIGCIS conference paper*.

[9] Abhipsa Basu, R. Venkatesh Babu, and Danish Pruthi. 2023. Inspecting the Geographical Representativeness of Images from Text-to-Image Models. arXiv:2305.11080 [cs.CV] https://arxiv.org/abs/2305.11080

[10] Marie Battiste and James Youngblood Henderson. 2000. *Protecting indigenous knowledge and heritage*. Purich Publishing, Saskatoon, SK, Canada.

[11] Advait Bhat, Saaket Agashe, Parth Oberoi, Niharika Mohile, Ravi Jangir, and Anirudha Joshi. 2023. Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 436–452. https://doi.org/10.1145/3581641.3584060

[12] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–21. https://doi.org/10.1145/3449287

[13] Daniel Buschek, Martin Zürn, and Malin Eiband. 2021. The Impact of Multiple Parallel Phrase Suggestions on Email Input and Composition Behaviour of Native and Non-Native English Writers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM. https://doi.org/10.1145/3411764.3445372

[14] Yong Cao, Li Zhou, Seolhwa Lee, Laura Cabello, Min Chen, and Daniel Hershcovich. 2023. Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study. arXiv:2303.17466 [cs.CL] https://arxiv.org/abs/2303.17466

[15] Tuhin Chakrabarty, Xurui Zhang, Smaranda Muresan, and Nanyun Peng. 2021. MERMAID: Metaphor Generation with Symbolism and Discriminative Decoding. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.336

[16] Mia Xu Chen, Benjamin N. Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M. Dai, Zhifeng Chen, Timothy Sohn, and Yonghui Wu. 2019. Gmail Smart Compose: Real-Time Assisted Writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. ACM. https://doi.org/10.1145/3292500.3330723

[17] Yu Ying Chiu, Liwei Jiang, Bill Yuchen Lin, Chan Young Park, Shuyue Stella Li, Sahithya Ravi, Mehar Bhatia, Maria Antoniak, Yulia Tsvetkov, Vered Shwartz, and Yejin Choi. 2024. CulturalBench: a Robust, Diverse and Challenging Benchmark on Measuring the (Lack of) Cultural Knowledge of LLMs. arXiv:2410.02677 [cs.CL] https://arxiv.org/abs/2410.02677

[18] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *23rd International Conference on Intelligent User Interfaces (IUI'18)*. ACM. https://doi.org/10.1145/3172944.3172983

[19] Nick Couldry and Ulises A. Mejias. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press, Stanford. 352 pages. http://www.sup.org/books/title/?id=28816 Accessed: 2024-09-07.

[20] Nick Couldry and Ulises A. Mejias. 2019. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media* 20, 4 (2019), 336–349. https://doi.org/10.1177/1527476418796632 arXiv:https://doi.org/10.1177/1527476418796632

[21] Paramveer S. Dhillon, Somayeh Molaei, Jiaqi Li, Maximilian Golub, Shaochun Zheng, and Lionel Peter Robert. 2024. Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-writing with Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24, Vol. 22)*. ACM, 1–18. https://doi.org/10.1145/3613904.3642134

[22] Paul Dourish. 2016. Algorithms and their others: Algorithmic culture in context. *Big Data & Society* 3, 2 (Aug. 2016), 205395171666512. https://doi.org/10.1177/2053951716665128

[23] Verena Eitle and Peter Buxmann. 2020. Cultural Differences in Machine Learning Adoption: An International Comparison between Germany and the United States. In *Proceedings of the 28th European Conference on Information Systems (ECIS)*. AIS, An Online AIS Conference.

