Journal of Creativity 33 (2023) 100066

Contents lists available at ScienceDirect

Journal of Creativity

journal homepage: www.elsevier.com/locate/yjoc

Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity☆

Jennifer Haase [a,b,*], Paul H.P. Hanel [c]

[a] Department of Computer Science, Humboldt University Berlin, Unter den Linden 6, Berlin 10099, Germany
[b] Weizenbaum Institute, Berlin, Germany
[c] Department of Psychology, University of Essex, 4.704, Colchester Campus, United Kingdom

ARTICLE INFO

Keywords:
Creativity
Originality
AI
Generative artificial intelligence

ABSTRACT

A widespread view is that Artificial Intelligence cannot be creative. We tested this assumption by comparing human-generated ideas with those generated by six Generative Artificial Intelligence (GAI) chatbots: alpa.ai, Copy.ai, ChatGPT (versions 3 and 4), Studio.ai, and YouChat. Humans and a specifically trained AI independently assessed the quality and quantity of ideas. We found no qualitative difference between AI and human-generated creativity, although there are differences in how ideas are generated. Interestingly, 9.4 % of humans were more creative than the most creative GAI, GPT-4. Our findings suggest that GAIs are valuable assistants in the creative process. Continued research and development of GAI in creative tasks is crucial to fully understand this technology's potential benefits and drawbacks in shaping the future of creativity. Finally, we discuss the question of whether GAIs are capable of being "truly" creative.

Introduction

While Artificial Intelligence has outperformed humans in numerous domains, including chess or GO (Miller, 2019), there is a prevailing sentiment that creativity remains one of the few arenas where humans still hold an advantage (Holford, 2019; Miller, 2019). However, recent generative artificial intelligence (GAI) developers have argued that their software is also creative (e.g. OpenAI advertises "GPT-4 is more creative and collaborative as ever before", https://openai.com/product, as of August 2023). We put this claim to the test by comparing whether humans are (still) more creative than six GAIs and let both humans and AI be the judge of this.

Artificial intelligence

The increasing use of GAI in daily life is changing how we work, communicate, and create (Frosio, 2023). The GAI is an innovative approach that allows machines to learn from previously collected data and adapt to new situations. This key technology is becoming increasingly important for organizations. It helps with automated decision-making processes, detects patterns in large data sets, and improves people's overall efficiency (Noy & Zhang, 2023). An increasing number of tasks are automated; therefore, workers are left with more complex and potentially more creative tasks that require human ingenuity and problem-solving skills (Pistrui, 2018).

As there is growing potential for GAI to perform complex tasks, there is also increasing interest in exploring how GAI can be used to support and enhance human creativity (Cope, 2005; Reddy, 2022). Recent research shows that working together with ChatGPT on complex creative tasks improves the individual's creative performance as well as their perceived self-efficacy beliefs (Urban et al., 2023). However, there is debate on whether AI can be genuinely creative or simply recombine existing knowledge to *appear* in new ways (Kirkpatrick, 2023; White, 2023). The discourse of GAI's potential usage and impact tends to reduce the dialog to a simple is or is not creative. However, scientific literature draws a much more detailed picture of creativity, as creative thinking and creative problem-solving are much more diverse. Typical examples include problem formulation, idea generation, idea selection, and potential idea implementation (Botella et al., 2019; Williams et al., 2016). GAI can generate vast amounts of new textual (e.g., ChatGPT, Dale, 2021) and visual (e.g., Dall-E, Marcus, Davis, & Aaronson, 2022) output based on written prompts through combining existing data in a

☆ Data-access: https://osf.io/9fctd/?view_only=6c8f02c6972b49319c12f87cfb3f76db

* Corresponding author.
E-mail address: Jennifer.haase@hu-berlin.de (J. Haase).

https://doi.org/10.1016/j.yjoc.2023.100066
Received 10 July 2023; Received in revised form 29 August 2023; Accepted 26 September 2023
Available online 9 October 2023

new way. The human counterpart to this is free-associative thinking, the cornerstone of creative processes (Steele et al., 2018).

The rapid emergence of new technologies generates a wealth of information that was not previously available. Unlike previous "smart" tools, which can aggregate existing knowledge, these new technologies can develop novel insights and solutions. This opens up possibilities for supporting human tasks in various domains, including healthcare, education, and entertainment (Seidel & Berente, 2020). However, it also raises important questions regarding the role of these technologies in facilitating human performance and how they might be designed to enhance and support creativity.

In recent years, there have been numerous demonstrations of AI's capacity for creativity. For instance, algorithms can compose music (Civit et al., 2022), which can be regarded as creative, following the general definition of "creating something new and useful" (Runco & Jaeger, 2012). Google's AlphaGo program, which defeated the human world champion in the ancient Chinese board game Go in 2016, is another remarkable example of AI's creative potential. AlphaGo mimicked human game players and generated new and sophisticated strategies to win the game (Miller, 2019). Additionally, DiPaola (2016) discussed the development of an AI system that emulates the creativity of a portrait painter, providing a new tool for artists and designers to explore novel creative paths and generate ideas that seem unlikely without AI assistance.

