



Does AI Chatbot Have A Conversation Style?

A Corpus-Based Analysis on AI-Generated Conversation Material

Hanliang Chen¹, Wanbo Ren*²

School of Foreign Languages, Northwest University, Xi'an, China

Email: ¹hlc2003@nwu.edu.cn, ²2423275980@qq.com

Abstract. 2023 has become a milestone for AI, yet the style of AI chatbots remain underexplored. This study takes the register framework proposed by Biber in investigating the conversation style of ChatGPT, Claude and Copilot. The result shows that AI chatbots vary in terms of ability in imitating conversation, and each AI chatbot has a specific conversation style.

Keywords: Artificial intelligence; multidimensional analysis; conversation style; discourse analysis

1 Introduction

2023 is said to be the first year where AI goes to people's life, with the release of ChatGPT at the end of 2022, the whole year of 2023 is filled with news about AI. While AI goes viral on the internet, it is indeed influencing our life at the same time. AI-generated content is frequently seen, and often acts as a sidekick for illegal activities like rumor spreading. Thus, it gradually emerges as necessary to investigate AI, not from a perspective to figure out its mechanism, but to get a better knowledge of its output to keep the vulnerable people's interest from being compromised.

One way to categorize AI is by modality. Certain models can only accept and generate text information, while others might be able to deal with both text and images. For the limited device of the researcher, current study only involves the study of text.

As there are many different AI models out there, it is necessary to consider their popularity and influence when deciding which could be our research subject. The researcher turned to Eldin for subject selection^[4]. Eldin shows that ChatGPT, Claude, Perplexity, Google Bard and Microsoft Bing Chat are the top 5 most popular AI chatbot in 2023^[4]. Among these 5 chatbots, Google Bard is still under test, and Perplexity shares the same model with Claude, thus, the current study selects ChatGPT, Claude and Microsoft Bing Chat as research subject. The current research aims to answer the following questions: how well could AI chatbot imitate daily conversation style? And do AI chatbots have a conversation style? That is, do different chatbots share the same style in terms of conversation, or do they each have a specific, peculiar style? To answer this question, the researcher examined AI related linguistic researches, which is shown in section 2. The whole process of the research is presented in section 3. Section 4 shows

© The Author(s) 2024

I. A. Khan et al. (eds.), *Proceedings of the 2024 2nd International Conference on Language, Innovative Education and Cultural Communication (CLEC 2024)*, Advances in Social Science, Education and Humanities Research 853, https://doi.org/10.2991/978-2-38476-263-7_26

the result, and section 5 extend a discussion based on the result. Finally, section 6 concludes the article with the conclusion.

2 Literature Review

Most linguistic studies on AI treat them as a new variable that could impact the societal system, and a new social agent that is either constructed through discourse or construct others. Anderson discussed the current metaphors used to construct the identity of ChatGPT, which pointed out that the tool metaphor and the collaborator metaphor are both deficit to some extent and fail to establish the necessary conceptual framework for student to use it^[1]. The author advocates the medical and surgical metaphors in conceptualizing the work that ChatGPT does on human input. Koh investigated the treatment an AI chatbot with a predisposed female identity received on online discussion on South Korea^[6]. The research reveals that the chatbot is used as a tool for online masculinity construction.

Compared with society-oriented studies, the style of AI chatbots received less attention, leaving this field underexplored. The first investigations concerning AI chatbot style is conducted between human and chatbots, whose aim is to discover if human can discern AI generated content and human generated content. In terms of academic texts, Ma et al. used ChatGPT as research article abstract generator, and asked human to judge if the text is generated by AI^[7]. Human judgement exhibited a high precision in this register. Similarly, Theocharopoulos et al. used NLP tools like word embedding in text classification, and the result is similarly accurate^[8]. However, when the register changes, the result also varies. Köbis & Mossink used GPT-2 to generate poems with first line given to it^[5]. The result shows that human cannot discern the generator of these poems, and even when told the generator, they like the poems generated by AI more. The review of researches concerning AI style indicates that AI is good at creating fictional and artistic content, while lacking enough technique in academic writing.

This leads to the research of the current research. While previous researches have examined academic genre and fictional genre, daily conversation remains uninvestigated. Thus, the current research takes daily conversation as research subject, and aims to answer the following 2 questions:

- (1): how well could AI chatbots imitate daily conversation?
- (2): do AI chatbots share the same style? Or do they each has a specific style?

3 Method

3.1 Data Collection

To obtain a corpus of conversation-style text generated by AI chatbots, a prompt was provided to ChatGPT, Claude, and Copilot instructing them to "Write a conversation between people of about 1000 words in length". This prompt-based approach of eliciting desired content from language models is commonly used in AI text generation research, as employed by Berber Sardinha^[2].

