

Exploring the Impact of Artificial Intelligence-Mediated Communication on Bias and Information Loss in Non-academic and Academic Writing Contexts

Bryce Petofi Towne

Abstract

This paper aims to explore the impact of Artificial Intelligence-Mediated Communication (AI-MC), specifically through large language models (LLMs) like ChatGPT, Claude, and Gemini, on positivity bias and information loss in non-academic and academic writing contexts. Through a dual-study approach involving both non-academic and academic texts, the research will assess whether texts refined by LLMs exhibit a significant positivity bias compared to original texts and if they exacerbate the problem of information loss in text-based communication. Empirical tests will determine the presence of these phenomena by analyzing participants' perceptions and understanding of AI-refined versus original texts. Confirmation of the hypotheses would highlight the nuanced effects of AI-MC on communication accuracy and integrity, while disconfirmation would suggest the potential of LLMs to maintain objectivity and information fidelity in written discourse.

Introduction

In an era where the digital landscape is constantly evolving, the fusion of artificial intelligence (AI) with communication processes has given rise to a novel paradigm known as Artificial Intelligence Mediated Communication (AI-MC). This emergent field stands at the nexus of technological advancement and human discourse, transforming the ways in which messages are constructed, disseminated, and interpreted. The significance of AI-MC extends beyond mere technological innovation; it heralds a profound shift in the dynamics of interpersonal interactions and the fabric of communicative practices. As AI technologies, particularly large language models (LLMs) like ChatGPT, Claude, and Gemini, become increasingly integrated into various facets of writing—including academic and scientific writing—the implications of such integration warrant meticulous scrutiny.

The literature on AI-MC, notably contributions from Hancock et al. (2020), Hohenstein & Jung (2018), and subsequent studies, underscores the intricate role of AI in enhancing, altering, or generating messages to achieve specific communicative objectives. These studies collectively highlight the depth of AI's integration into communication and its dual impact on efficiency and the reshaping of linguistic norms and social dynamics. Moreover, the exploration of AI-MC's influence on human cognition and interpersonal dynamics points to its capacity to both enrich and complicate human interactions, necessitating a nuanced understanding of its ethical, cultural, and policy implications.

However, despite the expansive discourse on AI-MC's broader implications, there exists a notable gap in research specifically addressing its role in the realm of non-academic and academic/scientific writing. This omission is peculiar, given the widespread recognition and application of AI-MC in these domains, even amidst ethical dilemmas concerning authenticity and academic integrity. The potential of AI to transform scholarly communication—by enhancing writing efficiency, overcoming language barriers, and fostering diversity in text creation—brings to light ethical concerns, accuracy doubts, and plagiarism risks. These concerns underscore the need for a cautious yet open approach to the integration of AI in scholarly endeavors.

This paper seeks to bridge this gap by positing that the employment of LLMs for tasks such as refining, rewriting, or directly composing articles or academic papers should be recognized as instances of AI-MC. It argues that utilizing LLMs for writing assistance not only epitomizes an AI-mediated stage in the communication process (writing to reading) between humans but also engages with the core issues of concern within AI-MC research, such as technology design, effectiveness, and accompanying ethical and societal impacts.

Given the transformative potential of AI-MC in academic settings, this introduction sets the stage for a detailed examination of the role of LLMs like ChatGPT, Claude, and Gemini in scholarly communication. It raises critical research questions regarding the presence of positivity bias in LLM-generated texts and the challenge of information loss in text-based communication. These research questions aim to elucidate the impact of advanced AI technologies on writing and academic writing, thereby contributing to the ongoing discourse on the ethical, cultural, and technological considerations of AI-MC in scholarly communication.

Literature Review

AI-MC stands at the confluence of technological innovation and human discourse, heralding a significant shift in the construction, transmission, and interpretation of messages. As delineated by Hancock et al. (2020), AI-MC emerges as a pivotal force in mediated communication, capable of enhancing, altering, or generating messages to fulfill distinct interpersonal or communicative objectives. This framework accentuates the intricate role of AI in both facilitating and complicating human interactions, wherein computational agents undertake actions on behalf of communicators to achieve varied communicative ends.

