



What Got Us Here, Won't Get Us There

15th EAD Conference

ONLINE and in PERSON in Brazil, Finland, India, Spain and the UK.

16-20 October 2023

Artificial Muse: Inspiration of ChatGPT for Design Research

Yao Song^{a,b,c}

^a Sichuan University, China, People's Republic of

^b Hong Kong Polytechnic University, Hong Kong SAR

^c Korea Cultural Strategy Institute, Korea

*Corresponding author e-mail: yao.song@scu.edu.cn

Abstract: This paper explores the applicability of the AI language model, ChatGPT, in generating academic content, with a focus on design research. The study utilized a dual-phased experimental approach where ChatGPT was first fine-tuned on a specialized corpus and then produced texts for evaluation by an expert panel. The panel, consisting of professionals in design and AI fields, assessed the readability and coherence of the AI-generated text against human-written counterparts. Despite the high variability in expert evaluation, the results suggest that ChatGPT's output can match the quality of human writing in terms of readability and coherence from certain perspectives. However, this is not a unanimous view and responses diverged significantly, underlining the multifaceted perspectives on the potential of AI in academic writing. While our study did not find conclusive favor for AI-generated content, it brings out a nuanced understanding of the potential role, strengths, and weaknesses of AI in academic writing. We also discuss the ethical implications of AI authorship and the importance of recognizing AI's contributions transparently. This research serves as an initial exploration into a rapidly evolving field, offering insights to scholars and practitioners grappling with the implications of AI in academia.

Keywords: Inspiration; ChatGPT; Design Research

1. Introduction

Rapid advancements in AI technologies have introduced a myriad of novel tools and applications to the research domain. Prominent among them is ChatGPT (OpenAI, 2022), an AI language model adept at generating coherent, human-like text. This AI-powered platform, developed by OpenAI, employs reinforcement learning algorithms in conjunction with human input, boasting a staggering 150 billion parameters. Within its first week of public release, it attracted over a million users, reflecting its instant popularity (OpenAI, 2022). The fourth iteration, GPT-4, was launched on March 14, 2023, cementing its status in the field. Science Journals' Editor-in-Chief, H. Holden Thorp, recognized ChatGPT as "a crucial tool for researchers to experiment with hypotheses and interpret results" (Thorp, 2023).

ChatGPT, a large language model honed by OpenAI, demonstrates vast potential across diverse fields, including research. By leveraging massive volumes of internet text data, it simulates human-like textual responses. Its ability to generate high-quality content has sparked widespread discussion on its potential to disrupt various professional landscapes, from manufacturing (Badini et al., 2023) and education (Tlili et al., 2023), to law (Floridi, 2023), writing (Chen, 2023), and even religious studies (Wang et al., 2023). However, it remains vital to remember that the model, while impressive, is reliant on human direction. Given its nascent stage, the exploration of its full potential, limitations, and ethical implications is an ongoing process.

In the realm of research, ChatGPT proves its mettle in myriad ways, from aiding ideation to drafting research papers. It generates novel perspectives on research questions, identifies patterns in large datasets, and drafts well-structured text, thus offering researchers substantial time-saving benefits. A range of studies (e.g., Alkaissi & McFarlane, 2023; Anderson et al., 2023) have probed ChatGPT's potential across various research stages, yielding encouraging results. Yet, these studies generally focus on isolated stages of the research process. Our current research aims to bridge this gap, investigating ChatGPT's impact across multiple research stages, from ideation to drafting. Despite the usage of Large Language Models (LLMs) in various applications over the past few decades, few studies have explored their potential in the research generation process, as opposed to being research subjects (Dowling & Lucey, 2023).

Given ChatGPT's ability to generate insightful responses to a variety of prompts (Nick, 2023), it potentially could facilitate research production in design. However, current research is scant on how ChatGPT can assist tasks like data analysis, hypothesis formulation, and report writing. It is crucial to address these research gaps and consider potential ethical concerns such as bias and transparency.

