

Creative Artificial Intelligence: Exploring the Qualities of Popular AI Art Tools to Determine Effectiveness

Ravia Bhullar

Received May 16, 2024

Accepted July 31, 2024

Electronic access August 15, 2024

Research into the effectiveness of AI art tools suggests that generated images are becoming increasingly indistinguishable from human-made artworks, yet perceived value has generated inconclusive results. This study implements a correlational method to explore the relationship between generated images from different AI tools and their perceived value and distinguishability by comparing them to human-made artworks with similar artistic styles. Value, in the given context, is defined as aesthetic appreciation, monetary worth, and perceived meaningfulness. Results imply that expensive tools provided the most human-like, though not the most valued images as they generated works that were most often mistaken as human artworks. Perceived value among different AI tools was similar and created considerable overlap between ratings for each tool. The average perceived value was higher for human artwork than generated works for all measured variables, though no significant difference was found. The provided results inspire further research into the effectiveness of AI tools. As AI art tools continue to develop to create more human-like images with greater perceived value, the current state of traditional art is threatened as some believe AI art may devalue human art. Regardless, the evolution of art through technology can break the current barriers of traditional art by allowing new degrees of freedom for digital artists.

Introduction

With no signs of slowing down, Generative AI is rapidly gaining traction in various fields. From the covers of famous musicians to poetry rivaling the works of Shakespeare¹, the products of generative AI are seemingly unheard of outside of human works until now. Yet, the development of AI is not as new as some may believe. AI development can be traced as far back as the 1950s when the Turing Test was first introduced by computer scientist Alan Turing. The Turing Test is a method used throughout the computing field to test whether a machine can imitate the qualities of a human well enough to be indistinguishable when compared to one². Since then, AI development has experienced many milestones leading to where it is today, one of the most notable being in 2016, when Google's AI-driven system, AlphaGo, beat the human grandmaster in the complex board game Go³.

With the common goal of maximizing the intellectual productivity of AI technology⁴, AI development has always prioritized generating products outside the realm of human capabilities. The public perception of these capabilities is limited, disregarding the creative abilities of AI in various fields. While AI has arguably mastered the art of problem-solving and logical reasoning, it is still too early to conclude whether it has even begun to rival the intrinsic human skill of creativity. Nevertheless, generative AI is present in various creative fields, including fashion, music, and the visual arts¹⁻⁴. The visual arts are one of the farthest along in development, with tools built specifically for

image creation. The field has long relied on the traditional skills of artists before the introduction of AI art, yet now criticism has also been cast upon AI art tools on whether their capabilities rival or even surpass the abilities of human artists. Therefore, exploring the effectiveness of several generative AI art tools through the distinguishability and value (aesthetics, meanings, and monetary worth) of generated images in comparison to traditional human artworks will provide further knowledge regarding the creative skills of generative AI.

Literature Review

In every field in which AI has grown a presence, there lie varying perspectives regarding its use, including the visual arts. With the rise in generative AI tools that use text inputs to generate corresponding images such as Stable Diffusion, Dalle, and Midjourney, fears have risen regarding the "death of the artist"⁵. However, many also believe that generative AI is an artistic tool for an artist rather than a replacement. Photography faced a similar adjustment, inflicting fear from artists when introduced, now photos have allowed for new aspects of artwork and creativity rather than a replacement. With the enhanced features of AI tools, artists allow themselves more room in the creative process as their boundaries have expanded^{3,4,6}. Even so, AI art currently lacks the intentionality and consciousness that human artists have, suppressing expression⁷. Further development is necessary for neural networks of generative AI as it will aid in

AI coexisting among creative developers⁸. Neural networks are the methods in AI that train machines to process data similar to how humans think^{1,3-8}. As a result of improving neural network models, AI art tools can provide more use in a given field as they adapt to human skill sets. For example, AI can help with art comprehension and appreciation, though it is currently considered a human capability⁹. Regardless, the majority of research focuses on image quality and representation. As AI continues to obtain human capabilities, it also embodies limitations such as (lack of) culture and aesthetics, authorship, labor economics of creative work, and impacts on media⁶. While these limitations are equally vital to address, bias is prominent. Bias includes a lack of racial, gender, and ethnic representations and stereotypes resulting from the images that an AI model is trained on^{6,10-13}. In addition to societal limitations, AI artwork is also limited in the quality of perception it receives, consequently, researchers tend to focus on measuring different perceptions of the generated artworks.