GAIs are becoming more competent and more capable of replicating information from the web, including a range of services for complex digital tasks such as coding, template creation, and business administration. However, reported inaccuracies in AI systems question their usage as a reliable knowledge-creation tool and fuel a debate on the precise application possibilities and limits of these systems (Dale, 2021; Else, 2023; "The AI Writing on the Wall,", 2023). As for ChatGPT, the language model is trained on massive amounts of text data sourced from the internet, allowing it to learn patterns and relations between words and phrases in a language. The produced text is unreliable when facts are involved but more valuable when fiction and accidental combinations are required (as in fiction, poetry, and game dialogues, Dale, 2021). A vast knowledge base, combined with a few factual specifications, can support creative thinking in humans. The currently advanced version of GPT-4 is advertised as leading to more comprehensive, correct, and creative results, than the prior version GPT-3 (OpenAI, 2023a).

Creativity

Creativity is considered a skill only humans possess (Kirkpatrick, 2023). Plucker provided a widely accepted definition of creativity as "the interaction of aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context" (Plucker et al., 2004). Creativity can be defined as creating and enacting something new, unique, and original (Sawyer, 2012). This pragmatic definition of creativity as "something new and useful" (Runco & Jaeger, 2012) is widely applied. However, these definitions are not linked to anything innately human, such as experience, emotions, or moral understanding. Thus, for machines, robots, and AI systems to be recognized as creative, they do not have to replicate the attitudes, behaviors, or actions of creative humans; rather, they need to replicate the cognitive process and the outcome to achieve something perceivable as "new and useful" (Cropley et al., 2023). Thus, whether GAI is creative is not the right question, as it is about the perceptually creative output. What is somewhat worth asking is the significance of their creative output.

Creativity can be distinguished from everyday level to higher, eminent levels of creative output with far-reaching consequences for a domain or a social area (Amabile, 2017; Richards, 2010). Whereas everyday creativity is mostly fast-paced, highly related to improvisation, and built into our everyday work and living, higher levels of creative achievement require significantly more time, specific knowledge, and often testing phases to determine whether a potential solution holds up (Benedek et al., 2020; Kaufman et al., 2016; Simonton, 2013). In our research we focus on everyday creativity. Since a chatbot only produces output in response to a written prompt, the creative work of a chatbot is dependent on human input. If the chatbot is used to solve a creative task to generate ideas on a specific topic, then the human needs to write a prompt that best represents the creative challenges core. The prompt therefore defines the creative problem.

Similarly, the further processing and potential implementation of the ideas generated by the GAI is also to be done or coordinated by humans. A chatbot can deliver a vast amount of generated output, from which users can and need to choose how to follow up. Thus, the GAI is inherently capable of "being creative", that is, able to generate ideas, but this does not resemble the full creative process observed with humans. Idea production can only be purposeful if the problem which precedes it is fully understood. If the creative problem or challenge is clear, criteria can be formed to recognize an idea as suitable for the problem (Lazar et al., 2022). This recognition of fit, meaningfulness, and situational novelty lies in individual human consideration (Cropley et al., 2023). Thus, the potential creative GAI has, at least at the moment, an assistance role, which can support a certain aspect of the holistic process: idea generation.

In this article, we ask to what extent GAI chatbots can produce new and useful output given an every-day creative task. Assuming its assistance role, these tools would require at least a medium performance to be considered a useful assistance for humans working on creative tasks. Thus, we test experimentally whether AI performance is comparable with human creative performance for everyday-creative tasks.

Method

Participants and chatbots

A power analysis revealed that we need at least 88 participants to detect a small-to-medium effect size of $d = 0.35$ with a power of 0.90. In total, 100 participants completed our study ($M_{age} = 41.00$, $SD = 12.25$, 50 women, 50 men), who were recruited through *Prolific Academic*. All participants provided informed consent at the beginning of the study. The study was approved by the Humboldt-University of Berlin. Participants were all native English speakers from the USA with full- or part-time work. We paid a prorated rate of US-$9 per hour. The average completion time was 17 min.

Initially, we selected five GAI chatbots: Alpa.ai, Copy.ai, ChatGPT version 3, Studio, and YouChat. We selected them based on their free usability and similar functions to ensure comparability. Alpa.ai is a system for training and deploying large-scale neural networks that have been made available as an open-source project. Its primary objective is to streamline the distributed training and deployment process of these networks, and it has been designed to do so with minimal code input required. A team of researchers created Alpa in the Sky Lab at UC Berkeley. We used the Chatbot function with maximum response length to collect the answers for the five prompts, respectively.

Copy.ai uses natural language processing (NLP) and machine learning (ML) algorithms to generate explicitly creative ideas. The tool is meant for content creation, such as social media posts, blogs, etc. It comes with a variety of features suited for specific content needs. Copy.ai has a chatbot function, which appeared rather limited in its output, so we used the "freestyle" template instead. This template can generate "more like this", which we used three times to generate more ideas.

ChatGPT is a language model based on the GPT-3 database developed by OpenAI, which can generate human-like responses to natural language inputs. It uses deep learning techniques to analyze and understand language patterns and can provide answers to a wide range of questions and prompts. After finishing our initial data collection and analyses, we added answers from the newer version, based on the GPT-4 database with more extensive training data, leading to more

comprehensive and (according to the developers' webpage) more creative output.

Studio is the *AI21 Studio,* a new developer platform developed by AI21 Labs based on their developed Large Language Model called Jurassic-1 and allows users to build their applications and services. We used *Playground* to interact freely, which comes closest to a chatbot tool.