[24] Gemma Team et al. 2024. Gemma 2: Improving Open Language Models at a Practical Size. arXiv:2408.00118 [cs.CL] https://arxiv.org/abs/2408.00118

[25] Marah Abdin et al. 2024. Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone. arXiv:2404.14219 [cs.CL] https://arxiv.org/abs/2404.14219

[26] Richard P. Gabriel, Jilin Chen, and Jeffrey Nichols. 2015. InkWell: A Creative Writer's Creative Assistant. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition (C&C '15)*. ACM. https://doi.org/10.1145/2757226.2757229

[27] 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* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 205–216. https:

//doi.org/10.1145/3593013.3593989
[28] Katy Ilonka Gero and Lydia B. Chilton. 2019. How a Stylistic, Machine-Generated Thesaurus Impacts a Writer's Process. In *Proceedings of the 2019 on Creativity and Cognition (C&C '19)*. ACM. https://doi.org/10.1145/3325480.3326573
[29] Katy Ilonka Gero and Lydia B. Chilton. 2019. Metaphoria: An Algorithmic Companion for Metaphor Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300526
[30] Katy Ilonka Gero, Vivian Liu, and Lydia B. Chilton. 2021. Sparks: Inspiration for Science Writing using Language Models. arXiv:2110.07640 [cs.HC] https://arxiv.org/abs/2110.07640
[31] Marjan Ghazvininejad, Xing Shi, Yejin Choi, and Kevin Knight. 2016. Generating Topical Poetry. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://doi.org/10.18653/v1/d16-1126
[32] Zhi Guo, Pei-Luen Patrick Rau, and Rüdiger Heimgärtner. 2022. *The "Onion Model of Human Factors": A Theoretical Framework for Cross-Cultural Design*. Springer International Publishing, 20–33. https://doi.org/10.1007/978-3-031-05434-1_2
[33] Sandra Halperin. 2024. Neocolonialism. https://www.britannica.com/topic/neocolonialism. https://www.britannica.com/topic/neocolonialism Accessed: 2024-09-07.
[34] Karen Hao. 2022. Artificial intelligence is creating a new colonial world order. https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/. https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/ Accessed: 2024-09-07.
[35] David J. Hauser and Norbert Schwarz. 2015. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods* 48, 1 (March 2015), 400–407. https://doi.org/10.3758/s13428-015-0578-z
[36] Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruixiang Cui, Constanza Fierro, Katerina Margatina, Phillip Rust, and Anders Søgaard. 2022. Challenges and Strategies in Cross-Cultural NLP. 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, 6997–7013. https://doi.org/10.18653/v1/2022.acl-long.482
[37] G. Hofstede. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. SAGE Publications. https://books.google.com/books?id=w6z18LJ_1VsC
[38] Geert Hofstede. 2011. Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture* 2, 1 (Dec. 2011). https://doi.org/10.9707/2307-0919.1014
[39] Geert H. Hofstede. 1991. *Cultures and organizations: Software of the mind*. McGraw-Hill, London and New York. xii, 279 pages.
[40] Jess Hohenstein, Rene F. Kizilcec, Dominic DiFranzo, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeffrey Hancock, and Malte F. Jung. 2023. Artificial intelligence in communication impacts language and social relationships. *Scientific Reports* 13, 1 (April 2023). https://doi.org/10.1038/s41598-023-30938-9
[41] Julie S. Hui, Darren Gergle, and Elizabeth M. Gerber. 2018. IntroAssist: A Tool to Support Writing Introductory Help Requests. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM. https://doi.org/10.1145/3173574.3173596
[42] Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E. Grinter. 2010. Postcolonial computing: a lens on design and development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM. https://doi.org/10.1145/1753326.1753522
[43] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM. https://doi.org/10.1145/3544548.3581196
[44] Rebecca L Johnson, Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. The Ghost in the Machine has an American accent: value conflict in GPT-3. arXiv:2203.07785 [cs.CL] https://arxiv.org/abs/2203.07785
[45] Kowe Kadoma, Marianne Aubin Le Quere, Xiyu Jenny Fu, Christin Munsch, Danaë Metaxa, and Mor Naaman. 2024. The Role of Inclusion, Control, and Ownership in Workplace AI-Mediated Communication. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1016, 10 pages. https://doi.org/10.1145/3613904.3642650
[46] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, and Vivek Ramavajjala. 2016. Smart Reply: Automated Response Suggestion for Email. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM. https://doi.org/10.1145/2939672.2939801
[47] Franki Y.H. Kung, Navio Kwok, and Douglas J. Brown. 2017. Are Attention Check Questions a Threat to Scale Validity? *Applied Psychology* 67, 2 (Aug. 2017), 264–283. https://doi.org/10.1111/apps.12108
[48] Moshe Landsman. 2014. Cultural Imperialism. In *Encyclopedia of Critical Psychology*, Thomas Teo (Ed.). Springer, New York, NY, 344–347. https://doi.org/10.1007/978-1-4614-5583-7_63