Conclusions vary in research focused on human responses to AI art. Some studies conclude that human art is valued more than AI-made art, while others deduce humans do not recognize the difference between AI and human art¹⁴. Variability in results is thanks to the format of the studies done in the field. A study published in *Cognitive Research: Principles and Implications* conducted an experiment where participants were randomly assigned to view AI-generated or “human-made” artworks, though both groups viewed AI artwork and answered questions regarding their views on the piece. Results emphasized the participants preferred human-made art over AI-created art due to perceived meaningfulness¹⁵, contesting the beliefs of⁵ as it illustrates AI art is unlikely to replace artists due to the value humans place on meaning in art. A study published in the *Applied Sciences Journal* supports the results of¹⁵. Authors found that aesthetic appreciation is more significant for human paintings due to greater emotional depth¹⁶. Both studies imply that while generative AI is typically effective in technical fields, it tends to fall behind in areas that involve creativity. Another study, published in the *National Center of Biotechnology Information*, investigates whether knowing if an artwork was human-made or generated by AI affects opinion of it. They found that art experts favored human-made art over AI-generated art concerning purchase and collection intentions. Meanwhile, the participants without professional art experience had no preference for either AI or human-made artwork, focusing on the style and aesthetics of the artwork itself¹⁷. These findings are contradictory to those in¹⁵, where they found that knowing whether the artwork was created by humans or AI affected views on it, regardless of their professional art experience. It can be argued that the findings in¹⁶ are in line with those in¹⁷ as 70% of participants had art experience and the results depicted that participants favored human-made paintings. Overall, it is unclear whether the findings on the value of AI artworks are generalizable given that

disclosing the origin of an artwork affects a participant’s view of it, hence the importance of incorporating distinguishability into the study¹⁵⁻¹⁷.

Aside from aesthetic appreciation and favorability, emotional depth and meaningfulness are equally significant in research. Researchers at Northwestern University discuss the need to treat AI art similarly to traditional human-creating art by partaking in the same aesthetic judgments when interacting with AI-generated images. With traditional artwork, historians and artists evaluate modern and contemporary works with similar themes in mind, themes which are often ignored when analyzing AI-generated artworks due to the common belief that AI-created images cannot exhibit intrinsic human traits¹⁸. Bai Liu, a researcher at the Winchester School of Art University of Southampton, explores the artistic and aesthetic value of AI artworks by interpreting the portrayal of four areas of interest: creativity, motivation, self-awareness, and emotion. Lui’s findings depict that AI can synthetically represent emotions through generated works¹⁹, which aligns with conclusions made in a study analyzing AI art emotional depth published in the *Journal of Computers in Human Behavior*. The study found that participants reported a presence of emotional value to the computer-generated artworks, though there was still more portrayed through human-generated artworks. Then, participants speculated what artworks were human-made and which were computer-generated. Researchers found that the mean of correct guesses was 63.8%, showing that participants distinguish between the two types of art²⁰. These results^{19,20} suggest that AI artwork is capable of displaying human emotion, which contests the assumptions and results of¹⁵⁻¹⁷ where most participants found a lack of emotional depth in AI works. Similar to why the results of studies measuring perceived value varied, the difference in results can be attributed to the researcher’s choice of disclosing the origin of an image before participants view it. Participants may have positive or negative biases toward AI art, affecting their perception of an image’s aesthetic, monetary, and emotional value, further reinforcing the need to measure distinguishability alongside value.

