AMERICAN PSYCHOLOGICAL ASSOCIATION
DIVISION 10
Psychology of Aesthetics, Creativity, and the Arts

© 2025 American Psychological Association
ISSN: 1931-3896
https://doi.org/10.1037/aca0000743

Beyond the Brush: Human Versus Artificial Intelligence Creativity in the Realm of Generative Art

Paul Seli[1], Anya Ragnhildstveit[2], William Orwig[3], Lucas Bellaiche[1], Sarah Spooner[1], and Nathaniel Barr[4]
[1] Department of Psychology and Neuroscience, Duke University
[2] Department of Psychiatry, University of Cambridge
[3] Department of Psychology, Harvard University
[4] Faculty of Humanities and Social Sciences, Sheridan College

Several decades ago, J. P. Guilford (1950) speculated that, although artificial intelligence (AI) might one day take over much of human thinking, creativity would remain a uniquely human faculty. However, rapid advancements in AI have led to the recent development of generative models that are capable of producing high-quality creative products, casting some doubt Guilford's prediction. Here, to shed light on this issue, we explored whether human creativity retains distinct value in the emerging domain of AI-assisted digital artworks. Using DALL-E 3, we generated images based on prompts crafted by professional artists, novice artists, and an AI chatbot (ChatGPT-4). Creativity ratings from 299 participants showed that images produced from professional artist prompts were rated as most creative, followed by those generated from AI chatbot prompts, with novice artist prompts rated lowest. Further analysis suggested that this pattern may be partially explained by the semantic distance of the prompts, which followed the same pattern (i.e., professionals > AI chatbot > novices). These findings suggest that, although AI demonstrates impressive creative abilities, the worth of human creativity has not been diminished in the domain of AI-assisted outputs, which supports Guilford's view on the enduring value of human creativity.

Keywords: creativity, generative artificial intelligence, human-artificial intelligence collaboration, creativity assessment, digital art

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In his 1950 Presidential Address to the American Psychological Association, J. P. Guilford foreshadowed the development of artificial intelligence (AI) in modern society, predicting that if advances in "thinking machines" (i.e., computers) reached their forecasted height, then "the only economic value of brains left would be the creative thinking of which they are capable" (Guilford, 1950, p. 446). Put differently, Guilford suggested that AI might eventually advance to the point at which it would replace human labor in countless domains, leaving creativity as the last bastion of uniquely human value. Almost 75 years later, Guilford's prediction appears prescient in many respects. Indeed, advancements in AI have now led to the automation of several tasks that were traditionally performed exclusively by humans, including manufacturing, medical diagnostics, legal practices, software programming, driving, and customer service (Buttazzo, 2023). Across these domains, AI has not only augmented human capabilities but has also in many cases replaced human labor altogether. As a result, the "economic value" of human contributions in such roles has clearly diminished as AI systems have increasingly expanded their capabilities and now perform a growing number of tasks with efficiency, precision, and cost-effectiveness beyond human reach.

While Guilford's prediction that AI would eventually replace human labor has become an undeniable reality, what is less clear is whether he was correct in suggesting that human creativity would remain uniquely valuable in a future wherein "thinking machines" reach their forecasted height. Over the past few years, this topic has become a focal point of debate among scientists, philosophers, and creatives (e.g., Krietzberg, 2023). And, the stakes are high: As AI technologies continue to grow more sophisticated, there is an increasing concern that at least some domains of human creativity may already be at risk of obsolescence.

In an effort to shed light on this issue, recent research has pit humans against AI in various creativity tasks to determine the victor. In one such study, Hubert et al. (2024) compared the performance of a popular AI model—ChatGPT-4—to human performance on tasks of divergent thinking, a widely recognized measure of creativity that involves generating multiple solutions or ideas in response to open-ended prompts (e.g., Runco & Acar, 2012). To assess divergent thinking, both AI and human participants completed a series of tasks requiring them to generate creative responses to prompts, such as identifying novel and useful applications of common objects (e.g., the alternate uses task; Guilford, 1967). Hubert et al. found that AI, on average, outperformed humans across all metrics of divergent thinking.

Given these findings, one might reasonably conclude—as Hubert et al. (2024) did—that AI is more creative than humans, at least

Dirk B. Walther served as action editor.
Paul Seli https://orcid.org/0000-0002-5398-6999
The hypotheses, methods, and analysis plan were preregistered at https://aspredicted.org/vu84d.pdf. All study materials, analysis scripts, and primary data are publicly available on the Open Science Framework at https://osf.io/njpkz/.
Correspondence concerning this article should be addressed to Paul Seli, Department of Psychology and Neuroscience, Duke University, Room 205, Reuben-Cooke Building, 417 Chapel Drive, Durham, NC 27708, United States. Email: paul.seli@duke.edu

in the realm of divergent thinking. However, in a separate study, Koivisto and Grassini (2023) reasoned that the story may be more nuanced. Indeed, whereas Hubert et al. demonstrated that, on average, AI showed superior divergent-thinking performance relative to humans, they did not explore whether some fraction of their human participants outperformed AI. Importantly, identifying such high-performing individuals would serve as a proof of concept that, while AI surpasses the average human on measures of divergent thinking, exceptional human creativity still retains unique value in this domain.

To explore this possibility, Koivisto and Grassini (2023) compared divergent-thinking performance of ChatGPT-4 and humans, following a similar approach to Hubert et al. As in Hubert et al., they found that, on average, AI indeed outperformed human participants. However, when examining individual performance, they also observed that the best human ideas matched or exceeded those generated by AI. Thus, at least within the domain of divergent thinking, Guilford's prediction finds partial support: among top performers, human creativity retains unique value.