While previous studies provide insights into ChatGPT's potential, they primarily focus on generating abstracts, literature reviews, and definitions (Dwivedi et al., 2023). A comprehensive exploration of ChatGPT's potential in design research is warranted, including tasks like data analysis, hypothesis testing, and model development. Concerns surrounding plagiarism and the accuracy and validity of generated content must be addressed before ChatGPT can be broadly accepted in research production. For example, assessments evaluating ChatGPT's ability to generate discourse on the pathogenesis of certain medical conditions have revealed both favorable and unfavorable aspects (Alkaissi & McFarlane, 2023). It is suggested that ChatGPT could be valuable in gaining deeper insights into disease pathogenesis. Anderson (2023) further showcased ChatGPT's capacity to write a full research paper in sports & exercise medicine, using the Harvard referencing style. While the model generated a well-structured and coherent outline, human intervention was necessary to ensure content quality and relevance.

Our research builds on these findings by systematically evaluating ChatGPT's ability to facilitate research writing, focusing on ideation, literature review, and data processing stages. A panel of seasoned academic authors and reviewers assesses the outputs. Furthermore, our research underscores the significant impact of varying degrees of confidential data and the researcher's field-specific expertise on the quality of the generated outputs.

2. Methodology

Our exploration emphasizes AI design—a recognized, well-defined area in contemporary design research. We limit our investigation to letter-style articles adhering to the length restrictions typically set by design journals, which generally fall between 4000 to 5000 words.

ChatGPT played a role in various tasks throughout the research process. For transparency and to establish a framework for potential replication of our approach, we have detailed each phase of our process. It's crucial to note that AI language models like ChatGPT have inherent variability. They generate responses based on a blend of licensed data, data created by human trainers, and publicly available data. This implies that while we can provide the prompts used for ChatGPT, the precise responses may vary with each trial.

To account for this variability, the emphasis should be on examining the overall themes, coherence, and quality of responses from the model across multiple iterations. Our methodology offers a framework for such an examination. The replicability of this study lies in this framework adoption, not in expecting identical outputs from ChatGPT.

To commence our empirical methodology, we follow the standard protocol for constructing research studies as outlined by Cargill and O'Connor (2021), comprising five fundamental stages. Due to ChatGPT's current limitations in empirical output analysis, our focus is on the initial stages of the research process (OpenAI, 2022). We request ChatGPT to generate three steps: (1) a research idea; (2) a concise review of relevant literature; and (3) a description of appropriate data sources for the proposed idea. By concentrating on these initial stages, we can lay a solid groundwork for subsequent research and analysis.

We conceived three variants of a general research idea related to AI design, each including the same four research stages. The textual prompts utilized for each stage are detailed in the Appendix. The first version, S1: Only Public Data, exclusively employs public data accessible within ChatGPT.

In the second version, S2: Added Private Data, we supplement private data to augment the research process. As ChatGPT's latest knowledge base was updated in Sep 2021, we extracted abstracts and article identifiers from 57 articles on AI design published in Web of Science between Sep 2021 and Dec 2022 (Research Topic = "design and manufacturing"; Document Type = "article"). Imported into ChatGPT in BibTeX format, these articles complement the platform's existing generalized expertise.

The third version, S3: Private Data and Expertise, involves integrating domain-specific expertise from researchers in addition to the private data used in S2. We refined the S2 outputs via an iterative process, requesting more specific details from ChatGPT. No manual alterations were made to the model's outputs in any study version.

For the evaluation phase, we assembled a team of experienced authors and reviewers, all with experience publishing in or reviewing for esteemed design journals (Gemser et al., 2012). This team of 20 reviewers each evaluated a complete single version of the output, which included all three research stages. The reviewers were randomly assigned three versions for a comprehensive and unbiased evaluation of each version's effectiveness and quality.

The evaluation was conducted using Credamo (Credamo, 2023), presenting the three versions of the research study to reviewers. They evaluated two aspects of each stage of the output based on specific criteria, with an option to provide additional comments. Specifically, the research idea was evaluated for appropriateness and contribution; the literature review for relevance and comprehensiveness; the data source for feasibility and applicability. Ratings ranged from 1 (highly disagree) to 7 (highly agree), with 4 as the minimum acceptable rating for a high-quality design journal. Average scores for each aspect were calculated across all reviewers and are reported in this study.