As research continues to emerge in the field of generative AI art, a trend is clear regarding the priorities of many researchers: measuring the perceived value of artworks after disclosing the origin. Unlike literature, where the Turing test is a popular method to test whether an AI-generated product is comparable to a human-created one², AI art research prioritizes the impact of origin on one’s perception of the artwork. As a result of the disregard for a distinguishability test, studies measuring the perception of AI-generated artworks¹⁵⁻²⁰ had different results. While one²⁰ did address this discrepancy, the structure of the study worked in favor of participants identifying the origin of each image. The subjects were already shown each artwork along with the origin in the first half of the study; when it came down to the distinguishability test, they were presented with the same images in randomized order without a label. Subjects may

have remembered or recognized patterns among AI-generated and human-created artworks in the first half of the study, aiding them in the distinguishability test.

With the intent of addressing the discrepancy regarding distinguishability and value emerges the research question: To what extent can AI art tools create artwork indistinguishable from and as valuable as human artwork, and how do variables differ by specific AI tool? Exploring the efficiency of different AI art tools through the research question allows for discoveries regarding the effectiveness of AI art tools, and it focuses on measuring value while keeping the image origin confidential. Throughout a handful of sources, the method employed to test distinguishability²⁰ or value^{15–20} only utilized a single AI art tool to provide the images. Assumably, the AI tools used to generate an image do not play impactful roles in the outcomes. If the findings are consistent with prior research in the same field, AI artwork will be distinguishable from human-created artwork while exhibiting similar perceived value.

Regardless of the outcome, answering the question provides real-world significance as it clarifies the current state of AI technology in visual arts. AI art can offer new perspectives and insights into the creative process, allowing artists to explore new possibilities. While it is unclear whether AI-created art can replicate the emotional depth of traditional art, if AI art tools can imitate the styles and themes of human artworks, then it has the potential to revolutionize the art world and expand creative possibilities.

Methodology

A correlational method design was implemented to address the research question, looking at the relationship between the variables: image origin, distinguishability, and value (aesthetic, monetary, and meaningfulness), though not accounting for causality. The method was composed of different aspects of previous research designs concerning generative AI art exploring specific aspects of the technology. It compared AI-generated images to human-made images, had participants identify the origin of the works, and judged the aesthetics, monetary value, and perceived meaningfulness in all the images provided^{15–17,19,20}. The aspects chosen to incorporate were meant to provide the most suitable setting to produce results that answer the question at hand as their corresponding studies explored similar questions regarding the effectiveness of AI art tools.

Exploring the effectiveness of AI art helped to achieve the purpose of assessing the current state of AI art tools as it allowed for open interpretation of results, as well as freedom regarding the types of AI tools and genre of artworks incorporated. A correlational method design was most appropriate as it allowed for the investigation of relationships between artistic perception and image origin which is not clearly defined, making room for exploratory research. Quantitative data was collected when

measuring perceived value as respondents were asked to rate each aspect of value on a 1-5 scale for each image, used to ensure that there was a common scale for all subjects to rate the value of the image. Qualitative data came from having participants distinguish between human art and AI art, allowing for a simple binary output. Responses were collected through surveys, consisting of multiple-choice questions and optional short-answer feedback. A multiple-choice survey was the most effective method as it ensured that the data collected was consistent and provided the measures necessary to analyze later on. Respondents were high school students in northeast Ohio with a broad range of experience in both art and AI tools, 169 responses were collected overall.