While the aforementioned research focused on divergent thinking, another domain of creativity that has recently garnered significant attention is AI-assisted digital artwork. This process involves human users collaborating with AI to generate visual art, often through platforms that interpret text descriptions to create images. One prominent example is DALL-E 3, which was developed by OpenAI. With this platform, users input descriptive text prompts—e.g., "a futuristic cityscape at sunset" or "a cat dressed as an astronaut in outer space"—and DALL-E 3 analyzes these prompts using complex language- and image-recognition algorithms. It then generates visual outputs that closely align with the user's description, effectively translating words into images. By combining human intention with AI's vast data set of visual knowledge, DALL-E 3 enables users to produce unique artworks that reflect a confluence of AI's interpretative abilities and human creative input.

Notably, the implications of AI-assisted digital art extend far beyond a casual tool for hobbyists. Indeed, this blend of human input and AI generation has already had a dramatic impact on the fine arts. For example, in 2022, Jason Allen's *Théâtre D'opéra Spatial*, a piece created through his collaboration with the AI tool Midjourney, clinched the first prize at the Colorado State Fair's annual arts competition, outperforming human artwork that was created without AI support. Perhaps unsurprisingly, this outcome sparked significant debate regarding the place of AI-generated art in traditional art spaces, raising questions about artistic originality and authorship (e.g., Bullis, 2024).

Beyond fine arts, AI-assisted digital artwork is also transforming industries like advertising. For instance, for Coca-Cola's *Create Real Magic* campaign, the brand invited digital artists to use an AI platform (powered by GPT-4 and DALL-E 3) to incorporate Coca-Cola's iconic symbols into original AI-assisted art (The Coca-Cola Company, 2023). Participants used this platform to create AI-driven visuals that blended Coca-Cola's brand imagery with their own designs. Select pieces were then showcased on prominent billboards, including in Times Square, illustrating how AI-assisted artwork is reshaping brand engagement and visual marketing.

AI-assisted art is also finding applications in education, where it has recently been introduced as a tool with which to teach and inspire students across multiple disciplines. For example, at the University of Illinois Urbana-Champaign, graphic design professors have begun to incorporate the use of AI tools in the classroom, with the goal of assisting students in visualizing their ideas and enhancing their creative projects (Heckel, 2023). Much like in fine arts and industry, AI-assisted art in education points to the broadening role of AI in contemporary society.

Given these recent advances, a reasonable concern among visual artists (and aspiring visual artists) is that traditional creative skill sets might have relatively diminished utility. If, for instance, AI-assisted outputs are similarly creative regardless of a user's artistic ability, then this would suggest that cultivating traditional artistic skills might hold limited value in this context. Indeed, if both novice and professional artists produced equally creative AI-assisted artworks, then "artistic ability," as traditionally understood, might be inconsequential in this burgeoning domain. Worse still, if AI chatbots were to independently outperform humans when collaborating with AI-art generators (e.g., ChatGPT generating prompts for DALL-E 3), then Guilford's claim would be in jeopardy: Human creativity might hold little economic value, at least in the realm of AI-assisted visual arts. As noted by Orwig et al. (2024), such an outcome could deeply affect how creativity is taught, appreciated, and valued, with implications that extend beyond individual artists to the broader creative ecosystem.

In light of these concerns, researchers have recently begun to explore whether human creativity still holds value in the realm of AI-assisted artworks. As suggested above, one key question within this context is whether individual differences in human creative ability are predictive of the quality of AI-assisted art—that is, whether more creative individuals tend to outperform less creative individuals. To address this question, Orwig et al. (2024) conducted a recent study in which they began by assessing participants' creative abilities through standard divergent-thinking tasks. Following these assessments, participants were instructed to generate creative text prompts for an AI art generator. The research team then entered each of these prompts into DALL-E 3 to produce corresponding digital images, which independent raters subsequently evaluated for creativity on a scale from 1 = *very uncreative* to 5 = *very creative*.[1] Results from the study revealed a modest, yet significant, positive correlation between participants' divergent-thinking performance and the rated creativity of their AI-assisted images. Further analysis revealed that this relationship was mediated by the semantic distance of the word prompts inputted to the AI: participants with higher creative ability (indexed as divergent thinking) tended to generate prompts with greater semantic diversity, which in turn led to more creative AI-assisted images. Critically, these results suggest that human creativity does appear to matter in the context of AI-assisted art, although perhaps only to a small extent.

While Orwig et al.'s (2024) study has shed valuable light on the role of human creative ability in the context of AI-assisted art, it leaves open some interesting questions for further investigation. One question is whether creative ability in the visual arts—rather than divergent-thinking ability—is predictive of the creativity of

[1] Raters were not provided with a specific definition of creativity; instead, they were instructed to apply their own interpretations when evaluating each image (as in Bellaiche, Smith, et al., 2023). This approach allows for a subjective assessment of creativity based on the individual perspectives of the raters. Notably, despite these relatively open-ended instructions, interrater reliability for creativity assessments was high, suggesting a shared understanding among raters regarding what constitutes creativity in the context of digital art.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

AI-assisted artwork. In other words, do experienced visual artists produce more creative AI-assisted artwork than those who are relatively inexperienced? Since expertise in the visual arts appears to more directly align with generating AI-assisted visual artwork than does divergent thinking, this seems a reasonable possibility. Such an outcome would likely be encouraging to artists, as it would suggest that their creative skills hold distinct value in the ever-developing field of AI-assisted visual art.

The second open question is perhaps even more pressing: How do humans—whether seasoned or novice artists—compare to AI chatbots in generating creative AI-assisted artworks? Even if experienced artists outperform novices, this distinction would prove inconsequential if AI ultimately surpasses both groups in creative tasks. Indeed, in this scenario, AI would likely take over roles traditionally reserved for human artists, regardless of their creative skill level. Such an outcome would thus challenge Guilford's assertion that creativity might be the last bastion of unique human value, suggesting that, at least within the domain of AI-assisted art, human creativity may not hold distinct economic value.