The progression of AI-MC from elementary text-based enhancements, such as auto-correct and predictive text, to sophisticated applications including smart replies and auto-completion (Hancock et al., 2020), underscores the depth of AI's integration into communicative practices. This evolution not only showcases the technological finesse involved but also prompts critical reflection on AI's impact on linguistic norms, interpersonal trust, and the ethical contours of communication. The advent of functionalities like Gmail's smart replies, which provide pre-generated email responses, illustrates AI-MC's dual impact by offering efficiency and potentially reshaping linguistic patterns and social dynamics (Hancock et al., 2020).

Moreover, the interplay between AI-MC and human cognition is profound. The interactive alignment model suggests that AI-generated text can significantly influence linguistic alignment, potentially altering not just lexical choices but also semantic content and social interactions (Pickering & Garrod, 2013; Hancock et al., 2020). This effect is further complicated by AI-MC systems' tendency towards a positivity bias, as evidenced by Hohenstein & Jung (2018), where suggestions for smart replies in text messaging exhibited an overly positive tone, potentially inducing shifts in language norms and interpersonal dynamics.

The exploration of AI-MC's role in interpersonal dynamics underscores its capacity to both complicate and enrich human interactions. Li, J., Chu, Y., & Xu, J. (2023) underscore the profound influence of AI's fairness within AI-MC contexts on human impression formation, indicating that AI behavior, when aligned with social norms of fairness, can significantly improve interpersonal perceptions and relationships. Conversely, Glikson & Asscher (2023) highlight the challenges AI-MC poses to perceived authenticity and forgiveness in multilingual work contexts, emphasizing the delicate equilibrium between leveraging AI's capabilities and preserving the authenticity of human expression.

Accessibility and equitable adoption of AI-MC tools surface as critical concerns, with Goldenthal et al. (2021) identifying barriers to AI-MC access and literacy that could impede the widespread and equitable utilization of AI technologies. This issue highlights the necessity of formulating inclusive strategies to ensure the benefits of AI-MC are accessible across all societal segments.

Thus, the integration of AI into mediated communication signifies a notable advancement with extensive implications for language usage, interpersonal relationships, and the ethical framework of communication. Although AI-MC presents unparalleled opportunities for enhancing communication efficiency and effectiveness, it concurrently demands meticulous consideration of its potential to modify social norms, affect interpersonal trust, and introduce ethical quandaries. Consequently, the scholarly examination of AI-MC must persist in evolving, tackling these challenges while exploiting AI's potential to enrich human communication.

The primary challenges encompassing AI-MC involve:

1. **The Impact on Human Language and Cognition:** AI-MC has the potential to transform human language usage and cognitive processes, guiding specific grammatical and semantic responses through functionalities like Gmail's smart replies, which could lead to shifts in language norms and expectations.
2. **Complexities in Interpersonal Dynamics and Impression Formation:** AI's intervention in human communication can affect interpersonal trust and the authenticity of expressions. Utilizing large language models may diminish the authenticity of communications, engendering skepticism towards AI-mediated apologies and impacting relationships.
3. **Reassessment of Online Self-presentation and Trust:** AI's involvement in crafting online profiles and messages adds complexity, potentially eliciting concerns over deceit and manipulation. Profiles considered to be AI-generated may be deemed untrustworthy, affecting impression formation and trust in online environments.
4. **Ethical, Cultural, and Policy Implications:** The replication of existing biases in AI systems could reinforce societal power structures and normalize certain modes of communication while marginalizing others. Additionally, balancing the need for transparency in AI-MC and protecting freedom of speech against ensuring the ethical use of AI-MC technologies presents significant ethical considerations.

5. Positivity Bias: AI's propensity to use overly positive language forms a core concern, as this positivity bias, exemplified by Hohenstein & Jung's study(2018) on "smart reply" suggestions in text messaging revealed that they were overly positive ("sounds great!"). could lead to overly positive perceptions of scientific articles, thereby influencing reader perceptions.