In the survey, subjects identified whether an image was AI-generated and then interpreted the value of each image. In terms of collecting the images, first, three major generative AI tools were selected to generate the images: DALL-E, Midjourney, and Stable Diffusion. Next three artists were chosen from the public domain of the Metropolitan Museum of Art database. The three artists were George Inness, William Michael Harnett, and Eastman Johnson, all artists with distinctive art styles from similar periods. For each artist, three artworks were selected to be representative of their work. For George Inness the works were as follows, Peace and Plenty, Evening at Medfield, Massachusetts, and Delaware Water Gap. All of the selected works by Inness were landscape paintings, therefore the prompt for the first set of AI-generated images was “Landscape painting in the style of George Inness, visible brushstrokes, tonalism style.” For William Michael Harnett the images were as follows, The Banker’s Table, New York Daily News, and Still Life—Violin and Music. All of the selected works by Harnett were still-life paintings, therefore the prompt for the second set of AI-generated images was “Still-life painting on canvas in the style of William Michael Harnett.” For Eastman Johnson, the images were as follows, The New Bonnet, The Funding Bill, and The Hatch Family. All of the selected works by Johnson were conversational paintings, therefore the prompt for the final set of AI-generated images was “Painting of conversation between several people, 1800s, inspired by Eastman Johnson.” Next, the prompt for each artist was input into each of the three generative tools in the order that follows, Stable Diffusion, Dalle, and Midjourney. For tools that produced several images (Dalle and Midjourney produce four images per prompt), a random number generator was used to select which image would be selected. At the end of this process of generation, 9 AI-generated images and 9 human-made images were collected. Next, each image was placed into a Google spreadsheet and labeled accordingly by the origin, name, copyright, imputed prompt, time to generate, and category. Then, the Google form was created, which first contained prefacing questions regarding the subject’s experience within art and AI tools, and demographic questions to ensure that there was a diverse group of subjects (age and gender). For

Image	Origin	Name	Copyright	Prompt	Time (sec)	Category	Date
	Stable Diffus	NA	NA	"landscape painting in the style of George Inness, visible brushstrokes, tonalism style"	56	Landscape	Oct 28, 2023
	Dalle	NA	NA	"landscape painting in the style of George Inness, visible brushstrokes, tonalism style"	16	Landscape	Oct 28, 2023
	Midjourney	NA	NA	"landscape painting in the style of George Inness, visible brushstrokes, tonalism style"	60	Landscape	Oct 28, 2023
	Human	Peace and Plenty, George Inness	Open access	NA	NA	Landscape	Oct 28, 2023
	Human	Evening at Medfield, Massachusetts	Open access	NA	NA	Landscape	Oct 28, 2023
	Human	Delaware Water Gap	Open access	NA	NA	Landscape	Oct 28, 2023
	Stable Diffus	NA	NA	"Still-life painting on canvas in the style of William Michael Harnett"	46	Still-life	Oct 28, 2023
	Dalle	NA	NA	"Still-life painting on canvas in the style of William Michael Harnett"	11	Still-life	Oct 28, 2023
	Midjourney	NA	NA	"Still-life painting on canvas in the style of William Michael Harnett"	41	Still-life	Oct 28, 2023
	Human	The Banker's Table	Open access	NA	NA	Still-life	Oct 28, 2023
	Human	New York Daily News	Open access	NA	NA	Still-life	Oct 28, 2023
	Human	Still Life—Violin and Music	Open access	NA	NA	Still-life	Oct 28, 2023
	Stable Diffus	NA	NA	"Painting of conversation between several people, 1800s, inspired by Eastman Johnson"	16	Conversation	Oct 28, 2023
	Dalle	NA	NA	"Painting of conversation between several people, 1800s, inspired by Eastman Johnson"	11	Conversation	Oct 28, 2023
	Midjourney	NA	NA	"Painting of conversation between several people, 1800s, inspired by Eastman Johnson"	48	Conversation	Oct 28, 2023
	Human	The New Bonnet	Open access	NA	NA	Conversation	Oct 28, 2023
	Human	The Funding Bill	Open access	NA	NA	Conversation	Oct 28, 2023
	Human	The Hatch Family	Open access	NA	NA	Conversation	Oct 28, 2023