The Present Study

Here, we directly address both of these questions in order to further explore the role of human creativity in AI-assisted art. Specifically, we investigate whether individuals with expertise in the visual arts produce more creative AI-assisted artworks than those with limited experience. Additionally, we test how these human–AI collaborations measure up against artworks generated exclusively by AI (absent human input). To this end, we had groups of novice and professional artists to generate creative text prompts and also obtained prompts that were generated by GPT-4. The prompts from each of these three sources were then input into an AI art generator (DALL-E 3) to produce corresponding AI-assisted images. Subsequently, we obtained creativity ratings for each of these images from a large, online sample.

Method

Research Transparency Statement

This research was approved by the Duke University Campus Institutional Review Board (2022-0534). The hypotheses, methods, and analysis plan were preregistered (https://aspredicted.org/vu84d.pdf) on 2024-03-18, prior to data collection, which began on 2024-03-26. While the preregistered analyses were conducted as planned, we also performed additional, exploratory analyses in response to reviewer comments. These supplemental analyses, which diverge from the preregistration, are distinguished from the preregistered analyses in the following sections. All study materials, analysis scripts, and primary data are publicly available (https://osf.io/njpkz/).

Stimulus Set

To begin our investigation, we first established the three groups for comparison: professional artists, novice artists, and AI (ChatGPT-4). Professional artists were defined as individuals who met the following criteria: (a) at least 5 years of experience in visual arts and (b) a minimum of 1,000 hr dedicated to visual arts ($M_{age} = 40.73$, 60% female). In contrast, novice artists were defined as individuals who did not meet either of these criteria ($M_{age} = 41.53$, 67% female). Importantly, for both groups, we opted to select only individuals with no prior experience using AI art generators. This criterion was implemented to ensure that differences observed between groups were not the result of varying levels of experience with AI tools.

Having established the defining criteria for our groups, we then pretested 500 participants online via Prolific (an online recruitment platform widely used in cognitive research), to recruit individuals for the professional and novice artist groups in our study. From this pool, we identified 15 professionals who met all three criteria. To ensure a balanced comparison, we then randomly selected 15 individuals—from the same pretest pool of 500—who met the criteria for novice artists. For comparison, AI (ChatGPT-4) was included as a distinct group in our analysis.

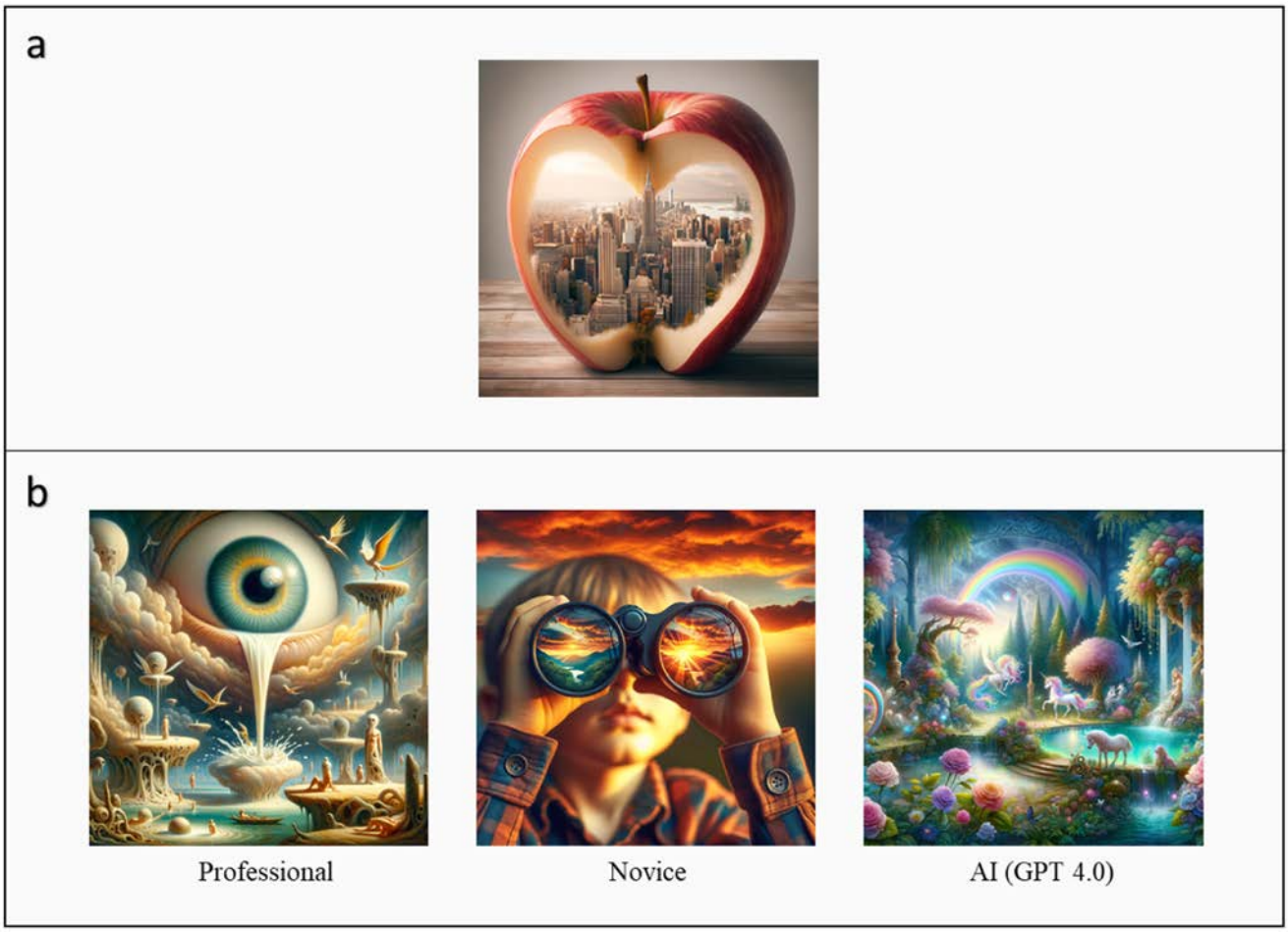
Each group was tasked with producing a creative word prompt, not exceeding 15 words, for a hypothetical art generator capable of transforming such word prompts into digital images. An example word prompt ("An apple bigger than New York City") and its corresponding image, created by the art generator (Figure 1a), were shown as a guide. Moreover, to avoid style-based confounding effects, participants were instructed not to include details about the artistic style (e.g., "oil painting," "abstract," "Cubist") in their word prompts; otherwise, they had complete freedom over the grammatical structure and wording of their prompts. Finally, they were informed that their word prompts would be used by the AI art generator, but they would not see the associated images produced. AI (ChatGPT-4) received identical instructions to produce 15 creative word prompts. After collecting all 45-word prompts from professional artists, novice artists, and AI (ChatGPT-4), a research assistant entered these prompts into DALL-E 3. This procedure took place on March 26, 2024. The 45 generated digital images were then aggregated to serve as the stimulus set for the present study. See Figure 1b for samples.