YouChat is a messaging platform and AI-powered search assistant created by You.com. Users can leverage its capabilities to ask various questions, receive helpful explanations and recommendations, translate text across different languages, summarize written content, and perform other useful tasks.

Materials

Participants completed the Alternative Use Test for five prompts. They were instructed to write down as many ideas as possible for a *ball, fork, pants, tire*, and *toothbrush*, respectively. These objects are commonly used in creativity tests (Reiter-Palmon et al., 2019; Silvia et al., 2008; Vartanian et al., 2019) and can therefore be reliably assessed by the AI-rating tool we used (the AI had been trained on many prompts, with highest reliability for the five we used, Organisciak et al., 2023). Human participants were given three minutes for each object to write down as many ideas as possible. The order in which the prompts were presented was randomized. To get responses from the six GAI chatbots, we used the same prompt: "What can you do with [prompt]?". We used separate chat sessions for each prompt, so prior answers would not impact the following ones. For all chatbots, responses were limited to a certain length of answers, which we increased by asking "What else?" up to three times (for Copy.ai, we used the option "more like this"). In some instances, a chatbot would also respond with something like "I can't think of anything." This is similar to what some humans reported. These kinds of no-answers were excluded from the data set. In other cases, the chatbots would report unrelated answers (e.g., "I am not a big fan of the toothbrush. I think it is overrated."). This is again similar to human answers, and those were also excluded from the data. In one case, when asking for the use of pants, alpa.ai could not bring up any alternate uses.

Procedure

Data were collected in early February 2023. All procedures performed in this study involving human participants were in accordance with the 1975 Helsinki Declaration and the ethics guidelines from the institution of the lead author. Six human raters rated the responses from human participants and five of the six GAI chatbots (GPT-4 was released on the 14th of March), blind to the origin of the responses. The order of the prompts was randomized throughout the raters, and the list of ideas was randomized. The six human raters were instructed to follow the CAT method (Amabile, 1982), which comprised using the full range of the originality scale from 1 to 5. The first author provided detailed instructions to all raters on how to evaluate individual ideas, which included defining non-relevant responses. Clear definitions for both 'originality' and 'creativity' were presented and deliberated upon. For consistency, all raters began by assessing an initial set of 100 randomly selected responses to resolve any ambiguities in their evaluations before proceeding to rate the entirety of the ideas. Additionally to human raters, we assessed originality scores for all human-generated as well as all six GAI chatbot-generated answers by a trained large language model for assessing AUT prompts (Organisciak et al., 2023). This measure derives from an extensive dataset comprising human evaluations of AUT prompts sourced from earlier creativity research. Notably, these human ratings exhibit a strong correlation with AI-generated assessments, with $r_m = 0.80$. To ensure the integrity of the AI's ratings, we selected five prompts that demonstrated a correlation exceeding 0.80 between human and AI evaluations. These prompts were ball, fork, pants, tire, and toothbrush (cf. Organisciak et al., 2023).

Fluency scores were calculated for the AI and the six raters as the sum of ideas from each participant and the GAI chatbots. The sum of ideas varied slightly, as the raters differed in their assessment of non-relevant answers coded as no-answer.

Results

Six humans and a specifically trained AI (Organisciak et al., 2023) independently rated the originality of each response produced by humans and GAI, blind to the creator of the response. Since most humans and GAIs produced more than one response, we averaged across responses separately for each creator and prompt to obtain an originality and a fluency score for each human and each GAI, separately for each of the five prompts. After we completed our data analyses in early March 2023, a GAI advertised as "more creative" than its prior versions, GPT-4 was released. We included it in some follow-up analyses but not the main analysis, as explained below. The data and the R-code to reproduce the analyses can be found at https://osf.io/9fctd/?view_only=6c8f02c6972b49319c12f87cfb3f76db

Originality

To estimate the interrater reliability between the six human raters, we computed the intraclass correlations using the R-package irr (Gamer, Lemon, & Singh, 2021). Interrater reliability was excellent (Cicchetti, 1994): Intraclass correlations ranged from 0.85 to 0.94 for the five prompts, indicating that human raters mostly agreed on which answers were original (supplemental materials, Table S1). To test whether ratings from humans and the creativity scoring AI align, we averaged across all six human raters and correlated the score with the score from the AI. Correlation coefficients were very high, $rs = 0.78$ - 0.94, $ps < 0.0001$, indicating that also humans and AI mostly agree on which response can be considered original.

To test whether humans or the GAI chatbots were more creative, we ran two linear mixed effects models with random intercepts and random slopes for the five prompts using the R-package lme4 (Bates et al., 2015). The first model, which included human-rated responses as the dependent variable, found no mean difference between human and GAI-generated ideas, $B = -0.21$, $SE = 0.15$, $p = .218$. The second model, which included AI-rated responses as the dependent variable, also found no mean difference between human and GAI-generated ideas, $B = -0.18$, $SE = 0.13$, $p = .241$. These results were mostly replicated in between-subject t-tests (Table 1). Only human-rated responses for forks and AI-rated responses for toothbrush humans outperformed the GAI (Figs. 1 and S1).

Given the number of comparisons and unequal sample sizes (100 vs. 5), we decided to additionally compute the number of participants who received a higher originality score than the most original GAI. For pants, for example, human raters rated 42 humans as more original than the most original GAI, whereas the AI rated 52 humans as more original than the most original GAI. Across all prompts, 32.8 humans were more original than the most original GAI.