Fig. 1 Google Spreadsheet for images

each image, 4 questions were asked: “Is this image AI-generated or human-created” and “Rank the [meaningfulness, aesthetics, or monetary worth] of the presented image on a scale of 1-5. 5 is most, 1 is least.” The images were placed in random order, therefore the prompt of the image did not affect its position. At the end of the form, the subjects were prompted to give feedback regarding their predicted accuracy, tactic of identifying images, and final perspective on AI art. To ensure that the generated images are as similar (themed) as possible, the same prompt was input for each AI tool. Additionally, to control for any confounding variables such as prior artistic experience and bias regarding AI art, subjects were asked prefacing questions regarding artistic and AI knowledge to use if blocks are needed later on when analyzing data.

All data was collected through surveys by Google Forms and then stored in Google Sheets, which helped interpret each response variable as a combination of all respondents because the two tools automatically transfer information. As images were generated and chosen, they were placed into a spreadsheet to ensure that the origin of each image was not lost anywhere during the process of data collection. Securing image origin was vital as it aided in the analysis of survey responses to ensure that

human-made images were not analyzed as AI-generated images and vice versa. Once data had been collected, the mean and standard deviation of value ratings (1-5) were found for each type of image. Analyzing the data in this matter allowed for comparison between the image origins and easier interpretation of responses. Due to the extensive process of image generation, comparison between blocks was justified because all AI tools were given prompts in similar structures. To interpret the distinguishability the proportion of correct identifications was calculated, allowing for a better understanding of how similar AI artworks are to human-made artworks. Both methods of analyzing the distinguishability and value of AI-created works allowed for a better understanding of the effectiveness of the tools utilized.

Results

Distinguishability was presented through frequency correctly and incorrectly guessed allowing for easy comparison between AI-generated and human-made artworks as the sample size was the same for each block. Next, z tests were utilized to find evi-



Fig. 2 Evening at Medfield, Massachusetts, George Inness (top); Dalle generated artwork with prompt “Landscape painting in the style of George Inness, visible brushstrokes, tonalism style.” (bottom)

dence supporting that participants would mistake an AI image for human more likely than mistake a human image for AI and whether the proportion of correct guesses for AI images was 0.5. Additionally, variables of value presented through average ratings were best fit as they efficiently described the center of each distribution (given there were no outliers) and allowed for easy comparison between means. Accordingly, a t test for a difference of means was utilized to find supporting evidence that human-made images are valued more than AI-generated images. For the z test, the large counts rule was met as there were greater than 10 successes and 10 failures in both samples. For the t test, all sample data lack apparent outliers or skewness in distribution, allowing for the use of the significance test. Having a sample size of 169 allowed for a diverse sample of respondents which also varied in AI and art experience (as provided by prefacing questions). Furthermore, the analysis of the study aligned with that of¹⁷ where significance tests were used to measure ratings, purchase intention, and collection intention. Utilizing a correlational method allowed for the best collection of data

27. Where do you believe this image originates? *
- Mark only one oval.
- Generative AI tool
- Human Artist
28. Does this image look nice to look at? Rate the aesthetic value. 1 is very unappealing, 5 is very appealing *
- Mark only one oval.
- 1 2 3 4 5
-
29. How meaningful do you find this image? 1 is meaningless, 5 is very meaningful *
- Mark only one oval.
- 1 2 3 4 5
-
30. How much money do you believe this image should be worth? 1 is least, 5 is most *
- Mark only one oval.
- 1 2 3 4 5
-
81. What was the largest signifier that an image was AI? (Optional)

Fig. 3 Sample of survey questions; questions 27-30 were given for all 18 images; question 81 was asked once at end of survey

as it allowed for insight into different relationships such as that between AI tool price and distinguishability. Furthermore, analysis of existing patterns was explored as interference within the process of data collection was unnecessary. The type of data collected also gave room to analyze the correlation between image origin and specific variables concerning value.