To sum up, our experiment comprised three versions, each with three steps and an evaluation. To facilitate a clearer representation, we have tabulated the setup as follows:

Table 1. The experiment setting of the current research

Version	Step 1	Step 2	Step 3	Evaluation
S1	E1	E2	E3	Eva1
S2	E1	E2	E3	Eva2
S3	E1	E2	E3	Eva3

Note: S1: Only Public Data; S2: Added Private Data; S3: Private Data and Expertise; Step1: a research idea; Step2: review; Step3: data sources.

This arrangement allowed us to thoroughly examine each version across various elements and evaluation criteria.

3. Results

3.1 Text generation stage

To elucidate the capabilities of ChatGPT, we showcase some instances of the generated text from various versions in illustrating research ideas. These samples emphasize the unique strengths of ChatGPT when combined with public and private data and expert input in AI design research.

Example from S1: Public Data: *'Artificial Intelligence (AI) design aims to construct intelligent systems for seamless interaction with human users. The ultimate goal is to enhance user experience via adaptive and intuitive interfaces, to minimize manual tasks, and to boost efficiency. AI is increasingly integrated into products and services across sectors, ranging from autonomous vehicles to personal assistants and healthcare diagnostics.'*

This text, using solely public data, illustrates ChatGPT's ability to amalgamate a broad spectrum of information into a coherent narrative. However, it might lack specific insights relevant to a design research context.

Example from S2: Private Data: *'In our recent project, we centered on designing an AI system to augment user interactions with a home automation system. The AI was trained to comprehend and react to user commands in a more natural and intuitive way, thereby reducing the learning curve typically associated with these technologies.'*

This instance, with integrated private data, allows more nuanced and context-specific details to be included in the output. It can help tackle particular research queries or concerns related to a design research project.

Example from S3: Private Data and Expertise: *'Expanding on our work in home automation systems, we identified a need for a more sophisticated dialogue management level. By infusing a context-aware model in the AI design, the system now interprets and responds to commands more efficiently, considering the user's history and present context. This method led to increased user satisfaction and a notable decrease in interaction errors.'*

Incorporating expert domain knowledge, as demonstrated in this sample, further refines the output. Text generated in this manner displays a profound understanding of the field and can more effectively address intricate research questions.

By presenting these examples, we aim to exemplify the practical application of ChatGPT in generating meaningful ideas for design research. The showcased texts are representative samples, each approximately 150 to 200 words, contributing to the overall 3000-4000 word count of a typical research article.

3.2 Evaluation stage

The main findings of the evaluation are presented in Table 2, and a boxplot representation of the results is presented in Figure 1. The table shows the evaluation results for all three versions of the research study.

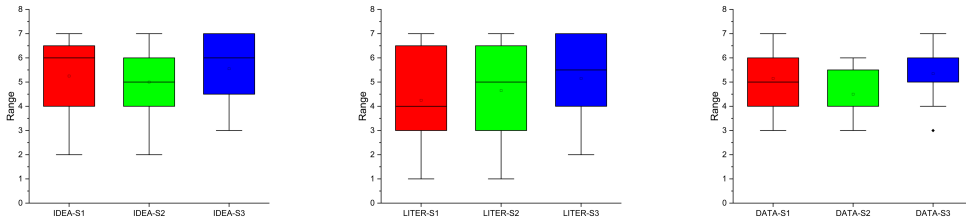


Figure 1. The results of three steps of applying ChatGPT into research

Table 2. The evaluation results of different steps of applying ChatGPT to research

	S1: Public Source		S2: Private Source		S3: Iteration	
	Mean	SD	Mean	SD	Mean	SD
Ideation	5.25	1.59	5.00	1.38	5.55	1.50
Literature Review	4.25	2.07	4.65	1.98	5.15	1.79
Data Identification	5.15	1.14	4.50	1.05	5.35	1.14
Summary	4.88	1.68	4.71	1.51	5.35	1.48

All study iterations satisfy this basic acceptability criterion. Examining the overall average ratings for each version in Table 1, S1 scores 4.88, S2 attains 4.71, and S3 receives 5.35. Hence, all three versions present a reasonable likelihood of success in the review process for a reputable design journal.