[49] Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault. 2010. *Automated Grammatical Error Detection for Language Learners*. Springer International Publishing. https://doi.org/10.1007/978-3-031-02137-4

[50] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM. https://doi.org/10.1145/3491102.3502030

[51] Sang-Gun Lee, Silvana Trimi, and Changsoo Kim. 2013. The impact of cultural differences on technology adoption. *Journal of World Business* 48, 1 (Jan. 2013), 20–29. https://doi.org/10.1016/j.jwb.2012.06.003

[52] Marjaana Lindeman and Markku Verkasalo. 2005. Measuring Values With the Short Schwartz's Value Survey. *Journal of Personality Assessment* 85, 2 (Oct. 2005), 170–178. https://doi.org/10.1207/s15327752jpa8502_09

[53] Zihan Liu, Han Li, Anfan Chen, Renwen Zhang, and Yi-Chieh Lee. 2024. Understanding Public Perceptions of AI Conversational Agents: A Cross-Cultural Analysis. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24, Vol. 1)*. ACM, 1–17. https://doi.org/10.1145/3613904.3642840

[54] Xi Lu, Eunkyung Jo, Seora Park, Hwajung Hong, Yunan Chen, and Daniel A. Epstein. 2022. Understanding Cultural Influence on Perspectives Around Contact Tracing Strategies. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–26. https://doi.org/10.1145/3555569

[55] Enrique Manjavacas, Folgert Karsdorp, Ben Burtenshaw, and Mike Kestemont. 2017. Synthetic Literature: Writing Science Fiction in a Co-Creative Process. In *Proceedings of the Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2017)*. Association for Computational Linguistics. https://doi.org/10.18653/v1/w17-3904

[56] Thomas Mejtoft, Sarah Hale, and Ulrik Söderström. 2019. Design Friction. In *Proceedings of the 31st European Conference on Cognitive Ergonomics (ECCE '19)*. Association for Computing Machinery, New York, NY, USA, 41–44. https://doi.org/10.1145/3335082.3335106

[57] Richard Menary. 2007. Writing as thinking. *Language Sciences* 29, 5 (Sept. 2007), 621–632. https://doi.org/10.1016/j.langsci.2007.01.005

[58] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23, Vol. 2021)*. ACM, 1–34. https://doi.org/10.1145/3544548.3581225

[59] Mazda Moayeri, Elham Tabassi, and Soheil Feizi. 2024. WorldBench: Quantifying Geographic Disparities in LLM Factual Recall. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1211–1228. https://doi.org/10.1145/3630106.3658967

[60] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology* 33, 4 (July 2020), 659–684. https://doi.org/10.1007/s13347-020-00405-8

[61] Tarek Naous, Michael J. Ryan, Alan Ritter, and Wei Xu. 2024. Having Beer after Prayer? Measuring Cultural Bias in Large Language Models. arXiv:2305.14456 [cs.CL] https://arxiv.org/abs/2305.14456

[62] Chinasa T. Okolo, Dhruv Agarwal, Nicola Dell, and Aditya Vashistha. 2024. "If it is easy to understand then it will have value": Examining Perceptions of Explainable AI with Community Health Workers in Rural India. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 1–28. https://doi.org/10.1145/3637348

[63] OpenAI. 2024. GPT-4o mini: advancing cost-efficient intelligence. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/ Accessed: 2024-09-07.