Participants

Following the development of our stimulus set, we examined whether creativity differed across the three groups: professional artists, novice artists, and AI (ChatGPT-4). As such, we recruited a convenience sample of 300 participants, through Prolific, to complete a 12-min questionnaire ($M_{age} = 36.31$, 49% female, 49% male, 0.7% "other," and 1.7% preferred not to disclose their gender).

We considered this sample size to be reasonable based on several factors. Firstly, a formal power analysis was conducted to determine the appropriate sample size for our repeated-measures analysis of variance (ANOVA) and follow-up paired-samples t tests across three conditions (see the Results section). For the repeated-measures ANOVA, we aimed to detect a medium effect size ($\eta_p^2 = .06$) with 80% power at a 5% significance level. According to standard power analysis guidelines (e.g., Cohen, 1988), a sample size of approximately 128 participants is sufficient for such analyses. However, given that our study involved multiple conditions and follow-up tests, we increased our sample size to 300 to ensure robust and reliable results. Additionally, for the paired-sample t tests, a medium effect size (Cohen's $d = 0.5$) with 80% power at a 5% significance level requires approximately 34 participants per pairwise comparison. Our sample size of 300 participants therefore allows for more than sufficient power to detect medium to large effects in these comparisons. Moreover, the sample size far exceeds the minimum requirements suggested by prior literature for similar research contexts, thus enhancing the robustness of our results.

Figure 1
Artwork Generated by DALL-E 3



Note. (a) Digital image, generated by DALL-E 3, shown to participants to exemplify what an art generator might produce when inputted with the creative word prompt: "An apple bigger than New York City." (b) Samples of artworks, generated by DALL-E 3, categorized as a function of the creative word prompts provided by professional artists, novice artists, and AI (ChatGPT-4). AI = artificial intelligence. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Participants received a total of US$2.30, in line with standard compensation rates for Prolific. To ensure high-quality responses, two attention checks were embedded within the questionnaire. One participant failed both attention checks, resulting in the exclusion of their data from all analyses (see preregistration for exclusion criteria); of the remaining participants, all correctly responded to both attention checks, and their data were included in the final analysis. Thus, the final sample size was 299.

Procedure

As part of the questionnaire, participants were asked to rate the creativity of 45 digital artworks (i.e., the stimulus set). No information was provided about the source of the digital images (cf. Bellaiche, Shahi, et al., 2023; Bellaiche, Smith, et al., 2023; Horton et al., 2023). To ensure that participant judgments were made in reference to the full stimulus set, each of the 45 images was initially displayed in random order for 2 s. During this phase, participants were simply asked to view each image. Next, participants were shown each of the 45 images, again in random order, and asked to indicate how creative they appeared, using a Likert scale from 1 = *not at all creative* to 5 = *very creative*. They were allotted as much time as needed to make each creativity judgment. As in Orwig et al. (2024), raters were not provided a strict definition of creativity but were instead instructed to evaluate each image based on their personal interpretation.