Finally, we compared the five GAIs. None of the GAI chatbots emerged as more original than the other four across all five prompts (Figs. 2, S2).

GPT-4

After we completed the analyses, including the five GAIs, GPT-4 was released in mid-March 2023. We had ChatGPT version 4 also complete the AUT. Its responses were only analyzed by the AI because the human raters would have likely known that the responses were not human, thus potentially biasing their ratings. Given the high correlations between human raters and the AI, we speculate that the findings would have been similar if humans had rated it. ChatGPT version 4 outperformed all five other GAIs, except for the prompt ball, for which it ranked second

Table 1
Descriptive and inferential statistics for originality.

	$Human_R$								AI_R							
Prompt	M_H	SD_H	M_{GAI}	SD_{GAI}	t	p	d	$H >$ GAI	M_H	SD_H	M_{GAI}	SD_{GAI}	t	p	d	$H >$ GAI
Pants	2.43	0.44	2.03	0.7	1.13	.339	0.88	42	2.4	0.46	2.3	0.34	0.6	.574	0.21	52
Ball	2.22	0.81	2.08	0.61	0.49	.644	0.17	26	2.63	0.46	2.27	0.47	1.69	.159	0.8	29
Tire	2.36	0.39	2.62	0.41	−1.37	.237	−0.66	7	2.51	0.49	2.02	0.69	1.43	.245	1	1
Fork	2.37	0.51	1.78	0.47	2.75	.045	1.15	43	2.78	0.47	3.01	0.42	−1.2	.291	−0.5	24
Tooth-brush	2.14	0.42	1.96	0.25	1.51	.190	0.44	39	2.72	0.36	2.49	0.1	3.99	.002	0.65	65

Note. $Human_R$: Ratings from humans. AI_R: Ratings from the AI. M_H: Arithmetic means from responses generated by humans. M_{GAI}: Arithmetic means from responses generated by GAI chatbots. *SD*: Standard deviation. *t*: t-value from the between-subjects *t*-test. *p*: p-value from the between-subjects *t*-test. *d*: Cohen's d. $H >$ GAI: Number of human participants who scored higher than the best GAI chatbot.

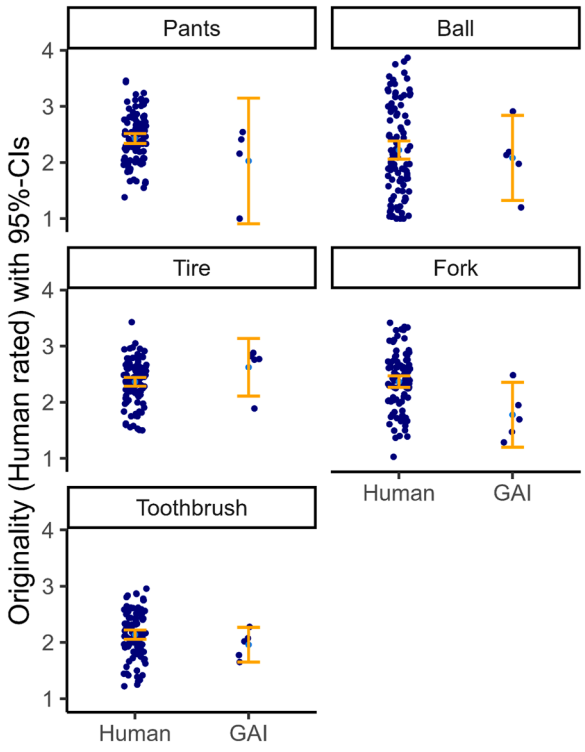

Fig. 1. Human-rated levels of originality for human and GAI-generated ideas.

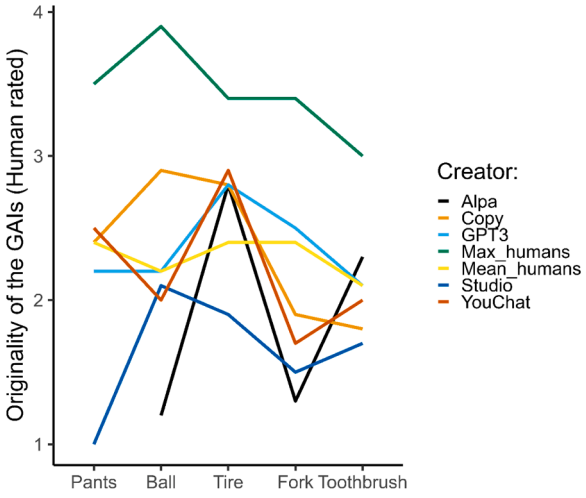

Fig. 2. Human-rated originality scores for each generative artificial intelligence (GAI), including the average score from humans and the score of the most creative human
Note. The alpa chatbot did not return any response for the prompt *pants*.

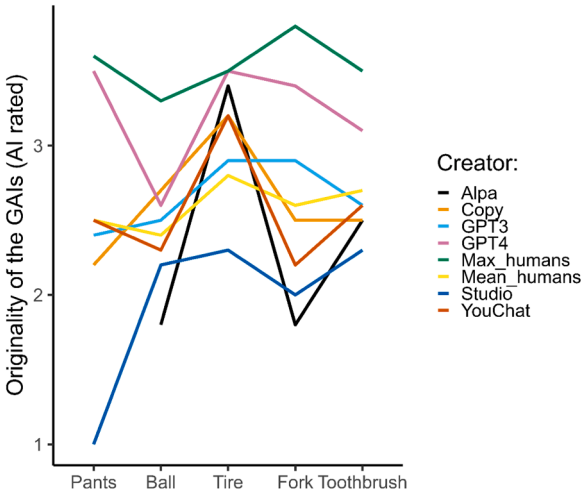

Fig. 3. AI-rated originality scores for each generative artificial intelligence (GAI), including the average score from humans and the score of the most creative human
Note. The alpa chatbot did not return any response for the prompt pants.