Each of the three chatbots received the same prompt and generated 10 distinct conversational samples in response, resulting in a total data set of 30 pieces of AI-generated conversational text - 10 from ChatGPT, 10 from Claude, and 10 from Copilot. The researcher took steps to ensure each response was fully completed without being cut off.

To facilitate analysis, the 30 conversational samples were saved into individual plain text (.txt) files, with one file per sample. Metadata such as the prompt, chatbot name, and any other identifying information was removed from the files to enable blind analysis solely based on the text content itself.

The use of 10 samples per chatbot was determined to provide a reasonably representative snapshot of each system's conversational output capabilities, while remaining a manageable corpus size for the multidimensional analysis. Future studies could consider analyzing more samples to increase statistical power.

Additionally, no other constraints or specifications were provided in the prompt beyond the target word count and conversational framing. This open-endedness allowed the chatbots to freely generate dialogue content based on their training without too many limiting parameters. However, it is possible that providing more scaffolding in the prompt could have influenced the resultant conversation styles.

With the 30 conversation text files prepared, the next stage involved processing them through the multidimensional analysis tagger (MAT) to quantify their linguistic characteristics and identify their closest genre associations, as detailed in the following section..

3.2 Multidimensional Analysis

To measure the style of certain text, the current study takes the framework proposed by Biber, which is originally a 6-dimension framework for register analysis^[3]. The framework measures the similarity and difference of a certain text or texts based on 6 dimensions: involved versus informational production, narrative versus non-narrative concerns, explicit versus situation-dependent reference, overt expression of persuasion, abstract versus non-abstract information and online information elaboration. The framework is based on lexical dispersion, which means each dimension includes certain lexicons, thus making it possible to measure the strength of a certain text based on the 6 dimensions.

This means the texts must be tagged before one can get the specific value of the 6 dimensions. The current research uses MAT (multidimensional analysis tagger) for text tagging. After the process is finished, the values of each chatbot are compared, first with other styles like conversation and academic writing, which assign the texts the closest style it is next to, and then, the values are compared within the 3 chatbots to see if they share the same style.

4 Result

The multidimensional analysis done by MAT first compared the AI-generated conversation data against 8 established text categories: conversation, broadcasts, prepared speeches, personal letters, general fiction, press reportage, academic prose, and official documents. For each of the 6 dimensions, the AI outputs were scored based on their similarity to the typical characteristics of those 8 genres. Ultimately, MAT determines which of the 8 text types is the closest stylistic match for the analyzed material.

According to MAT's analysis, the conversational output generated by ChatGPT showed no meaningful similarity to the actual conversation genre across any of the 6 dimensions. Instead, ChatGPT's closest textual resemblance was to "general narrative exposition" prose. This suggests ChatGPT's responses tended towards explanatory, storytelling language rather than capturing the interactive real-time dynamics of dialogue.

Claude's conversational output also failed to align with any dimensions of the conversation genre. The closest match for Claude was the "involved persuasion" category, indicating its responses featured traits of making claims, arguments, and attempts to influence the reader's perspective. However, key components of back-and-forth conversational pragmatics appear to be lacking.

In contrast, Copilot demonstrated substantial overlap with actual conversation style on two of the dimensions: 1) Involved vs Informational Production, and 2) Abstract vs Non-abstract Information. This means Copilot's responses exhibited many of the interpersonal involvement, situational grounding, and tangible content characteristics typical of authentic conversational exchanges. As a result, MAT identified "informational interaction" as the closest overall text type match for Copilot's output.

The specific dimensional scores and closest text type categorizations for each chatbot are presented in Table 1. These quantitative results clearly delineate that while Copilot came closest to approximating conversational style, ChatGPT and Claude deviated significantly, albeit in different ways based on their respective strongest dimensionalities.

Table 1. Style information of the 3 chatbots

Model Name	Dimension1	Dimension2	Dimension3	Dimension4	Dimension5	Dimension6	Closest Text Type
ChatGPT	-2.03	3.73	2.5	-4.31	-1.07	-1.57	General narrative exposition
Claude	16.51	-1.99	-2.49	1.79	-2.09	-1.87	Involved persuasion
Copilot	29.61	-4.77	-0.37	-3.23	-3.15	-3.01	Informational interaction

5 Discussion

With the data presented in section 4, we are now able to answer the two research questions. Certain AI chatbots demonstrate stronger abilities than others in imitating natural conversation style. This variation in performance directly indicates that AI chatbots do not share a uniform conversation style, but rather each one exhibits its own distinct stylistic patterns when generating conversational text.