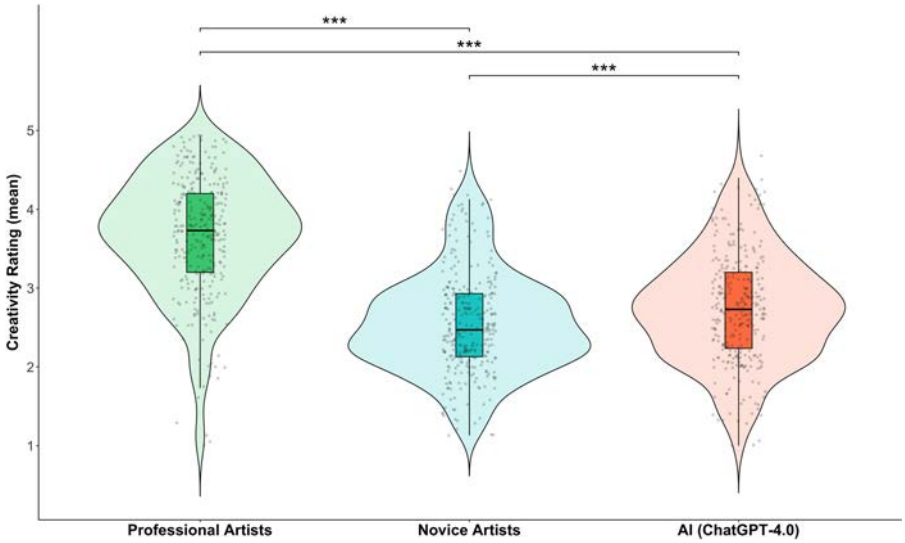
Results

Preregistered Analyses

A repeated measures ANOVA was conducted to examine differences in creativity for the digital images produced across the three groups: professional artists, novice artists, and AI (ChatGPT-4; Figure 2). Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(2) = 119.224$, $p < .001$. To correct for this violation, the Greenhouse–Geisser correction was applied. Results showed a significant main effect of group on creativity ratings, $F(1.503, 596) = 578.52$, $\eta_p^2 = .66$, $p < .001$.

Figure 2

Creativity Ratings Across Conditions



Note. Violin plots illustrating the mean creativity ratings ($N = 299$) for digital images generated by professional artists, novice artists, and AI (ChatGPT-4). The median is indicated by the line within the box of each violin plot, and the IQR is represented by the box (25th to 75th percentile). The whiskers extend to the furthest data points within 1.5 times the IQR from the lower and upper quartiles. AI = artificial intelligence; IQR = interquartile range. See the online article for the color version of this figure.
*** $p < .001$.

Next, we conducted three paired-samples t tests to statistically compare these creativity ratings across the three groups. These tests revealed significant differences across all three groups: digital images created by professional artists were rated as significantly more creative than those by novice artists, $t(298) = 29.94$, $d = 1.73$ $p < .001$, and by AI, $t(298) = 22.21$, $d = 1.29$, $p < .001$. Furthermore, digital images generated by AI were rated as significantly more creative than those by novice artists, $t(298) = -7.69$, $d = 0.45$, $p < .001$. Thus, participants' perceptions of creativity strongly favored professional artists' works, followed by AI-generated works, and then by novice artists' works.

Exploratory Analyses

In light of thoughtful reviewer feedback, we decided to conduct additional exploratory analyses to better understand what might be driving the observed creativity differences across the three groups. Specifically, we explored two potential factors: (a) the word count of each group's prompts and (b) the semantic distance of their prompts.

Word Count

We first assessed whether differences in creativity ratings might be attributed to the length of prompts. Here, we reasoned that longer prompts could offer more context, potentially guiding the AI to generate images with richer details that might, in turn, enhance perceptions of creativity. To assess this possibility, we compared the word counts of the prompts generated by professional artists, novice artists, and AI. Results revealed a significant main effect of group, $F(2, 42) = 38.56$, $\eta^2 = .65$, $p < .001$, indicating that word count varied significantly across groups (Figure 3a).

Next, we conducted three independent-samples t tests to statistically compare word counts across the three groups. Professional artists' prompts did not differ significantly in word count from those generated by AI, $t(28) = 1.02$, $p = .32$, $d = 0.37$. However, both professional artists, $t(28) = 7.86$, $p < .001$, $d = 2.87$, and AI, $t(28) = 7.35$, $p < .001$, $d = 2.68$, generated prompts with significantly higher word counts than those created by novice artists. Thus, our analyses indicate that prompts generated by both professional artists and AI had similar word counts, which were significantly greater than those generated by novice artists.

Semantic Distance

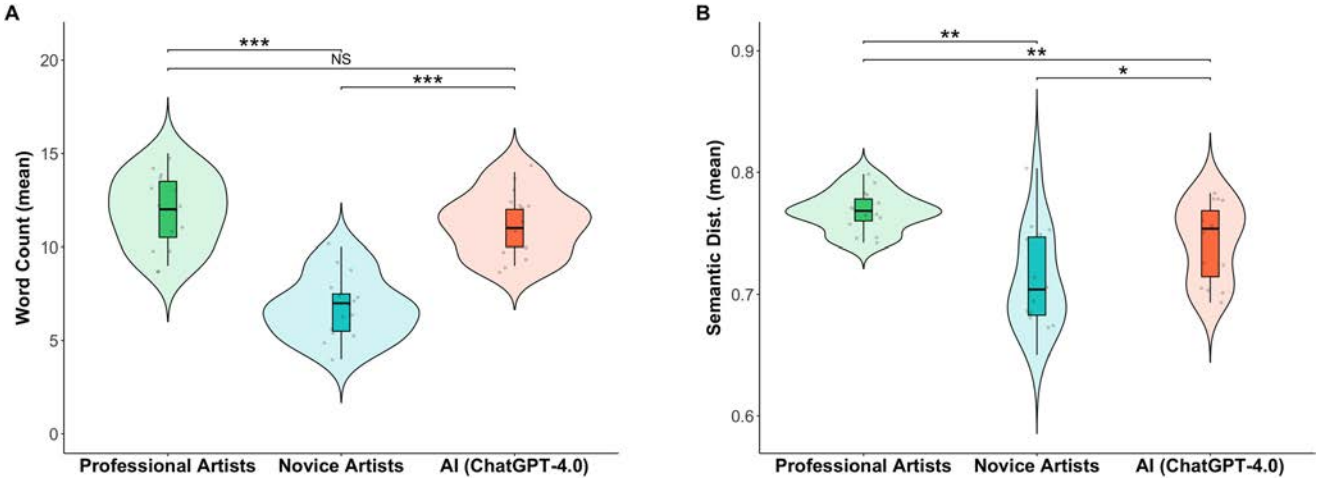
To further investigate potential factors influencing creativity ratings, we examined the semantic distance of the prompts generated by professional artists, novice artists, and AI. Previous research has identified semantic distance as a potential marker of creativity, with studies suggesting that responses containing a broader range of associations between diverse concepts tend to be associated with higher creativity ratings (Beaty & Johnson, 2021; Orwig et al., 2021). In addition, Orwig et al. (2024) found that the semantic distance of DALL-E 3 prompts partially mediated the relationship between individual creative ability (i.e., divergent thinking) and rated visual creativity of DALL-E 3 output. Thus, here, we might expect to find that professionals' prompts have the highest semantic distance, followed by AI, and then novice artists.