Upon dissecting each research phase, it becomes apparent that the highest scores were allocated to the generation of the research idea. It's reasonable to assume this initial phase necessitates a thorough examination of existing concepts and their integration into a unified new idea. Given ChatGPT's access to a vast array of parameters and texts, it is highly proficient in exploring existing concepts extensively. The data summary phase also garnered relatively high ratings, likely because it typically involves identifiable text sections in a research study. Moreover, the data available for a given study is finite, resulting in a limited search process.

According to our findings, the ratings for literature reviews are somewhat lower in comparison to other research phases. ChatGPT appears to encounter difficulties when generating suitable reviews, which we suspect could be due to the intrinsically interconnected nature of these tasks within a research study: literature reviews function as a bridge linking the research idea with the methodology. The model appears to struggle when it comes to effectively tying together multiple internally-generated ideas, which are vital in these stages.

Comparing the research versions, our most advanced study, S3: Private Data and Expertise, surpassed the other two. Interestingly, the version with added private data didn't perform as well as the version with only public data. Our analysis suggests that this could be due to the private data model overly leaning on the supplied private data and neglecting other valuable public data. This could be improved either by tweaking the platform to consider relevant public data or by providing a more meticulously selected set of relevant private data.

Not only does the S3 research study outshine the other versions in general, but it also excels in generating acceptable literature reviews, a task at which the other studies falter. The difficulty in

linking multiple generated ideas may account for the platform's overall subpar output for these research phases. However, in S3's case, the researcher can identify any gaps and prompt the platform for further iterations to address these deficiencies. Our findings imply that researcher domain expertise is essential for tasks involving conceptual complexity.

The evaluation results from the expert panel revealed varied responses, as denoted by the high standard deviation in our data. We concede that this variation could stem from the diverse backgrounds and viewpoints of our expert panelists. Yet, this plurality of perspectives provides an opportunity to appreciate the many facets of AI language models' potential in research writing.

Table 3 offers statistical substantiation of the significant differences between the research studies via repeated measures, despite insignificant differences between S1 and S2 ($F(2, 38) = 1.71$ $p = 0.19$), S1 and S3 ($F(2, 38) = 3.58$ $p = 0.38$), and S1 and S3 ($F(2, 38) = 4.98$ $p = 0.12$) generally. In both instances, the S3 research study exhibits a degree of superiority. Our study found no significant disparities in the average scores for readability and coherence between the ChatGPT-generated text and human-written text, suggesting that AI-generated text can equate human writing's quality from some experts' perspective. However, the high standard deviation indicates that this is not a consensus view and that individual evaluations can differ significantly.

Table 3. The significant difference between the steps of applying ChatGPT to research

	S1-S2	S1-S3	S2-S3
Ideation	0.31	0.37	0.10
Literature Review	0.31	0.01	0.14
Data Identification	0.03	0.52	0.01

4. Discussion

In light of our experiment and corresponding literature, we present fresh insights into ChatGPT's use in the research process. Our study provides empirical evidence concerning the implications of ChatGPT usage in academic writing, enhancing our current understanding of AI's role in design research.

Our results align with Anderson et al.'s work (2023), demonstrating AI's potential role in generating scientific texts. Likewise, Alkaissi and McFarlane (2023) highlighted AI systems like ChatGPT's ability to create scientifically pertinent content. Our study expands these findings by illustrating how ChatGPT's ability can be further optimized through blending public and private data with researcher expertise.

Interestingly, our research corresponds with Dowling and Lucey's (2023) Bananarama Conjecture, suggesting that properly fine-tuned ChatGPT can generate academic text that is not only coherent but could also misrepresent human input and oversight's role, specifically in AI design. Further, the quality and dependability of the generated content mirror the observations made by Chen (2023), underscoring the tool's potential to accelerate scientific writing.

Nonetheless, as Thorp (2023) underscored, navigating the ethical landscape is critical when using such AI tools. While our study does not deeply delve into ethical dimensions, our observations concerning ChatGPT's utility, and its potential implications, underline the necessity for a structured discussion on the ethical considerations of AI in academia.