(Fig. 3).

When we compared the performance of ChatGPT version 4 to humans, 2 humans were more creative than the most creative AI for the prompt pants, 29 were more creative for the prompt ball, none were more creative for tire, 3 were more creative for fork, and 13 more creative for toothbrush. On average, 9.4 humans were more creative than ChatGPT version 4 across all prompts.

Fluency

Results for fluency are reported in the supplemental materials. In short, intraclass correlations and correlations between human raters and the AI were between 0.98 and 1.00. Since most of the GAI chatbots were prompted multiple times, the GAI chatbots came up with 2–3 times more ideas than humans (Figs. S3, S4). Fluency and originality were mostly unrelated across prompts and raters (humans vs AI), $rs = -0.28$ to 0.26.

Discussion

The question of whether GAIs such as ChatGPT, Studio.ai, and You.com can be considered creative is complex. Our research showed that their output for a standardized creativity measure for broad-associative "thinking" is as original as the human-generated ideas. Thus, from a scientific perspective, these chatbots are creative, as their output was judged as comparable to human output by human raters and AI. Some critics have argued that chatbots cannot replicate the creativity of humans, as human creativity is a combination of real-world experience,

emotion, and inspiration (Kirkpatrick, 2023; White, 2023). However, the definition and common measurement of creativity do not require these elements. It is defined as the ability to produce something new and useful (Runco & Jaeger, 2012), which can be judged by those engaging with the potentially creative output. We believe that this definition can also be applied to GAIs. Our results show that when chatbots are asked the same question as humans, they generate more ideas, which are, on average, as original as ideas generated by humans. As the sheer number of ideas is less important, and the assessment style between humans and chatbot conversations is less comparable, we do not want to stress the results for fluency too much. However, GAI chatbots can recombine knowledge so that the ideas presented are considered original.

The argument against GAIs' creative potential stems from two distinct but linked arguments: GAI is missing (so far) a connection to the real world, with emotions and imagination, and second, GAI is thus not capable of "actual" creativity, as high level creative endeavors. Although we cannot speak against both positions, we aim to advance this debate by closely looking into human creativity: generating creative output is much closer to recombining existing knowledge than actually developing anything new (Corazza, 2016; Frosio, 2023), and secondly, most humans do not come close to creative acts which are leading to eminent, high level creativity. Instead, we use and apply our human creativity to improve (and improvise) everyday tasks (Benedek et al., 2020; Reddy, 2022; Richards, 2010). This is not to belittle human creativity but instead aims to show the GAI chatbot's potential to be comparable to human creative abilities.

Especially art as a creative output seems driven by our human ability to dream, visualize and imagine potential futures. However, developing new ideas, which can serve a specific intention, solve an issue, or deliver an abstract meaning, is always built on a cumulative tradition of knowledge within the domain of art (Baer, 2015; Bruno, 2022). Most creativity-support systems in businesses thus focus on generating, processing, and retrieving knowledge (Maiden, 2019; Müller-Wienbergen et al., 2011). Brain scan analyses showed that idea generation is similar to knowledge retrieval (Benedek, 2018). Thus, similar to GAIs, we retrieve and recombine existing knowledge to make it appear new. Arguably, current databases of these chatbots do have a much larger knowledge base than any human being could possess, which makes the potential idea recombination that chatbots can provide much wider (Gruner & Csikszentmihalyi, 2019).

The second argument, the missing potential for high level or eminent creativity, seems unjust against the GAIs: human's ability of high level of creativity – bringing forward actual world-changing ideas – is also minimal. Mostly, we generate something new and useful for us in a specific and thus limited context. Our study shows GAI chatbots can compete with human ideation skills when it comes to everyday creativity. The prompts we used for the idea generation are very generic. When we consider more complex problems, a proper solution is achieved by including several factors, such as intense domain knowledge and creative thinking, individual subjective experiences, emotions, cultural background, and the capacity for abstract thinking. Here, current GAI chatbots appear to perform well on complex knowledge-intensive tasks, such as complex coding tasks: ChatGPT can free up coders on tedious work (Bellaiche et al., 2023) so that the coder can focus on more complex, creative work aspects (Dell'Aversana, 2023). However, ChatGPT is shown to be rather limited in emotional responses and evaluations and shows less reliable performance with more complex tasks (Kocoń et al., 2023).

Overall, GAI chatbots show a convincing human-like or potentially above-human performance for some tasks which was also found in research published after the present study was conducted (Guzik et al., 2023), whereas their performance is limited in others. Concerning creative performance, GAI can generate ideas based on specific input but cannot create the need to ideate. The motivation to engage with a specific creative task and problem understanding must come from the human interacting with the tool (Cropley et al., 2023). Thus, GAI is limited considering the overall creative process: it would not trigger the creative process. It can only respond to a prompt that is given. Thus, the problem definition is currently still uniquely human, as is evaluating whether an idea fits a problem. Although, for particular contexts, such as the assessment of the AUT output, an AI sufficiently assessed the quality of the generated ideas (Organisciak et al., 2023). An intriguing solution to the debate whether AI can be creative was recently proposed by Runco (2023), who suggested that we should refer to "Artificial Creativity" as we are referring to "Artificial Intelligence". Runco challenges the notion of "creative AI" and suggests labeling computer-generated outputs as "artificial creativity," emphasizing that while such outputs may be original and effective, they lack essential human creative qualities.