To measure semantic distance, we used the well-established divergent semantic integration (DSI) metric (Johnson et al., 2023). DSI

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 3

Mean Word Count and Semantic Distance Across Artist Groups



Note. Violin plots illustrating the mean word count (A) and mean semantic distance (B) for digital images generated by professional artists, novice artists, and AI (ChatGPT-4, $N = 45$). The median is indicated by the line within the box of each violin plot, and the IQR is represented by the box (25th to 75th percentile). The whiskers extend to the furthest data points within 1.5 times the IQR from the lower and upper quartiles. NS = not significant; Dist. = distance; AI = artificial intelligence; IQR = interquartile range. See the online article for the color version of this figure.
* $p < .05$. ** $p < .01$. *** $p < .001$.

scores were calculated using bidirectional encoder representations from transformers (Devlin et al., 2018), a bidirectional transformer-based language model that generates context-dependent word embeddings by assigning unique weights to each word based on its surrounding context. For each word prompt, bidirectional encoder representations from transformers produced embeddings that reflect the semantic relationships between words, capturing how much priority should be placed on each word relative to its context. We then computed DSI by taking the pairwise cosine semantic distances between all word embeddings. The range of DSI values is from 0 to 1, though most scores tend to fall between 0.70 and 0.90. Higher DSI scores indicate that the wordset connects more divergent concepts.

Having computed DSI scores for each wordsets, we then conducted a one-way ANOVA to examine potential differences in semantic distance across groups. Results showed a significant main effect of group on semantic distance, $F(2, 42) = 12.23$, $\eta^2 = 37$, $p < .001$, indicating that semantic distance varied significantly among the professional artists, novice artists, and AI-generated prompts (Figure 3b).

Next, we conducted three independent-samples t tests to compare DSI scores across the three groups. Professional artists' prompts had significantly higher DSI scores than those generated by AI, $t(28) = 2.76$, $p = .010$, $d = 1.01$, and by novice artists, $t(28) = 4.94$, $p < .001$, $d = 1.80$. Furthermore, AI-generated prompts showed higher DSI scores than those generated by novice artists, $t(28) = 2.34$, $p = .03$, $d = 0.85$. Thus, our analysis of DSI scores indicates that prompts generated by professional artists exhibited the highest semantic distance, followed by AI, with novice artists' prompts displaying the lowest semantic distance.

Correlational Analyses

To examine potential predictors of creativity ratings in greater detail, we conducted Pearson correlation analyses to assess the relationships between word count, semantic distance (DSI), and creativity ratings across all prompts. The results indicated a significant positive correlation between word count and creativity ratings, $r(43) = .37$, $p = .01$, suggesting that increases in prompt length tended to be associated with higher creativity ratings. Additionally, we observed a significant positive correlation between semantic distance (DSI) and creativity ratings, $r(43) = .55$, $p < .001$, suggesting that prompts with higher semantic distance tended to be associated with images that received higher creativity ratings.

Overall, these exploratory analyses suggest that both prompt length and semantic distance may play a role in the observed differences in creativity ratings across the groups. Although prompt length did not differ significantly between professional artists and AI, both groups had significantly longer prompts than novice artists. Prompt length was also significantly correlated with creativity ratings across the groups, indicating that longer prompts may help generate images perceived as more creative. Additionally, when examining semantic distance across the three groups, the pattern of results (professional artists > AI > novices) mirrored the pattern observed for creativity ratings. Furthermore, semantic distance was positively associated with creativity ratings across groups. Collectively, these findings suggest a potential role for both prompt length and semantic distance in enhancing the creative quality of AI-assisted artworks.

While informative, these exploratory results should be interpreted with some caution. Indeed, because these analyses were conducted post hoc, with a relatively small sample size ($N = 45$), there are concerns surrounding the reliability and generalizability of the findings, and they should therefore be viewed as preliminary. Nonetheless, they provide initial insights that could inform future research on the factors influencing creativity in AI-assisted artworks and offer a foundation for more robust, preregistered studies with larger sample sizes to validate and extend these findings.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Discussion

Within the realm of generative art, we find that professional artists, when assisted by AI, produce more creative digital images than AI working independently. However, this outcome was observed solely among professional artists: novice artists, who were also assisted by AI, were outperformed by AI alone. This indicates that while professional artists maintain a creative edge, AI has already surpassed less experienced artists in this domain. This pattern of results aligns well with Koivisto and Grassini's (2023) findings in the domain of divergent thinking, where AI outperformed humans on average but was surpassed by top performers. Similarly, in our study, AI generated higher rated creative outputs than novice artists—who plausibly reflect the average person's expertise in this domain—yet was outperformed by professional artists with relatively advanced expertise. Together, these findings suggest that while AI excels relative to the average individual, the contributions of top performers continue to hold unique value.

An intriguing question that arises from our findings is what, exactly, might have contributed to the different creativity ratings observed across the groups. Prior research suggests that semantic distance may play a key role, given its established positive correlation with AI-generated image creativity (Orwig et al., 2024). Consistent with this view, we observed that professional artists' prompts had the highest semantic distance, followed by those generated by AI, and then by novices—mirroring the pattern of creativity ratings across these groups. Moreover, and replicating Orwig et al. (2024), we also found a significant positive correlation between creativity ratings and semantic distance scores across the full sample, indicating that prompts with greater semantic distance tended to be associated with images rated as more creative. Thus, it appears plausible that differences in semantic distance across the three groups may help to explain the observed differences in creativity. However, future research is needed to more firmly establish this relationship.

