10^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2024, Philadelphia, USA

Evaluating the Impact of Generative AI on Human Creativity: A Randomized Experiment

Keywords: creativity, large language models, generative artificial intelligence, Human-AI Collaboration, experiments

Extended Abstract

Since the release of advanced deep learning models like GPT, the landscape of human-AI co-creation has significantly expanded, fostering innovation across various domains, such as poetry and music [1]. The state-of-the-art Generative AI (GAI) systems have reached a level of proficiency where they not only parallel human creativity in tests [2], but have also been shown to enhance the creative capabilities of knowledge workers [3], allowing the production of a multitude of ideas in minutes. Despite these advancements, there are concerns about the potential long-term effects of sustained GAI tool usage, such as ChatGPT, on inherent human creativity. While existing research has explored the comparative creative abilities of humans and AI, and how humans can leverage AI for improved performance in creative tasks, an unaddressed question remains [4].

> Does the use of Generative AI diminish our innate creativity, similar to a steroid effect, or does it instead bolster it, acting as a coach that enhances our creativity?

In this ongoing research, we are conducting a randomized controlled experiment to examine the impact of exposure to various types of LLM responses on human creativity. Our study specifically targets divergent thinking, a key component of creativity, utilizing the Alternate Uses Task (AUT)—the most commonly used method to measure this aspect [5].

Method We will recruit participants from Mechanical Turk and Prolific. Participants in this task are asked to come up with novel and creative uses for common everyday objects within a two-minute timeframe per item. They will be told *"The goal is to come up with creative ideas, which are ideas that strike people as clever, unusual, interesting, uncommon, humorous, innovative, or different. Your ideas don't have to be practical or realistic; they can be silly or strange, even, so long as they are CREATIVE uses rather than ordinary uses..."*. Our experimental design (shown in Figure 1) includes two main phases: a practice phase and a test phase. During the practice phase, participants are introduced to objects (such as a tire, backpack, or lightbulb) are then randomly assigned to receive three different types of LLM response:

- *No LLM Response:* Control group.
- *List of Ideas:* GPT-4 was asked to generate a list of alternate uses for the given object. Seven uses are shown to the participant.
- *List of Strategies:* GPT-4 with a specialized system prompt (shown in Figure 2) was asked to generate seven helpful strategies based on the SCAMPER [6] technique to help users come up with ideas. Each technique includes an example use of the object.

Figure 3 shows the LLM responses shown to participants for both LLM conditions. Following the practice phase, participants engage in a brief interlude, playing a game of Snake for one minute to act as a mental reset before proceeding to the test phase. In this subsequent phase, individuals are tasked with the AUT for a new object selected at random, this time without LLM support, to measure the effect of prior exposure to LLM responses. We collect participants' self-assessments of their creativity levels and their attitudes towards AI both before and after the experiment.

Analysis We will conduct a detailed comparison of the responses generated during the test phase across the three experimental conditions. The evaluation of these responses will be based on two critical dimensions of creativity: originality, which assesses the uniqueness of the ideas presented, and fluency, which quantifies the volume of ideas generated. To objectively measure originality, we will follow a long tradition of scoring responses to the AUT computationally [7, 8]. Specifically, we will utilize a fine-tuned GPT-3 classifier [9] to measure the originality of the ideas generated during AUT. This classifier has shown a strong overall correlation (r=0.81) with human judgments of originality, providing a reliable means to evaluate the novelty of participant responses.

Discussion AI can now generate, or "create," artifacts that can be useful and/or enjoyable to people. This prompts an essential question: How does this AI advancement impact people's ability to be creative? The findings from our study will shed light on this critical question. For example, it might show how different forms of interaction with LLMs, such as receiving strategies rather than direct ideas (the status quo), can enhance human creativity. This knowledge will be instrumental in guiding the design of human-AI co-creative systems. The implications of our research extend far and wide, potentially affecting cultural norms [10] and practices related to creativity and innovation.

Our study has limitations. The focus on divergent thinking through the Alternate Uses Task, while significant, represents just a fraction of the broad spectrum of creativity. Creativity is inherently complex and challenging to accurately quantify. Despite these constraints, our investigation will be one of the initial attempts to understand the effects of using LLMs on human creativity. In presenting this work, we will discuss results from this experiment, along with additional studies that we have planned on other aspects of creativity and in real-world settings (such as classrooms, as opposed to with online crowdworkers).