[64] James L. Peterson. 1980. Computer programs for detecting and correcting spelling errors. *Commun. ACM* 23, 12 (Dec. 1980), 676–687. https://doi.org/10.1145/359038.359041

[65] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23, Vol. 87)*. ACM, 1–16. https://doi.org/10.1145/3544548.3580907

[66] Ritika Poddar, Rashmi Sinha, Mor Naaman, and Maurice Jakesch. 2023. AI Writing Assistants Influence Topic Choice in Self-Presentation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM. https://doi.org/10.1145/3544549.3585893

[67] Vinodkumar Prabhakaran, Rida Qadri, and Ben Hutchinson. 2022. Cultural Incongruencies in Artificial Intelligence. arXiv:2211.13069 [cs.CY] https://arxiv.org/abs/2211.13069

[68] Rida Qadri, Renee Shelby, Cynthia L. Bennett, and Emily Denton. 2023. AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 506–517. https://doi.org/10.1145/3593013.3594016

[69] Philip Quinn and Shumin Zhai. 2016. A Cost-Benefit Study of Text Entry Suggestion Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16)*. ACM. https://doi.org/10.1145/2858036.2858305

[70] Abhinav Rao, Akhila Yerukola, Vishwa Shah, Katharina Reinecke, and Maarten Sap. 2024. NormAd: A Framework for Measuring the Cultural Adaptability of Large Language Models. arXiv:2404.12464 [cs.CL] https://arxiv.org/abs/2404.12464

[71] Melissa Roemmele and Andrew S. Gordon. 2018. Automated Assistance for Creative Writing with an RNN Language Model. In *Companion Proceedings of the 23rd International Conference on Intelligent User Interfaces (IUI'18)*. ACM. https://doi.org/10.1145/3180308.3180329

[72] Niamh Rowe. 2023. Millions of Workers Are Training AI Models for Pennies. *Wired* (Oct. 2023). https://www.wired.com/story/millions-of-workers-are-training-ai-models-for-pennies Accessed: 2024-09-07.
[73] Shalom H. Schwartz. 1992. *Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries*. Elsevier, 1–65. https://doi.org/10.1016/s0065-2601(08)60281-6
[74] Pola Schwöbel, Jacek Golebiowski, Michele Donini, Cedric Archambeau, and Danish Pruthi. 2023. Geographical Erasure in Language Generation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.823
[75] John Scott and Gordon Marshall. 2009. *A Dictionary of Sociology*. Oxford University Press. https://doi.org/10.1093/acref/9780199533008.001.0001
[76] Farhana Shahid and Aditya Vashistha. 2023. Decolonizing Content Moderation: Does Uniform Global Community Standard Resemble Utopian Equality or Western Power Hegemony?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23, Vol. 26)*. ACM, 1–18. https://doi.org/10.1145/3544548.3581538
[77] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, and Gurleen Virk. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23, Vol. 24)*. ACM, 723–741. https://doi.org/10.1145/3600211.3604673
[78] Donghee Shin, Veerisa Chotiyaputta, and Bouziane Zaid. 2022. The effects of cultural dimensions on algorithmic news: How do cultural value orientations affect how people perceive algorithms? *Computers in Human Behavior* 126 (Jan. 2022), 107007. https://doi.org/10.1016/j.chb.2021.107007
[79] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2023. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Transactions on Computer-Human Interaction* 30, 5 (Sept. 2023), 1–57. https://doi.org/10.1145/3511599
[80] Ian René Solano-Kamaiko, Dibyendu Mishra, Nicola Dell, and Aditya Vashistha. 2024. Explorable Explainable AI: Improving AI Understanding for Community Health Workers in India. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3613904.3642733
[81] Jasmina Tacheva and Srividya Ramasubramanian. 2023. AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order. *Big Data & Society* 10, 2 (July 2023), 20539517231219241. https://doi.org/10.1177/20539517231219241 Publisher: SAGE Publications Ltd.
[82] Alex Tamkin, Miles Brundage, Jack Clark, and Deep Ganguli. 2021. Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models. arXiv:2102.02503 [cs.CL] https://arxiv.org/abs/2102.02503
[83] Yan Tao, Olga Viberg, Ryan S Baker, and René F Kizilcec. 2024. Cultural bias and cultural alignment of large language models. *PNAS Nexus* 3, 9 (Sept. 2024). https://doi.org/10.1093/pnasnexus/pgae346
[84] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S. Bernstein, and Ranjay Krishna. 2023. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–38. https://doi.org/10.1145/3579605
[85] Pranav Narayanan Venkit, Mukund Srinath, and Shomir Wilson. 2022. A Study of Implicit Bias in Pretrained Language Models against People with Disabilities. In *Proceedings of the 29th International Conference on Computational Linguistics*, Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na (Eds.). International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 1324–1332. https://aclanthology.org/2022.coling-1.113
[86] Sterling Williams-Ceci, Maurice Jakesch, Advait Bhat, Kowe Kadoma, Lior Zalmanson, and Mor Naaman. 2024. Bias in AI Autocomplete Suggestions Leads to Attitude Shift on Societal Issues. https://doi.org/10.31234/osf.io/mhjn6
[87] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and LC Ray. 2022. AI as an Active Writer: Interaction Strategies with Generated Text in Human-AI Collaborative Fiction Writing 56-65. In *IUI Workshops*. https://api.semanticscholar.org/CorpusID:248301902
[88] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces (IUI '22)*. ACM. https://doi.org/10.1145/3490099.3511105