GAI chatbots possess the capability to uncover novel connections by leveraging their extensive knowledge base at specific stages of the creative process. Human involvement is crucial in framing a pertinent problem and executing the chosen solution. Our study shows that chatbots can generate ideas on the same level as humans, especially on the level of everyday creativity (with ChatGPT version 4 showing the best results, followed by Copy.ai, ChatGPT version 3.5, and YouChat scoring all similarly high in terms of the originality of ideas). Whether the person interacting with the GAI achieves every day or eminent levels of creative achievement are more up to the person than the GAI. GAI can successfully support the creative process and generate ideas, but it remains the task of humans to make sense of it and embed this in physical reality (Verganti et al., 2020). This speaks for the merging of human and artificial competencies in the form of augmented creativity – which is the most optimistic and benevolent future development for the usage of AI (Vinchon et al., 2023).

There are some limitations to our research. Our experimental design likely led to an underestimation of the creativity of humans and GAIs. We paid participants to generate ideas for creative tasks they might not care about. However, intrinsic motivation strongly contributes to creative performance (Amabile & Pratt, 2016; Luria & Kaufman, 2017), potentially leading to an overall human underperformance. Regarding chatbots, smart prompting is how the best answers are obtained, which we did not use to allow a direct comparison between humans and chatbots. The actual potential for chatbots as creative assistants is likely much higher. Tailored prompts and reshaping answers given by the chatbots will likely lead to much more concise and, thus, relevant answers. Also, chatbots can be used to get information from a specific angle, such as for a certain profession, which can improve the quality of answers a user seeks, which we did not test here either.

Although the AUT is a widely used creativity measurement, there is an ongoing debate regarding its validity (Karwowski, 2015; Runco et al., 2016). As chatbots use wide parts of the internet as a source of their data, it could be the case these databases include test material and thus previously given human answers to the prompts used by the AUT. It is essential to recognize that the responses of chatbots are predominantly determined by statistical predictions of the most appropriate subsequent words (OpenAI, 2023b). Hence, even if exact human responses to AUT prompts are embedded in a chatbot's foundational training dataset, they would not necessarily manifest in the produced output.

We did not measure usefulness to assess the reported ideas because originality is the more important part of the "new and useful" definition (Diedrich et al., 2015). Further, judging an idea's usefulness is difficult without a proper real-life application to serve as an anchor. When we assessed the AUT with the chatbots, we pushed for more answers, with a relatively arbitrary number maximum of three times. In our past interactions with chatbots, a recurrent theme emerged: most chatbots tended to yield a fairly standard volume of content in response to open-ended queries. While each began with a predetermined set of ideas, their adaptability and depth differed when nudged for more, using prompts like "what else." It's worth noting that some platforms, such as Alpa.ai, reached their output threshold rather swiftly, either stalling with no additional feedback or generating incongruous replies. To

institute uniformity in our evaluations, we limited the succession of probing questions to three. This limitation, albeit somewhat arbitrary, complicates direct comparisons between the prolificacy of chatbot-generated content and human responses, if solely based on the number of responses.

Research has shown that exposure to other people's creative ideas can stimulate cognitive activity and enhance creativity (Fink et al., 2012). Participants who were prompted with highly creative ideas generated more creativity than those who were given random, unrelated words. In this study, the comparison between the most original humans vs. chatbot shows that humans had the most creative answers in all but one case. Thus, humans can (still) serve as ideation partners. However, on a more pragmatic note, it might be easier to ask a chatbot than to find a motivated human to run ideas by. Our study and a lot of anecdotal evidence on the web show the possibility of generating creative output in combination with a GAI, be it a writing tool, chatbot, or picture generation. The potential is real for GAIs to properly support human (creative) work. However, the ethical dilemma needs to be addressed, as the potential for misuse ("The AI Writing on the Wall,", 2023) or harmful application of the GAIs is present as with any potent technology ("Much to Discuss in AI Ethics,", 2022).

Taken together, the question whether GAI is creative can be answered pragmatically with: yes, on an everyday level as much or as little as humans. We recommend avoiding viewing GAI chatbots as omnipotent tools that may replace human performance. Instead, they can be valuable assistants in provoking, listing, and reviewing thoughts and ideas. The extensive knowledge base they build upon can be very useful in expanding one's ideas. The more our (working) life's are automated, and the more authority automation acquires, the more important the human role with its creative abilities and motivation becomes (Parasuraman et al., 2008; Runco, 2023).

Funding

This work was partly supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DII133 (Weizenbaum-Institute). The responsibility for the content of this publication remains with the authors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.yjoc.2023.100066.

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*(5), 997.