References

[1] Ziv Epstein, Aaron Hertzmann, Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R Frank, Matthew Groh, Laura Herman, Neil Leach, et al. Art and the science of generative ai. *Science*, 380(6650):1110–1111, 2023.

[2] Mika Koivisto and Simone Grassini. Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports*, 13(1):13601, 2023.

[3] Nan Jia, Xueming Luo, Zheng Fang, and Chengcheng Liao. When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, (ja), 2023.

[4] Jake M Hofman, Daniel G Goldstein, and David M Rothschild. Steroids, sneakers, coach: The spectrum of human-ai relationships. *Available at SSRN 4578180*, 2023.

[5] Joy P Guilford. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1):3–14, 1967.

[6] Olivier Serrat and Olivier Serrat. The scamper technique. *Knowledge solutions: tools, methods, and approaches to drive organizational performance*, pages 311–314, 2017.

[7] Roger E Beaty and Dan R Johnson. Automating creativity assessment with semdis: An open platform for computing semantic distance. *Behavior research methods*, 53(2):757–780, 2021.

[8] Roger E Beaty, Dan R Johnson, Daniel C Zeitlen, and Boris Forthmann. Semantic distance and the alternate uses task: Recommendations for reliable automated assessment of originality. *Creativity Research Journal*, 34(3):245–260, 2022.

[9] Peter Organisciak, Selcuk Acar, Denis Dumas, and Kelly Berthiaume. Beyond semantic distance: automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity*, page 101356, 2023.

[10] Joshua Ashkinaze, Julia Mendelsohn, Li Qiwei, Ceren Budak, and Eric Gilbert. How ai ideas affect the creativity, diversity, and evolution of human ideas: Evidence from a large, dynamic experiment. *arXiv preprint arXiv:2401.13481*, 2024.