However, the challenges of AI-MC involvement in writing and academic/scientific writing are much less discussed in the field of communication studies, despite AI-Mediated writing is quite an important topic in recent scholarly literature.

This paper argues that employing large language models (LLMs) such as ChatGPT and Claude for refining, rewriting, or directly composing articles or academic papers should be recognized as instances of AI-MC. AI-MC involves the use of artificial intelligence systems to modify, enhance, or generate content to achieve communication and relational goals. The use of LLMs for writing assistance exemplifies an AI-mediated stage in the text-based communication process between human writers and readers, raising concerns over the quality and efficiency of information expression and sparking profound discussions on how AI impacts human communication methods, content creation quality, and recipient perception.

Moreover, AI-MC research focuses on the design of these technologies and their psychological, linguistic, interpersonal, policy, and ethical impacts on human communication. Thus, utilizing LLMs for writing or editing tasks not only embodies the definition of AI-MC but also engages with the core issues of concern within AI-MC research, such as technology design, effectiveness, and accompanying ethical and societal impacts (Hancock, Naaman, & Levy, 2020).

Despite extensive discourse on AI-MC's broader implications, there remains a notable scarcity of research specifically targeting its role in facilitating the writing of articles and academic papers. This gap is peculiar, considering the widespread recognition and application of AI-MC in these domains, even amid ethical dilemmas concerning authenticity and academic integrity (Fitria, T. N., 2023; Chen, T.-J., 2023; Miao et al., 2024).

AI-MC has transformed various facets of human interaction, especially in writing and scholarly activities. While research in communications and human-computer interaction often emphasizes the challenges and ethical concerns associated with AI-MC, such as its influence on language and thought, ethical and policy implications, and the reevaluation of trust and authenticity online, there is a notable discrepancy in the embrace of AI tools like ChatGPT in academic writing (Fitria, T. N., 2023; Chen, T.-J., 2023; Miao et al., 2024).

Studies highlight AI's efficacy in enhancing writing efficiency, overcoming language barriers, and generating diverse text versions (Chen, T.-J., 2023; Kacena, M. A. et al., 2024). However, these advancements are accompanied by ethical concerns, accuracy doubts, and plagiarism risks, advocating for a cautious yet open approach to AI integration in scholarly endeavors (Miao et al., 2024; AlAfnan et al., 2023).

The ethical challenges and implications of AI integration in academic settings are profound, with discussions on academic integrity, transparency, and the formulation of ethical guidelines taking precedence (Miao et al., 2024; Thorp, 2023). Furthermore, AI's potential to exhibit a positive bias raises questions about its impact on scholarly discourse and the integrity of scientific communication.

The trust in AI-MC, particularly in academic writing, necessitates a reevaluation. The acceptance of AI tools in scientific writing, in spite of known challenges, indicates a significant paradigm shift in the scholarly community's perception of trust and credibility. This shift demands a thorough understanding of AI's capabilities and limitations to ensure responsible usage (Herbold et al., 2023; Balel, 2023).

The divergent perspectives on AI-MC in communication studies versus academic writing underline a complex interplay of ethical, cultural, and technological considerations. The widespread acceptance of AI tools like ChatGPT in academic settings, despite existing challenges, suggests an evolving landscape of scholarly communication. This scenario underscores the importance of ongoing research,

ethical deliberation, and policy development to navigate effectively the future of AI-MC in academic discourse (Dwivedi et al., 2023; Thorp & Vinson, 2023).

Research Gaps and Research Hypothesis

The exploration of AI-MC in academic settings, particularly concerning the role of LLMs like ChatGPT, Claude, and Gemini, introduces a complex paradigm shift in scholarly communication. Following the foundational insights provided in the literature review section on the diverse implications and challenges of AI-MC, this chapter delves into positivity bias and information loss within writing and academic writing aspects of AI-MC.