A PARTICIPANT DEMOGRAPHIC DETAILS

Please refer Table 6.

B FAVORITE CELEBRITIES SELECTED BY PARTICIPANTS

Please refer Table 7.

	Languages	Occupations
India	Bengali, English, French, Garhwali, German, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Urdu, Nepali Marathi, Punjabi, Sanskrit, Tamil, Telugu	Academician/Professor, Accountant/Finance, Doctor, Business Professional, Counsellor, Data Analyst, Designer, Employee (General), IT Professional, Developer/Software Engineer, Entrepreneur, Intern, Freelancer/Consultant, Student, Self-employed, Teacher/Tutor, Manager (Marketing/Sales), Writer
US	Chinese, English, French, German, Indonesian, Italian, Japanese, Korean, Spanish, Vietnamese	Academician, Construction Worker, Consultant, Cook, Copywriter, Data Analyst, Driver, Teacher, Developer/Software Engineer, Finance, Artist, Business Professional, Employee (General), Health Care Worker, Homemaker, IT Professional, Manager (Marketing/Sales), Military, Real Estate, Project Engineer, Retail Worker, Retired, Security, Self-Employed, Student, Transcriber, Transportation, Unemployed, Veterinary Assistant, Video Editor

Table 6. Full list of languages and occupations self-disclosed by our participants

C CHANGING CULTURAL EXPRESSION: ADDITIONAL EXAMPLES

Please refer Table 8.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

No AI (n=24)	**Count**	**AI** (n=36)	**Count**
Shah Rukh Khan	3	Shah Rukh Khan	5
Virat Kohli	2	APJ Abdul Kalam	4
Rajinikanth	2	Priyanka Chopra	1
Lakshay Sen	1	Mahatma Gandhi	1
Konidela Pavan Kalyan	1	Diljit Dosanjh	1
Kangana Ranaut	1	Amitabh Bachchan	1
Salman Khan	1	Allu Arjun	1
Arijit Singh	1	Alia Bhatt	1
P Sainath	1	A.R. Rahman	1
Warren Buffet	1	Yogi Adityanath	1
Tom Hardy	1	Vishwanathan Anand	1
Terrence Tao	1	Virat Kohli	1
Sebastian Vettel	1	Vinesh Phogat	1
Barack Obama	1	Samay Raina	1
Miley Cyrus	1	Sadh Guru	1
Kim Namjoon	1	Sachin Tendulkar	1
Elon Musk	1	Sylvester Stallone	2
Cristiano Ronaldo	1	Ryan Reynolds	2
Zayn Saifi	1	Cristiano Ronaldo	2
Emma Watson	1	Selena Gomez	1
		Keanu Reeves	1
		Rebecca Solnit	1
		Leon Marchand	1
		Christo Xavier	1
		Chamath Palihapitiya	1
		Adam Khoo	1

Table 7. Public figures chosen by Indian participants across the AI and No AI conditions. Blue cells indicate Indian public figures, yellow cells represent non-Indian public figures.