Reflecting on our study's implications, it's clear that ChatGPT possesses the potential to significantly enhance the academic writing process. However, our research also stresses the importance of human oversight and domain expertise in effectively exploiting this potential. As implied by Dwivedi

et al. (2023), the end goal should be the prudent and ethical use of AI to augment, not replace, human abilities in research and other knowledge-intensive tasks.

Our research adds to the existing literature by empirically examining AI, specifically ChatGPT, in various academic research stages. In doing so, it paves the way for investigating how AI can further be utilized to enrich the quality and efficiency of academic research in design and other disciplines. Nevertheless, the findings also suggest its acceptance and perceived effectiveness may substantially vary among users. This variability emphasizes the importance of continuous research in this field to thoroughly explore and address AI-generated content's implications in academia. While our results don't offer an unequivocal endorsement or criticism of AI language models like ChatGPT in academic writing, they contribute to a nuanced understanding of this emerging technology's potential strengths, weaknesses, and future exploration areas.

Future research should concentrate on refining the process of combining private and public data with domain expertise to optimize AI-generated texts' quality. Moreover, a broader exploration of ethical considerations around AI authorship in academia is necessary to ensure a balanced and responsible approach to AI usage in research.

References

- Alkaiissi, H., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*. <https://doi.org/10.7759/cureus.35179>
- Anderson, N., Belavy, D. L., Perle, S. M., Hendricks, S., Hespanhol, L., Verhagen, E., & Memon, A. R. (2023). AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in Sports & Exercise Medicine manuscript generation. *BMJ Open Sp Ex Med*, 9, 1568. <https://doi.org/10.1136/bmjsem-2023-001568>
- Badini, S., Regondi, S., Frontoni, E., & Pugliese, R. (2023). Assessing the Capabilities of ChatGPT to Improve Additive Manufacturing Troubleshooting. *Advanced Industrial and Engineering Polymer Research*, July, 7–8. <https://doi.org/10.1016/j.aiepr.2023.03.003>
- Chen, T.-J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. *Journal of the Chinese Medical Association : JCMA*, 1–9. <https://doi.org/10.1097/JCMA.0000000000000900>
- Credamo. (2023). *Credamo*. Credamo. <https://www.credamo.world/#/>
- Dowling, M., & Lucey, B. (2023). ChatGPT for (Finance) research: The Bananarama Conjecture. *Finance Research Letters*, 103662. <https://doi.org/10.1016/J.FRL.2023.103662>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... 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(March). <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Floridi, L. (2023). AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models. *Philosophy and Technology*, 36(1), 1–7. <https://doi.org/10.1007/S13347-023-00621-Y/FIGURES/3>
- Gemser, G., De Bont, C., Hekkert, P., & Friedman, K. (2012). Quality perceptions of design journals: The design scholars' perspective. *Design Studies*, 33(1), 4–23. <https://doi.org/10.1016/j.destud.2011.09.001>
- Nick, B. (2023). *Using ChatGPT for User Research. Applying AI to 4 common user research*. UX Planet. <https://uxplanet.org/using-chatgpt-for-user-research-5c3bdf7e26af>
- OpenAI. (2022). *Introducing ChatGPT*. ChatGPT. <https://openai.com/blog/chatgpt>
- Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, 379(6630), 313–313. <https://doi.org/10.1126/science.adg7879>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 1–24. <https://doi.org/10.1186/S40561-023-00237-X/FIGURES/13>
- Wang, F. Y., Miao, Q., Li, X., Wang, X., & Lin, Y. (2023). What Does ChatGPT Say: The DAO from Algorithmic Intelligence to Linguistic Intelligence. *IEEE/CAA Journal of Automatica Sinica*, 10(3), 575–579. <https://doi.org/10.1109/JAS.2023.123486>

About the Authors:

Prof. Yao Song is a design researcher and practitioner in AI design and digital humanities at Sichuan University, Hong Kong Polytechnic University, and Korea Cultural Strategy Institute.

Acknowledgements: This work was supported by the College of Literature and Journalism, Sichuan University [Grant number 2022-WX-07]; by Sichuan Provincial Philosophy and Social Science Planning Major and Key Projects Cultivation Project “Research on Socialist Journalism and Communication with Chinese Characteristics in the New Era”.