Amabile, T. M. (2017). In pursuit of everyday creativity. *The Journal of Creative Behavior, 51*(4), 335–337. https://doi.org/10.1002/jocb.200

Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior, 36*, 157–183. https://doi.org/10.1016/j.riob.2016.10.001

Baer, J. (2015). The importance of domain-specific expertise in creativity. *Roeper Review, 37*(3), 165–178. https://doi.org/10.1080/02783193.2015.1047480

Bates D., Maechler M., Bolker B., Walker S., Christensen R.H., Singmann H., & Dai B. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014.

Bellaiche L., Shahi R., Turpin M., Ragnhildstveit A., Sprockett S., Barr N., Christensen A., & Seli, P. (2023). Humans vs. AI: Whether and why we prefer human-created compared to AI-created artwork. 10.31234/osf.io/f9upm.

Benedek, M. (2018). The Neuroscience of Creative Idea Generation. In Z. Kapoula, E. Volle, J. Renoult, & M. Andreatta (Eds.), *Exploring Transdisciplinarity in Art and Sciences* (pp. 31–48). Springer International Publishing. https://doi.org/10.1007/978-3-319-76054-4_2.

Benedek, M., Bruckdorfer, R., & Jauk, E. (2020). Motives for creativity: Exploring the what and why of everyday creativity. *The Journal of Creative Behavior, 54*(3), 610–625. https://doi.org/10.1002/jocb.396

Botella, M., Nelson, J., & Zenasni, F. (2019). It is time to observe the creative process: how to use a creative process report diary (CRD). *The Journal of Creative Behavior, 53* (2), 211–221. https://doi.org/10.1002/jocb.172

Bruno, C., & Bruno, C. (2022). Digital creativity dimension: A new domain for creativity. *Creativity in the design process: exploring the influences of the digital evolution* (pp. 29–42). Springer International Publishing.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6* (4), 284.

Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications, 209*, Article 118190. https://doi.org/10.1016/j.eswa.2022.118190

Cope, D. (2005). *Computer models of musical creativity*. Mit Press Cambridge.

Corazza, G. E. (2016). Potential originality and effectiveness: The dynamic definition of creativity. *Creativity Research Journal, 28*(3), 258–267. https://doi.org/10.1080/10400419.2016.1195627

Cropley, D. H., Medeiros, K. E., Damadzic, A., Henriksen, D., & Mishra, P. (2023). The intersection of human and artificial creativity. *Creative provocations: speculations on the future of creativity, technology & learning* (pp. 19–34). Springer International Publishing. 10.1007/978-3-031-14549-0_2.

Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering, 27*(1), 113–118. https://doi.org/10.1017/S1351324920000601

Dell'Aversana P. (2023). GPT-3: A new cooperation scenario between humans and machines. Benefits and limitations of GPT-3 as a coding virtual assistant. ResearchGate. 10.13140/RG.2.2.32450.04800.

Diedrich, J., Benedek, M., Jauk, E., & Neubauer, A. C. (2015). Are creative ideas novel and useful? *Psychology of Aesthetics Creativity and the Arts, 9*, 35–40. https://doi.org/10.1037/a0038688

DiPaola, S., & McCaig, G. (2016). *Using artificial intelligence techniques to emulate the creativity of a portrait painter*. Electronic Visualisation and the Arts. https://doi.org/10.14236/ewic/EVA2016.32

Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*. https://doi.org/10.1038/d41586-023-00056-7

Fink, A., Koschutnig, K., Benedek, M., Reishofer, G., Ischebeck, A., Weiss, E. M., & Ebner, F. (2012). Stimulating creativity via the exposure to other people's ideas. *Human Brain Mapping, 33*(11), 2603–2610. https://doi.org/10.1002/hbm.21387

Frosio, G. (2023). The Artificial Creatives: The Rise of Combinatorial Creativity from Dall-E to GPT-3. Martha Garcia-Murillo, Ian MacInnes, and Andrea Renda. In *Handbook of Artificial Intelligence at Work: Interconnections and Policy Implications*. Edward Elgar, Forthcoming), Queen's University Belfast Law Research Paper. https://papers.ssrn.com/abstract=4350802.

Gamer, M., Lemon, J., & Singh, I. F. P.. irr: Various coefficients of interrater reliability and agreement. R package version 0.84. 1. 2019. https://Cran.r-Project.

Gruner, D. T., & Csikszentmihalyi, M. (2019). Engineering creativity in an age of artificial intelligence. *The Palgrave Handbook of Social Creativity Research*, 447–462.

Guzik, E. E., Byrge, C., & Gilde, C. (2023). The originality of machines: AI takes the torrance test. *Journal of Creativity, 33*(3), Article 100065. https://doi.org/10.1016/j.yjoc.2023.100065

Holford, W. D. (2019). The future of human creative knowledge work within the digital economy. *Futures, 105*, 143–154. 10.1016/j.futures.2018.10.002.

Karwowski, M. (2015). Notes on creative potential and its measurement. *Creativity. Theories Research Applications, 2*(1), 4–16. https://doi.org/10.1515/ctra-2015-0001

Kaufman, J. C., Beghetto, R. A., & Watson, C. (2016). Creative metacognition and self-ratings of creative performance: A 4-C perspective. *Learning and Individual Differences, 51*, 394–399. https://doi.org/10.1016/j.lindif.2015.05.004

Kirkpatrick, K. (2023). Can AI demonstrate creativity? *Communications of the ACM, 66* (2), 21–23.

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., … Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion, 99*, 101861. 10.1016/j.inffus.2023.101861.