Positivity bias, a concern previously flagged in various communicative contexts (Hohenstein & Jung, 2018), warrants a reevaluation in the context of advanced LLMs' application to academic writing. The optimistic stance of the academic community towards LLMs in scholarly writing, as indicated by the relative absence of concern for positivity bias in previous research, prompts an inquiry into whether advancements in AI capabilities have mitigated this issue.

Diamond (2024) argues that systems like autocorrect, autocomplete, and smart replies have become cornerstones of modern text communication. While these systems provide significant assistance day-to-day, they primarily focus on simple tasks like response prediction, spelling corrections, or sentence completion. With the sudden rise in advanced generative AI—namely large language models (LLMs) like GPT-4 and LLaMa 2—the door has opened for smarter and more capable AI assistance systems for digital writing composition.

However, the question remains: Does the advanced technology of LLMs perpetuate or mitigate positivity bias and its associated risk of information loss in academic/scientific writing? This concern is crucial because academic/scientific writing demands precision and objectivity, with any form of bias potentially skewing reader perception and distorting the author's intended message.

Moreover, the challenge of information loss in text-based communication—a phenomenon well-documented in literary and communication studies—gains a new dimension with the intervention of LLMs. Studies have shown that discrepancies between authorial intent and reader interpretation are commonplace, leading to varied understandings of the same text (Pisanty, 2015; Gibbs, 2001; Rosebury, 1997; Katz & Lee, 1993; Horváth, 2015). This discrepancy, termed in this paper as *information loss in text-based communication*, raises pertinent questions about LLMs' role in either exacerbating or alleviating this fundamental challenge of communication.

To address these concerns, this chapter proposes two research questions aimed at critically examining the impact of LLMs like ChatGPT, Claude, and Gemini on writing and academic writing:

Research Questions 1 (RQ1)

RQ1: Does texts generated by LLMs, such as ChatGPT, Claude, and Gemini, exhibit a significant bias when compared to the original texts?

This research question and its null counterpart allow for an empirical test of whether LLM-generated texts are characterized by a tendency towards more positive language compared to original human-authored texts. This is grounded in the observation of potential positivity bias in AI-generated content, as noted in prior research.

Research Questions 2 (RQ2)

RQ2: Does LLM-generated texts, such as those from ChatGPT, Claude, and Gemini, in comparison to original texts, exacerbate the problem of information loss in text-based communication?

This research question is designed to investigate the effect of LLMs on the fidelity of information transmission in text-based communication. Specifically, they aim to determine if texts generated by LLMs lead to greater or lesser information loss compared to original texts, addressing concerns about the accuracy and integrity of AI-mediated communication.

Methods

Ethics information

This research complies with all relevant ethical regulations for work with human participants and has been approved by the Institutional Review Board (IRB) of Hephaestus Education Technology Ltd. Informed consent will be obtained from all participants involved in the study. Participants will be compensated with money compensation for their time and effort. The informed consent form will detail the study's purpose, procedures, potential risks, benefits, and the confidentiality of participant data.

Design

The experimental design includes two separate studies, each aimed at exploring the impact of AI-MC on positivity/negativity bias and information loss in both non-academic and academic writing contexts.

Participants

Participants for both studies will be recruited online via Credamo, a professional survey platform in China. Inclusion criteria include being 18 years of age or older, with Study 1 requiring native speakers of Chinese, and Study 2 participants additionally requiring experience in academic writing or peer review. Exclusion criteria include prior participation in similar studies to avoid learning effects.

Participants will be randomly assigned to either the experimental or control group by the platform Credamo to ensure equal distribution across conditions and mitigate selection bias. This randomization process will be documented and verifiable.

Dual Design

Study 1: Non-Academic Text Writing and Reading

Design Type: Mixed design, with participants randomly divided into two main groups: those reading AI-refined texts (experimental) and those reading original texts (control).

Blinding: To ensure objectivity, participants will not be informed whether the texts they read have been AI-refined until the end of the experiment. However, data collection and analysis will not be conducted blind to the conditions of the experiments as the scoring standards are objective and transparent.