Topic	Essay without AI	Essay with AI
Festival (Onam)	My favorite festival is Onam which is a festival celebrated in my home state, Kerala. One of the specialties of Onam is the sadya. In northern Kerala Onam sadya includes non- vegetarian dishes and in central and southern Kerala it is mainly a vegetarian affair. I usually prepare the non-vegetarian sadya along with two types of payasam . We also wear new clothes on the day called "onakkodi" . Onam is a harvest festival.	I celebrate onam which is a state festival of kerala. **We decorate our homes with flowers**. and prepare traditional meals **called Onasadya. We also wear traditional attire.** The amain highlight is get-together with family and friends **and participating in cultural events .** In every nook and corner of the state there will cultural events and games which can be enjoyed without any discrimination and **with great enthusiasm .**
Festival (Pongal)	My Favorite Festival is Pongal which is celebrated in January. We celebrate with our family relatives for 3 days- Boggi, Kanuma, Makara Sankranti. 1st Boggi is celebrated abound firewood with relatives and neighbours and then kanuma Sankranthi which is celebrated by drawing colours on the grounds and visiting the temple and Makara Sankranthi which we eat together and enjoy together.	My cultural festival is Pongal. We celebrate this festival to **honor the sun god and the harvest.** We have three days to celebrate three pongal **days: Bhogi, Surya, and Mattu Pongal.** We celebrate with **delicious food, family gatherings, and traditional dances .** All wear new **clothes and decorate their homes .** We clean **our homes thoroughly and draw beautiful kolams.**
Festival (Ugadi)	My favorite festival is Ugadi. Being the first year for Telugu people, it is a major festival. We used to wear new dress and prepare and consume Ugadi Pachhadi . This helps me to identify and accept both difficulties and happiness are part of life and need to accept them as a part. Also there is a panchanga shravanam where I listen to this to find how my future looks in the present year.	My favorite festival is Ugadi because it is new year day **for many South Indian states.** I ce**lebrate it with traditional dishes and family gatherings.** I **also enjoy decorating the house with flowers .** I get up early and take an head bath wear new clothes and do puja or anustana at my home. **with devotion and joy** . I offer **prayers and special dishes to the deities** . A special portion called Ugadi Pachadi is prepared and offered first to my presiding deities **to seek their blessings for the new year.** Then it will be consumed **by everyone in the family as a ritual.** It **marks the beginning of a prosperous year for us.**
Food (Dosa)	My favorite food is ghee masala dosa , a South Indian cuisine. I like it because it is very tasty and healthy. Firstly, it is made with clarified butter which is healthy for digestion. Secondly, it is accompanied with lentil soup called sambar which has lots of yummy flavor and a good mix of healthy veggies. It is also accompanied with a number of chutneys, which are a paste of gram dal , coconut, tomato , onion , garlic and coriander . It is very delicious and healthy. Finally. I like ghee masala dosa because it is easy to digest and at the dame time full of gastronomic flavors.	My favorite food is Dosa **because it's crispy and delicious.** It can be filled with different ingredients **like potatoes or cheese .** It's nutritious, easy to digest, low in fat and culturally significant, making it a popular choice for many **people in India and beyond .** D**osa is also versatile and can be enjoyed with various chutneys and sambar.**

Table 8. Additional examples from the content analysis comparing cultural artifacts described by Indian participants with and without AI suggestions. AI-generated text is shown in **bold**. Highlights indicate key differences: green for cultural details present without AI but missing with AI, yellow for generic phrasing introduced by AI, blue for Westernized descriptions.