Table 1 Value among Human and AI artworks (Mean ± Standard Deviation)

Origin	Aesthetic Rating (M ± SD)	Meaningfulness Rating (M ± SD)	Monetary Rating (M ± SD)
Human	3.395 ± 1.027	2.991 ± 1.075	3.202 ± 1.050
AI	3.294 ± 1.133	2.803 ± 1.147	2.980 ± 1.147

Trends in the data depict that some tools appeared to produce images comparable to human works better than others. Table 4 shows that images from Stable Diffusion were correctly identified over half as much as images produced by Dalle. Furthermore, Table 6 illustrates that when image origin was incorrectly guessed by the majority, it was more likely that the respondent perceived an AI image as a human-made image rather than vice

Table 2 Value among AI tools (Mean ± Standard Deviation)

AI Tool	Aesthetic Rating (M ± SD)	Meaningfulness Rating (M ± SD)	Monetary Rating (M ± SD)
Stable Diffusion	3.274 ± 1.138	2.665 ± 1.108	2.854 ± 1.136
Midjourney	3.578 ± 1.013	2.998 ± 1.162	3.250 ± 1.081
Dalle	3.03 ± 1.175	2.748 ± 1.147	2.834 ± 1.175

Table 3 Frequency of Correct and Incorrect Image Identification for Human and AI (number of images = 1521)

Origin	Frequency Correctly Identified	Frequency Incorrectly Identified
Human	68.2%	31.8%
AI	63%	37%

Table 4 Frequency of Correct and Incorrect Image Identification for Specific AI Tools (number of images per tool = 507)

AI Tool	Frequency Correctly Identified	Frequency Incorrectly Identified
Stable Diffusion	84.2%	15.8%
Midjourney	63.3%	36.7%
Dalle	41.4%	58.6%

versa. For value, average aesthetic, meaningfulness, and monetary ratings were greatest for human images, as shown in Table 1. Table 2 then goes to show how Midjourney performed the best in value within the specific tools, followed by Stable Diffusion, and lastly Dalle. This outcome was surprising given that Dalle provided the images most comparable to human artworks. While not depicted, when answering the optional free-response question regarding how they were able to distinguish between AI and human-generated images, responses varied though common claims included: image texture, construction of human features (face and hands), and the placement of objects.

Discussion

In concern to the inquiry, AI tools are at times indistinguishable from human artworks and are mistaken for human artworks more often than human artworks are mistaken for AI work. Furthermore, there is no significant difference in the perceived value of AI and human artwork, given the large p-values for each t test done (Table 5). Results presented¹, which examined the effectiveness of AI literature, found that for human poems the proportion of correct guesses was 51.4% while for AI-generated poems the portion of correct guesses was only 46.2%, aligning with the results of this project as it was evident that the propor-

tion of correct guesses for human artworks (68.2%) was greater than than of generated artworks (63%), as supported by Table 3.

Though distinguishability results were not entirely in favor of indistinguishable AI works, as the one sample z test provided significant results with evidence supporting that proportion of correct and incorrect guesses was not evenly split. The results would be evenly split (0.5 correct and 0.5 incorrect), if participants randomly guessed the origin, however the z test suggested that there is \neq 0.5 proportion of correct guesses (Table 6). Limitations may arise due to the restricted population of participants, as only high schools from northeast Ohio were surveyed, which may not be representative of the general population of high schoolers. Furthermore, age is a limiting factor in distinguishing between AI and human tools as the method design only accounted for high school aged respondents, the age range was limited to 14 to 18 years old. Teenagers are oftentimes more familiar with AI images due to the use of AI design in social media platforms. This confounding was accounted for by asking prefacing questions in the survey, such as how familiar they are with AI art, the majority of respondents have never used AI art tools and are more familiar with AI writing tools, which could be the reason for the distinguishability of the AI images. For specific AI tools a significance test was not applied to avoid false significant results due to multiple tests. Even so, simply comparing the proportion of incorrect and correct guesses provides an effective perspective regarding the distinguishability of the AI tools utilized. Dalle proved to provide the most human-like images with a frequency of 41.4% correct guesses, indicating that respondents more often than not believe that an image generated by Dalle was actually human-created. Stable Diffusion provided the least human-like images with a frequency of 84.2% correct guesses, more than double the proportion from Dalle. Images produced by Midjourney were in the center in terms of distinguishability with a frequency of 63.3% correct guesses. These trends in distinguishability also align with the expense of the respective AI tool, as Dalle was the most expensive while Stable Diffusion was free. Price could also be a significant limitation as more expensive AI tools seem to provide better images in terms of comparability to human artworks. Distinguishability results were similar to those²⁰, where researchers found that the mean of correct guesses was 63.8%, comparable to the 68.2% for human images and 63% for AI images. Furthermore, as the method design utilized²⁰ was similar to the presented project, as both tested distinguishability and value concurrently, the re-