In addition to examining the role of semantic distance, our exploratory analyses considered whether the word count of the prompt might also contribute to higher creativity ratings. Supporting this possibility, our findings revealed that both professional artists and AI-generated prompts with significantly higher word counts than novice artists, and word count was positively associated with creativity ratings across the sample. Although word count did not differentiate professional artists from AI, its positive correlation with creativity ratings suggests it may be a relevant factor in the observed differences in creativity ratings across groups. One plausible explanation for this relationship is that prompts with higher word counts provide additional context or detail, which could in turn guide the AI to generate richer, more complex images that may enhance their perceived creativity.

Another factor that may have contributed to the observed differences in creativity ratings across groups is the extent to which prompts conveyed emotional depth. While speculative, this idea aligns with findings by Messingschlager and Appel (2023), who showed that artworks perceived as higher in emotional content tended to be rated as more engaging and creative than those lower in such content. In the context of the present study, this raises the possibility that professional artists produced prompts with more nuanced emotional content than either novice artists or AI. Interestingly, if emotionality does indeed contribute to creativity ratings, our findings would suggest that AI may use more emotional content in its prompts than novice human artists, as AI-generated prompts received higher creativity ratings than those generated by novice artists.

Ultimately, a host of additional factors could potentially influence creativity ratings for AI-assisted artworks. For instance, conceptual coherence (a clear underlying theme or message), specificity of visual descriptions (providing detailed elements that guide the AI toward more intricate imagery), or a focus on existential themes (such as purpose, mortality, freedom, and isolation) may resonate more deeply with viewers, thereby enhancing the perceived creative value of the artworks. These and other possibilities open exciting avenues for future research to better understand how prompt characteristics shape the creative outcomes of AI-assisted collaborations.

Reflecting more broadly on the current findings, had AI demonstrated a superior ability to generate creative artwork, this outcome would have posed significant challenges for artists at risk of displacement by autonomous AI art generators. However, the fact that professional artists collaborating with AI maintained an advantage over AI working independently supports the continued, vital role of human creativity and artistic expertise in the domain of generative AI. For now, at least, in these creative domains, our results align with Guilford's (1950) early prediction that uniquely human skills and expertise would continue to hold value, even as technology advances. Importantly, these findings are relevant to discussions surrounding the future of jobs (e.g., World Economic Forum, 2020), skills that are resistant to automation (e.g., Kosslyn, 2019), and imperatives to adapt higher education to changing technology (Barr et al., 2022). Indeed, our findings suggest that nurturing human creativity and artistic expertise remains crucial in a time increasingly dominated by generative AI systems.

While our results are encouraging for those advocating for a unique place for humans in the artistic domain, historical developments remind us that technological advancements have sometimes transformed human involvement from an asset to a hindrance in performance. A notable example is the evolution of AI in chess. Initially, collaborations between humans and AI, known as "Centaur Chess," were highly successful, combining human intuition with AI's computational power to outperform either working alone (Case, 2018). However, as AI systems like DeepMind's AlphaZero advanced, they ultimately surpassed both humans and human–AI teams (Silver et al., 2018). Bearing this in mind, we should not lose sight of the possibility that, as generative AI continues to evolve, the landscape of digital art may likewise shift, possibly revealing a future where AI alone outperforms human–AI collaborations in generating creative digital images.

Notwithstanding this possibility, it is heartening to note that some recent studies have found that human raters tend to prefer artworks that ostensibly involve human participation—whether labeled as being created solely by humans (Bellaiche, Shahi, et al., 2023; Bellaiche, Smith, et al., 2023) or in collaboration with AI (Horton et al., 2023)—over those produced by AI alone, a preference termed the "anti-AI bias." This bias suggests that if participants in our study were informed of human involvement in the creation of the artworks they rated, such pieces might have garnered more favorable judgments of creativity. Consequently, the superior creativity ratings we observed for professional artists compared to AI alone may have been further amplified. Additionally, artworks created by novice artists might have received higher creativity assessments than those generated solely by AI. Hence, despite the imminent artistic advancements that will be made by AI, the anti-AI bias may

continue to favor artworks that include human contributions, positioning them as more creatively valuable than those produced solely by the increasingly sophisticated AI.

Finally, it is worth noting that the artists who contributed to our stimulus set—both professionals and novices—reported no prior experience with generative art models like DALL-E 3. Furthermore, these contributors did not have the opportunity to view the images generated from their text prompts, nor could they engage in an iterative process to refine their prompts based on initial visual outputs. Thus, if future studies were to explore performance among artists with substantial experience using generative art models, and if these artists were allowed to iteratively adjust their prompts, we anticipate that the creativity ratings of their works would be even higher than those observed in the present study. This combination of increased familiarity with generative AI technology and iterative refinement of artistic inputs could substantially amplify the perceived creativity and value of the artworks, further solidifying the role of human creativity in the era of advanced AI.