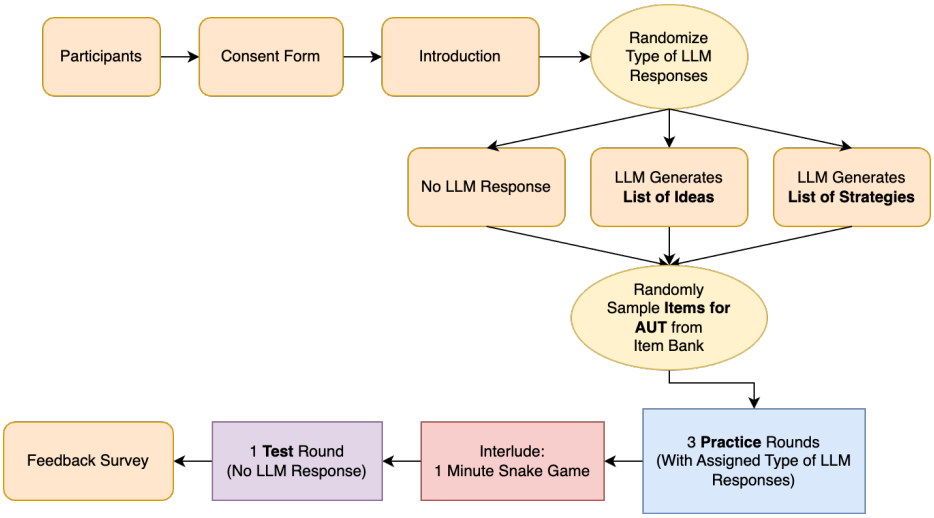


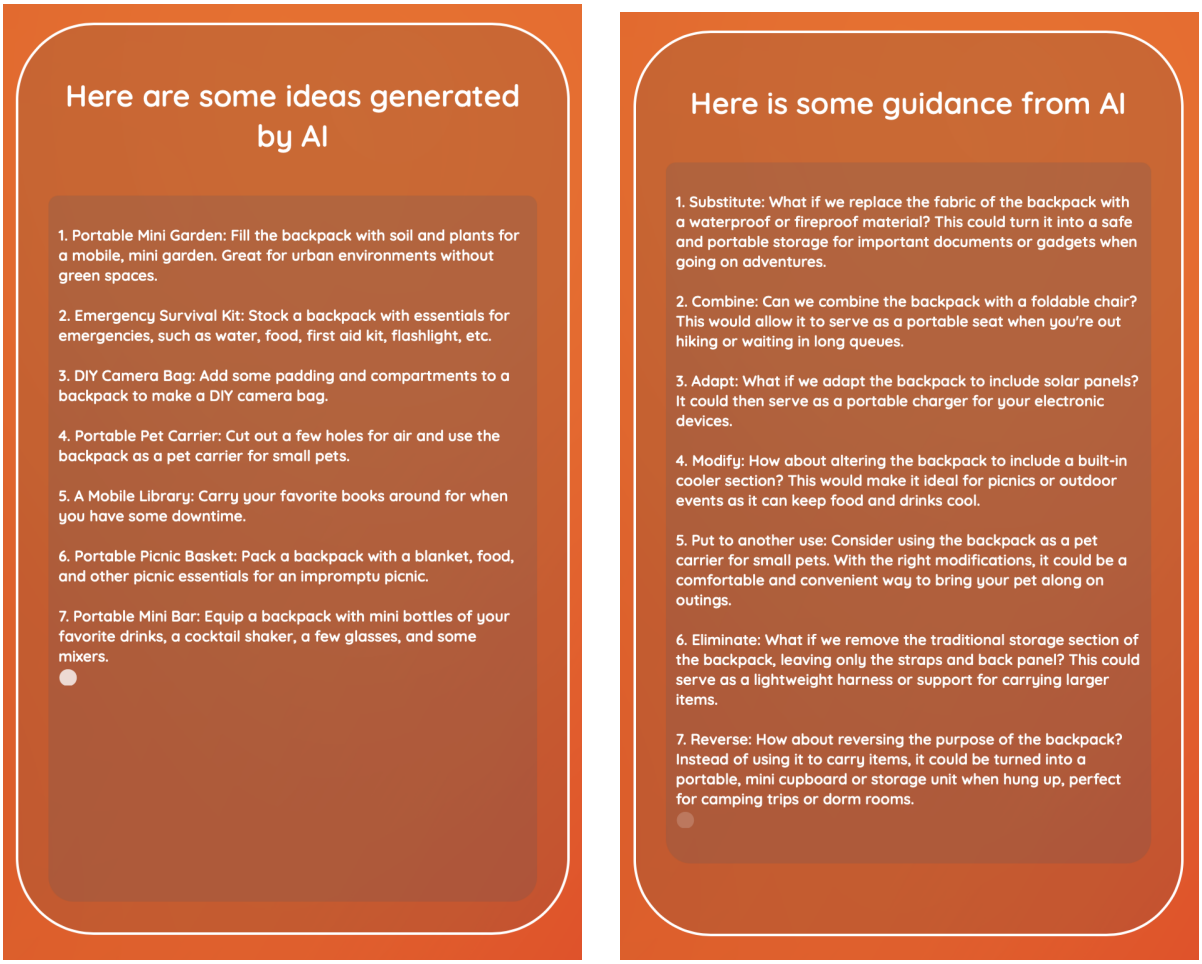
Figure 1: Schematic of the experiment design.

You are an AI assistant tasked with the specific goal of teaching users to be more creative by applying the SCAMPER technique. SCAMPER is a powerful tool for creative thinking that involves exploring various ways to modify an object or idea to generate new possibilities. Your interactions should guide users through the SCAMPER process, which includes the following steps:

- **Substitute:** Encourage users to think about what elements of the object could be replaced with something else to improve it or give it a new use.
- **Combine:** Prompt users to consider how they might combine the object with other items or ideas to create something new.
- **Adapt:** Ask users how the object could be adapted or modified to serve another purpose, or how it might be inspired by something else.
- **Modify:** Challenge users to think about how altering the shape, size, color, or any other attribute of the object could create new uses or improvements.
- **Put to another use:** Encourage users to brainstorm completely different uses for the object, beyond its original purpose.
- **Eliminate:** Guide users to consider what might happen if parts of the object were removed or simplified. How does this change its potential uses?
- **Reverse:** Ask users to think about what would happen if they reversed the object's purpose or the way it's used. Can it function differently or serve a new role?

In your guidance, encourage users to iterate through these steps, not necessarily in order, to explore the full range of creative possibilities. Provide examples to illustrate how each step of the SCAMPER technique can lead to innovative solutions and new perspectives on everyday objects. Don't mention SCAMPER explicitly in your response. Output a list that the user can follow.

Figure 2: The system prompt (pre-prompt) for GPT-4 for generating *list of strategies* (not visible to participants).



(a) LLM generates *List of Ideas*.

(b) LLM generates *List of Strategies*.

Figure 3: LLM responses for the object "backpack" from the practice phase in different LLM conditions. Participants are given 2 minutes per object.