Lazar, M., Miron-Spektor, E., & Mueller, J. S. (2022). Love at first insight: An attachment perspective on early-phase idea selection. *Organizational Behavior and Human Decision Processes, 172*, Article 104168. https://doi.org/10.1016/j.obhdp.2022.104168

Luria, S. R., & Kaufman, J. C. (2017). The dynamic force before intrinsic motivation: Exploring creative needs. *The creative self* (pp. 317–325). Elsevier.

Maiden, N. (2019, June). *Digital creativity support: designing AI to augment human creativity. In The Human Position in an Artificial World* (pp. 145–146). Creativity, Ethics and AI in Knowledge Organization.

Miller, A. I. (2019). *The artist in the machine: the world of AI-powered creativity*. MIT Press.

Much to discuss in AI ethics. *Nature Machine Intelligence, 4*(12), (2022). https://doi.org/10.1038/s42256-022-00598-x

Müller-Wienbergen, F., Müller, O., Seidel, S., & Becker. (2011). Leaving the beaten tracks in creative work – A design theory for systems that support convergent and divergent thinking. *Journal of the Association for Information Systems, 12*(11). https://doi.org/10.17705/1jais.00280

Marcus, G., Davis, E., & Aaronson, S. (2022). A very preliminary analysis of DALL-E 2. *arXiv preprint* arXiv:2204.13807.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283*.

OpenAI. (2023a). *ChatGPT-4* (Mar 14 version) [Large language model].
OpenAI. (2023b). *GPT-4 Technical Report* arXiv. 10.48550/arXiv.2303.08774.
Organisciak, P., Acar, S., Dumas, D., & Berthiaume, K. (2023). Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity, 101356*. https://doi.org/10.1016/j.tsc.2023.101356
Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140–160. https://doi.org/10.1518/155534308X284417
Pistrui, J. (2018). The future of human work is imagination, creativity, and strategy. *Harvard Business Review*.
Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*(2), 83–96. https://doi.org/10.1207/s15326985ep3902_1
Reddy, A. (2022). Artificial everyday creativity: Creative leaps with AI through critical making. *Digital Creativity, 33*(4), 295–313. https://doi.org/10.1080/14626268.2022.2138452
Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 144.
Richards, R., Kaufman, J. C., & Sternberg, R. J. (2010). Everyday creativity: process and way of life – Four key issues. *The cambridge handbook of creativity* (pp. 189–215). Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.013
Runco, & Jaeger, G. J (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092
Runco, M. A. (2023). AI can only produce artificial creativity. *Journal of Creativity*. , Article 100063. https://doi.org/10.1016/j.yjoc.2023.100063
Runco, M. A., Abdulla, A. M., Paek, S. H., Al-Jasim, F. A., & Alsuwaidi, H. N. (2016). Which test of divergent thinking is best? *Creativity Theories Research Applications, 3*(1), 4–18. https://doi.org/10.1515/ctra-2016-0001
Sawyer, K. (2012). Extending sociocultural theory to group creativity. *Vocations and Learning, 5*(1), 59–75.
Seidel, S., & Berente, N. (2020). Automate, informate, and generate: Affordance primitives of smart devices and the Internet of Things. *Handbook of Digital Innovation*, 198–210.
Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity and the Arts, 2*(2), 68–85.
Simonton, D. K. (2013). What is a creative idea? Little-c versus Big-C creativity. *Handbook of Research on Creativity, 2*, 69–83.
Steele, L. M., Johnson, G., & Medeiros, K. E. (2018). Looking beyond the generation of creative ideas: Confidence in evaluating ideas predicts creative outcomes. *Personality and Individual Differences, 125*, 21–29. https://doi.org/10.1016/j.paid.2017.12.028
The AI writing on the wall. *Nature Machine Intelligence, 5*(1), (2023). https://doi.org/10.1038/s42256-023-00613-9
Urban, M., Dechterenko, F., Lukavsky, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2023). *ChatGPT Improves Creative Problem-Solving Performance in University Students: An Experimental Study*. Psy. arXiv:10.31234/osf.io/9z2tc.
Vartanian, O., Beatty, E. L., Smith, I., Forbes, S., Rice, E., & Crocker, J. (2019). Measurement matters: The relationship between methods of scoring the alternate uses task and brain activation. *Current Opinion in Behavioral Sciences, 27*, 109–115. https://doi.org/10.1016/j.cobeha.2018.10.012
Verganti, R., Vendraminelli, L., & Iansiti, M. (2020). Innovation and design in the age of artificial intelligence. *Journal of Product Innovation Management, 37*(3), 212–227. https://doi.org/10.1111/jpim.12523
Vinchon, F., Lubart, T., Bartolotta, S., Gironnay, V., Botella, M., Bourgeois-Bougrine, S., … Gaggioli, A. (2023). Artificial Intelligence & Creativity: A Manifesto for Collaboration. *The Journal of Creative Behavior, 597*. 10.1002/jocb.597.
White, C. (2023). *Opinion: artificial intelligence can't reproduce the wonders of original human creativity*. The Star. https://www.thestar.com.my/tech/tech-news/2023/01/18/opinion-artificial-intelligence-cant-reproduce-the-wonders-of-original-human-creativity.
Williams, R., Runco, M., & Berlow, E. (2016). Mapping the themes, impact, and cohesion of creativity research over the last 25 years. *Creativity Research Journal, 28*, 385–394. https://doi.org/10.1080/10400419.2016.1230358