Study 2: Academic Text Writing and Reading

Design Type: Mixed design, similar to Study 1, but with texts of an academic nature and participants are required to read both AI-refined texts and original texts. There is no control group.

Blinding: Similar to Study 1, Participants will not be informed whether texts they read have been AI-refined until the end of the experiment. However, data collection and analysis will not be conducted blind to the conditions of the experiments as the scoring standards are objective and transparent.

Procedures

Study 1:

1. **Writing Phase:** Writers will write texts conveying specific emotions (Study 1A) or on assigned topics (Study 1B). Writing experts will create texts based on fact description, opinion statement, and emotional expression.

2. **Refinement Phase:** Selected texts will be refined using AI models with their latest versions (ChatGPT 4.0, Claude 3 Opus, and Gemini Advanced) based on predefined prompts incorporating the authors' intentions.
3. **Reading and Evaluation Phase:** Randomly assigned readers will evaluate the texts, answering questions designed to assess their perception of the text's emotional content or clarity.

Study 2:

1. **Writing Phase:** Academic researchers will write texts containing literature discussions and/or introductions.
2. **Refinement Phase:** Texts will be refined by the same AI models using prompts that include the authors' key points and findings.
3. **Reading and Evaluation Phase:** Participants will read the original and AI-refined texts, answering questions that assess their understanding and the texts' academic suitability.

Quality Checks

- **Outcome-neutral criteria:** The distribution of scores on participant responses will be monitored for the absence of floor or ceiling effects.
- **Positive controls:** Sample texts known to convey clear emotional or factual content will be used to validate the questionnaires.

Sampling Size

The sample size for both studies was calculated using the formula ($n=Z^2 \cdot p \cdot (1-p)/E^2$), with $Z=1.96$ for a 95% confidence level, $p=0.5$ to maximize sample size, and $E=0.05$ for a 5% margin of error, resulting in a required sample size of approximately 385. To account for potential dropouts and ensure statistical power, we aim for 400 participants for each study.

Inclusion and exclusion criteria are defined objectively. Participants must be aged 18 or older and have proficiency in the language of the study (either English or Chinese, depending on the participant group). Exclusion criteria include incomplete consent forms, incomplete participation, or technical issues that prevent full engagement with the study materials. Data exclusion will be predetermined, and replacement of participants will occur in cases of technical failure or withdrawal.

Data availability

Data collected during this study, including de-identified participant responses and AI-modified texts, will be stored in Figshare. Data availability will be subject to ethical considerations and participant consent, with the intention of making the dataset available to other researchers upon reasonable request.

Code availability

The AI modifications will be made using publicly available versions of ChatGPT 4.0 Workplace, Claude 3 Opus, and Gemini Advanced, without any custom code. Details of the AI tool versions and settings will be included in the Supplementary Information file, ensuring reproducibility of the AI-mediated modifications.

Results

The results are separated as Study 1 and Study 2.

Study 1

Study 1 investigated the effects of AI-mediated communication, specifically through text refinement by LLMs: ChatGPT 4.0, Claude 3 Opus, and Gemini Advanced, on perceived bias and information loss in non-academic text contexts. The study employed a mixed-methods design, incorporating both normality tests and nonparametric tests to analyze the data collected from a control group (original texts) and three experiment groups (texts edited by each LLM).

Normality of Data Distribution

Initial analysis using the Kolmogorov-Smirnov and Shapiro-Wilk tests revealed a significant departure from normal distribution across all datasets ($p < 0.01$), indicating a pronounced non-normality in the distribution of scores related to perceived bias and information loss. Specifically, skewness and kurtosis values further affirmed the non-normal distribution of data, necessitating the adoption of nonparametric methods for subsequent analyses.