The finding that ChatGPT performs the worst at conversational style is somewhat surprising given its immense popularity and visibility as one of the pioneering viral AI chatbots in 2023. There could be several potential reasons underlying ChatGPT's struggles with this communicative register. One possibility is that the training data used skewed more towards formal, expository text rather than colloquial dialogue. Another factor could be the size and complexity of the underlying language model itself - while powerful, it may still have limitations in fully capturing the nuances and pragmatics of casual conversation.

In contrast, Copilot's stronger conversational abilities suggest its training prioritized more naturalistic dialogue sources. The "informational interaction" textual similarity implies Copilot's responses have many of the involvement, situational grounding, and non-abstract qualities characteristic of actual spoken exchanges between people.

Claude occupying a middle ground of "involved persuasion" hints that its conversational outputs tend towards trying to sway the user's stance, perhaps making them more argumentative or opinionated compared to simple information transfer. This could arise from the model being tuned for tasks requiring reasoning and stance-taking.

Importantly, the performance fluctuations and stylistic idiosyncrasies observed between chatbots are likely impacted by frequent model updates and refinements by their developers. It's quite possible that the conversational abilities of any given chatbot evaluated here could shift in future iterations as the AI systems evolve rapidly.

Beyond the technical factors, these findings raise important implications around AI system design and optimization for targeted use cases. If an AI chatbot is intended for open-ended conversational interactions, clearly mimicking naturalistic dialogue should be a key objective. However, if the goal is question-answering or task-oriented commands, then prioritizing clarity and preciseness over conversational naturalness may be preferable. These contrasting objectives pose fascinating challenges for future AI development.

6 Conclusion

By applying Biber's multidimensional analysis to AI chatbots' output, this study investigated the conversation style of 3 popular chatbots - ChatGPT, Claude, and Copilot. The results revealed that some AI chatbots are better than others at imitating natural conversation style, and each chatbot exhibits a distinct style when generating conversational text.

Copilot demonstrated the strongest ability to mimic authentic conversation, with its output showing similarity to the "informational interaction" text type in terms of the

involved vs informational production and abstract vs non-abstract information dimensions. On the other hand, ChatGPT performed the worst, with its closest textual similarity being "general narrative exposition" - quite far from conversational prose. Claude fell somewhere in between, producing output most akin to "involved persuasion" text.

The variation in conversational abilities likely stems from differences in the training data and underlying language models used by each chatbot. As "black box" systems, the exact factors and processes that lead to these stylistic divergences remain unclear. However, the findings underscore that while AI can generate conversational text, the quality and naturalness can vary significantly depending on the specific system.

One limitation of this research is the relatively small corpus size of 30 conversation samples per chatbot. Future studies with larger datasets would provide more robust and generalizable results. Additionally, investigating other facets beyond style, such as coherence, context-appropriateness, and avoidance of problematic content could further illuminate the strengths and shortcomings of current AI conversation generation capabilities.

Overall, this exploratory study highlights that as AI language systems proliferate, understanding their output characteristics - including conversational styles - will be crucial for properly interpreting and leveraging the content they produce. Multidimensional analysis offers a valuable framework for such textual evaluations across AI systems and use cases.

Reference

1. Anderson, S. S. (2023). "Places to stand": Multiple metaphors for framing ChatGPT's corpus. *Computers and Composition*, 68, 102778. <https://doi.org/10.1016/j.comp-com.2023.102778>
2. Berber Sardinha, T. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, 4(1), 100083. <https://doi.org/10.1016/j.acorp.2023.100083>
3. Biber, D. (1991). *Variation Across Speech and Writing*. Cambridge University Press.
4. Eldin, A. B. (2023, August 23). The Top AI Chatbots in 2023: Capabilities, Use Cases, and Limitations. *Medium*. <https://medium.com/@neltac33/the-top-ai-chatbots-in-2023-capabilities-use-cases-and-limitations-a3bfa31321aa>
5. Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, 114, 106553. <https://doi.org/10.1016/j.chb.2020.106553>
6. Koh, J. (2023). "Date me date me": AI chatbot interactions as a resource for the online construction of masculinity. *Discourse, Context & Media*, 52, 100681. <https://doi.org/10.1016/j.dcm.2023.100681>
7. Ma, Y., Liu, J., Yi, F., Cheng, Q., Huang, Y., Lu, W., & Liu, X. (2023). *AI vs. Human—Differentiation Analysis of Scientific Content Generation* (arXiv:2301.10416). arXiv. <http://arxiv.org/abs/2301.10416>
8. Theocharopoulos, P. C., Anagnostou, P., Tsoukala, A., Georgakopoulos, S. V., Tasoulis, S. K., & Plagianakos, V. P. (2023). *Detection of Fake Generated Scientific Abstracts* (arXiv:2304.06148). arXiv. <https://doi.org/10.48550/arXiv.2304.06148>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