Table 5 Two Sample T Test for Difference of Mean Between Human (μ_1) and AI (μ_2) where $\alpha = 0.05$

Variable	H_0 (null hypothesis)	H_a (alternative hypothesis)	Test Statistic	P-value
Aesthetic	$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$t = 0.180$	$p = 0.430$
Meaningfulness	$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$t = 0.359$	$p = 0.362$
Monetary	$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$t = 0.428$	$p = 0.337$

Table 6 Z Test for Proportions where $\alpha = 0.05$

Test	H_0 (null hypothesis)	H_a (alternative hypothesis)	Test Statistic	P-value
Two sample Z test for difference of proportions ($p_1 =$ proportion of correctly guessed human images; $p_2 =$ proportion of correctly guessed AI images)	$p_1 = p_2$	$p_1 > p_2$	$z = 3.05$	$p = 0.001$
One sample Z test for proportions ($p =$ proportion of correctly guessed AI images)	$p = 50$	$p > 50$	$z = 10.1$	$p = 2.11 \times 10^{-24}$

sults²⁰ supported that image origin influenced the perception of an artwork, specifically in terms of evoked emotion. The difference between results can be explained by the different variables being measured as well as method design, used an experimental method instead of correlational²⁰.

In terms of value, human images received a higher average rating for all three categories of value, though no significant difference was found in the rating for human vs the ratings for AI images (Table 5). The lack of difference could be due to the limited scale of rating, as the participants were only given a range of 1 to 5, alternately it could be due to the comparable value of AI generated images, though further research would be necessary to support that claim. Midjourney produced the images with the highest average ratings for all variables of value in comparison to the other AI tools. Stable Diffusion revived greater average ratings for both aesthetic and monetary variables over Dalle, however Dalle received greater average rating for meaningfulness. In a study¹⁷ the ratings for evaluations, purchase intention, and collection intention toward paintings were greater for human images than AI images, however the results were not significantly significant, supporting the results of this study as well.

For both distinguishability and value, results are strongly influenced by the structure of the prompts input into the AI tools utilized. Due to the design of different AI art tools, there are certain phrases and sentence structures that produce “better” images than others, and the ones implemented in the design when producing the images for the surveys may not have been the most effective. Consequently, the images produced may

not be representative of the best images formed by the selected tools. Additionally, all artworks chosen were of similar time periods, not accounting for the nuances between human-made art throughout generations. Increasing the number of artists as well as the types of artworks would result in more conclusive results. Nevertheless, research into the effectiveness of AI art tools progresses the current body of knowledge present regarding AI development as it discusses the current limitations of popular AI tools while also providing supporting evidence for the efficiency of current tools. With further research, AI tools can continue to improve their precision and accuracy to serve a greater purpose in the arts.