Conclusion

For generations, humans have offloaded numerous tasks to technology, extending the "scope of our sense organs, our brains or our limbs" through devices such as "telescopes and microscopes, wireless, calculating machines, typewriters, motor-cars, ships and aeroplanes" (Craik, 1943, p. 61). In recent years, more advanced cognitive functions have been delegated to increasingly sophisticated technologies (e.g., smartphones; Barr et al., 2015), with the latest developments in generative AI providing novel capabilities for transferring aspects, or even the entirety of the creative and artistic process to machines. As the evolution of these capabilities marches on, the "continuing development of aesthetically sensitive machines will challenge our notions of beauty, creativity, and the nature of art" (Chatterjee, 2022, p. 1), and will redefine the role of humans in the creative and artistic process. As we continue to witness technological advancements, it is imperative for humanity to (a) assess whether and how we can most effectively leverage these innovations to push the boundaries of creativity, (b) gauge how human artistic expertise interacts with novel modes of creation, and (c) determine whether collaboration with AI enhances aspects of creative production in ways that autonomous AI alone may not achieve. Ultimately, the ongoing synergy between human creativity and machine intelligence promises to shape the future of art and, in many ways, redefine the creative process.

References

Barr, N., Hartley, K., Lopata, J. A., McFarlane, B., & Mcnamara, M. J. (2022). Learning in an uncertain world: Transforming higher education for the anthropocene. In R. A. Beghetto & G. J. Jaeger (Eds.), *Uncertainty: A catalyst for creativity, learning and development* (pp. 337–357). Springer International Publishing. https://doi.org/10.1007/978-3-030-98729-9_18

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that smartphones are used to supplant thinking. *Computers in Human Behavior*, *48*, 473–480. https://doi.org/10.1016/j.chb.2015.02.029

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, *53*(2), 757–780. https://doi.org/10.3758/s13428-020-01453-w

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications*, *8*(1), Article 42. https://doi.org/10.1186/s41235-023-00499-6

Bellaiche, L., Smith, A. P., Barr, N., Christensen, A., Williams, C., Ragnhildstveit, A., Schooler, J., Beaty, R., Chatterjee, A., & Seli, P. (2023). Back to the basics: Abstract painting as an index of creativity. *Creativity Research Journal*, *35*(4), 698–713. https://doi.org/10.1080/10400419.2023.2243100

Bullis, E. (2024, November 1). Artist challenges copyright office refusal to register award-winning AI-assisted work. *Law Journal Newsletters*. https://www.lawjournalnewsletters.com/2024/11/01/artist-challenges-copyright-office-refusal-to-register-award-winning-ai-assisted-work/?srsltid=AfmBOooIkwAQhH1-epb6YBIxyBZ_DN8sRKmBMps388CA6QjBEp_tmplP&slreturn=20241101142415

Buttazzo, G. (2023). Rise of artificial general intelligence: Risks and opportunities. *Frontiers in Artificial Intelligence*, *6*, Article 1226990. https://doi.org/10.3389/frai.2023.1226990

Case, N. (2018). How to become a centaur. *Journal of Design and Science*. https://doi.org/10.21428/61b2215c

Chatterjee, A. (2022). Art in an age of artificial intelligence. *Frontiers in Psychology*, *13*, Article 1024449. https://doi.org/10.3389/fpsyg.2022.1024449

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Craik, K. J. W. (1943). *The nature of explanation*. CUP Archive.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805.

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*(9), 444–454. https://doi.org/10.1037/h0063487

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, *1*(1), 3–14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x

Heckel, J. (2023, December 15). *Design students' use of AI in projects teaches career skills*. University of Illinois News Bureau. https://blogs.illinois.edu/view/6367/436842224

Horton, C. B., Jr., White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific Reports*, *13*(1), Article 19001. https://doi.org/10.1038/s41598-023-45202-3

Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, *14*(1), Article 3440. https://doi.org/10.1038/s41598-024-53303-w

Johnson, D. R., Kaufman, J. C., Baker, B. S., Patterson, J. D., Barbot, B., Green, A. E., van Hell, J., Kennedy, E., Sullivan, G. F., Taylor, C. L., Ward, T., & Beaty, R. E. (2023). Divergent semantic integration (DSI): Extracting creativity from narratives with distributional semantic modeling. *Behavior Research Methods*, *55*(7), 3726–3759. https://doi.org/10.3758/s13428-022-01986-2

Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific Reports*, *13*(1), Article 13601. https://doi.org/10.1038/s41598-023-40858-3

Kosslyn, S. M. (2019, September). Are you developing skills that won't be automated? *Harvard Business Review*.

Krietzberg, I. (2023, December 23). Human creativity persists in the era of generative AI. *TheStreet*. Retrieved from https://www.thestreet.com/technology/human-creativity-persists-era-of-generative-artificial-intelligence

Messingschlager, T. V., & Appel, M. (2023). Mind ascribed to AI and the appreciation of AI-generated art. *New Media & Society*. Advance online publicaiton. https://doi.org/10.1177/14614448231200248

Orwig, W., Bellaiche, L., Spooner, S., Vo, A., Baig, Z., Ragnhildstveit, A., Schacter, D. L., Barr, N., & Seli, P. (2024). Using AI to generate visual

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

art: Do individual differences in creativity predict AI-assisted art quality? *Creativity Research Journal*. Advance online publication. https://doi.org/10.1080/10400419.2024.2440691

Orwig, W., Diez, I., Vannini, P., Beaty, R., & Sepulcre, J. (2021). Creative connections: Computational semantic distance captures individual creativity and resting-state functional connectivity. *Journal of Cognitive Neuroscience*, *33*(3), 499–509. https://doi.org/10.1162/jocn_a_01658

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, *24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140–1144. https://doi.org/10.1126/science.aar6404

The Coca-Cola Company. (2023, August 22). *Coca-Cola hosts digital artists and tech enthusiasts for generative AI summit*. Coca-Cola Company. https://www.coca-colacompany.com/media-center/coca-cola-invites-digital-artists-to-create-real-magic-using-new-ai-platform

World Economic Forum. (2020). *The future of jobs report*. WEF.

Received July 16, 2024
Revision received November 2, 2024
Accepted December 5, 2024 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.