Bias Perception

The nonparametric Kruskal-Wallis H test was utilized to compare the median scores of correct responses across the control and experiment groups. Results indicated no statistically significant differences in bias perception among the original texts and those edited by Gemini Advanced, Claude 3 Opus, and ChatGPT 4.0 ($p > 0.05$). This outcome suggests that the intervention of LLMs in text refinement does not significantly alter the perceived bias in non-academic texts from the perspective of the readers.

Information Loss

Similarly, the analysis of information loss employed the Kruskal-Wallis H test to evaluate differences among the groups. Consistent with the findings on bias perception, the results showed no significant differences in information loss between the control group and the texts edited by the respective LLMs ($p > 0.05$). This finding indicates that the text refinement by LLMs neither exacerbates nor mitigates the issue of information loss in non-academic texts, as perceived by the readers.

References

1. AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). Chatgpt as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, 3(2), 60-68.
2. Balel, Y. (2023). The Role of Artificial Intelligence in Academic Paper Writing and Its Potential as a Co-Author: Letter to the Editor. *European Journal of Therapeutics*, 29(4), 984-985.
3. Chen, T.-J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. *Journal of Chinese Medical Association*, 86(4), 351-353.
<https://doi.org/10.1097/JCMA.0000000000000900>
4. Diamond, N. (2024). AI Does Not Alter Perceptions of Text Messages. arXiv:2402.01726v2.
5. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.
6. Fitria, T. N. (2023, March). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. In *ELT Forum: Journal of English Language Teaching* (Vol. 12, No. 1, pp. 44-58).
7. Glikson, E., & Asscher, O. (2023). AI-mediated apology in a multilingual work context: Implications for perceived authenticity and willingness to forgive. *Computers in Human Behavior*, 140, 107592.
8. Goldenthal, E., Park, J., Liu, S. X., Mieczkowski, H., & Hancock, J. T. (2021). Not all AI are equal: Exploring the accessibility of AI-mediated communication technology. *Computers in Human Behavior*, 125, 106975.
9. Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, 25(1), 89-100.
10. Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13(18617).
11. Hohenstein, J., & Jung, M. (2018, April). AI-supported messaging: An investigation of human-human text conversation with AI support. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems* (pp. 1-6).
12. Kacena, M. A., Plotkin, L. I., & Fehrenbacher, J. C. (2024). The Use of Artificial Intelligence in Writing Scientific Review Articles. *Current Osteoporosis Reports*, 22(115-121).
13. Li, J., Chu, Y., & Xu, J. (2023). Impression transference from AI to human: The impact of AI's fairness on interpersonal perception in AI-Mediated communication. *International Journal of Human-Computer Studies*, 179, 103119.
14. Miao, J., Thongprayoon, C., Suppadungsuk, S., Garcia Valencia, O. A., Qureshi, F., & Cheungpasitporn, W. (2024). Ethical Dilemmas in Using AI for Academic Writing and an Example Framework for Peer Review in Nephrology Academia: A Narrative Review. *Clin. Pract.*, 14(1), 89-105.
15. Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, 36(4), 329-347.
16. Thorp, H. H. (2023). ChatGPT is fun but not an author. *Science*, 379(6630), 313.
17. Thorp, H. H., & Vinson, V. (2023, November 16). Change to policy on the use of generative AI and large language models. *Science Blog*. Retrieved from <https://www.science.org/content/blog-post/change-policy-use-generative-ai-and-large-language-models>
18. Gibbs, R. (2001). Authorial Intentions in Text Understanding. *Discourse Processes*, 32(73-80).
19. Horváth, M. (2015). Authorial intention and global coherence in fictional text comprehension: A cognitive approach. *Semiotica*, 2015(39-51).

20. Katz, A., & Lee, C. J. (1993). The Role of Authorial Intent in Determining Verbal Irony and Metaphor. *Metaphor and Symbol*, 8(257-279).
21. Pisanty, V. (2015). From the model reader to the limits of interpretation. *Semiotica*, 2015(37-61).
22. Rosebury, B. (1997). Irrecoverable intentions and literary interpretation. *British Journal of Aesthetics*, 37(15-30).