Conclusion

The results of the study provide inspiration for further research into the effectiveness of AI art tools, specifically regarding the value of generated works. While evidence was found to support that individuals in the study were able to differentiate between AI and human artworks, results may be limited to the efficiency of prompts in AI tools and the era of artwork chosen to compare. Even so, the gap of information regarding the effectiveness of several AI tools with comparison to human artworks was addressed as evidence was found supporting the greater indistinguishability of Dalle-generated images and the greater perceived value of Midjourney-generated images. Furthermore, participants were more likely to mistake AI images as human than vice versa, implying that with further research and progress, AI art tools may become indistinguishable from human-made works.

Midjourney performed the best in terms of perceived value, then followed by Stable Diffusion and Dalle, respectively. Human images performed better in all variables of value measured, however there was not a significant difference for any of the mean ratings. Utilizing an experimental method to block by image origin more efficiently may provide better results to infer causation regarding the relationship between image origin and the measured variable. Regardless of limiting factors of inference, the research provided further knowledge into the effectiveness of several AI tools and allowed room for improvement. As AI works become indistinguishable from human works, the originality of works is threatened as AI algorithms must use pre-existing datasets to create images. Perceived value remains one of the key variables in determining the success of AI artworks, therefore if further research can support that AI images are comparable in value, there could be negative implications for traditional artists as creating AI images may be a more affordable option. With the evolution of AI, continuous research will be necessary to keep up with the efficiency of AI art tools. A crucial aspect of the creative arts is the human sentiment put into artworks, therefore research delving into the views of AI art versus human art would be worth considering for further inquiry.

References

- 1 J. Hopkins and D. Kiela, *ACL*, 168–178.
- 2 W. Kevin and SHuma, *Journal of Experimental Theoretical Artificial Intelligence*, **27**, 123–135.
- 3 A. Kaplan, *Oxford Research Encyclopedia of Business and Management*.
- 4 J. Doyle and T. Dean, *AI magazine*, **18**, 87, .
- 5 J. Hutson and M. Harper-Nichols, *Generative AI and Algorithmic Art: Disrupting the Framing of Meaning and Rethinking the Subject-Object Dilemma*, <https://digitalcommons.lindenwood.edu/cgi/viewcontent.cgi?article=1463>, Accessed: Oct. 27, 2023. [Online]. Available:.
- 6 Z. Epstein, *Art and the science of generative AI: A deeper dive*, [Online]. Available: arXiv:2306.04141v1 [cs.AI].
- 7 T. Feng, *Journal of Cultural Media Studies*, **36**, 110–125,.
- 8 R. Wendrich, *Int. Des. Conf.*
- 9 J. She and E. Cetinic, *ACM Transactions on Multimedia Computing, Communications and Applications*, **18**, 1–22,.
- 10 R. Srinivasan and K. Uchino, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 41–51, .
- 11 A. Field, *Interdisciplinary Science Reviews*, **47**, 185–198,.
- 12 M. Zameshina, O. Teytaud and L. Najman, *Diverse Diffusion: Enhancing Image Diversity in Text-to-Image Generation*, [Online].rXiv:2310.12583v1 [cs.CV].
- 13 S. Chatterjee, *Front Psychol.*
- 14 A. Oksanen, *Science Direct*, **1**, year.
- 15 L. Bellaiche, *Humans versus AI: whether and why we prefer human-created compared to AI-created artwork*, *Cognitive Research: Principles and Implications*, <https://cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-023-00499-6#Sec4>, Accessed: Sept. 26, 2023. [Online]. Available:.
- 16 Y. Sun, C. Yang, Y. Lyu and R. Lin, *Applied Sciences*, **12**, year.
- 17 L. Gu and Y. Li, *Front Psychol.*
- 18 J. Hullman, A. Holtzman and A. Gelman, *Artificial intelligence and aesthetic judgment*, Northwestern University.
- 19 B. Liu, *Arte Individuo y Sociedad*, **35**, year.
- 20 T. Demmer, C. Kühnapfel, J. Fingerhut and M. Pelowski, *Science Direct*, **148**, year.