

Human or Algorithm? The Visual Turing Test of AI-Generated Images

Changsheng Wang*

Abstract

With the advancement of artificial intelligence (AI) technology, the application of AI in generating digital art images has become increasingly common. However, whether people can distinguish images created by the latest AI painting tools from those created by humans, as well as the strategies and success rate of such differentiation, remains to be explored. This study employs a double-blind experimental method, combining a visual Turing test and in-depth interviews, to investigate participants' ability to distinguish between human-created and AI-generated images, the strategies they use, and the success rate of their distinctions. The results show that participants' average accuracy in recognizing AI-generated images was 61.67%, higher than the traditional Turing test benchmark of 30%, but 38.33% of participants still failed to accurately distinguish between the images. Participants primarily use three strategies to differentiate the images: details and logic, aesthetic experience, and Human-like characteristics and material properties, with recognition success rates of 75.7%, 73.05%, and 64.5%, respectively. This study reveals the potential of AI in the field of visual arts while also highlighting the advantages of human observation and logical reasoning. It provides empirical evidence for future AI art creation and recognition.

Key Words: AI Painting, AI-Generated Images, Turing Test, Human-Computer Interaction, AI Art.

I. INTRODUCTION

In recent years, the application of artificial intelligence (AI) technology in the field of artistic creation has become increasingly common [1-2], and it is gradually becoming a popular means of creating digital art images [3-4]. Some scholars have explored the potential applications of AI art technology [5-7], the relationship between traditional painting and AI painting [8], whether AI possesses imagination [9], visual uncertainty in AI abstract art [10], and comparisons between human and AI painting [11]. Other scholars have also examined perceptions of AI art, the extent to which it is liked or appreciated, and its acceptability [12-15]. Researchers have proposed a trend that AI will become increasingly proficient in understanding artists' intentions and generating more human-like artistic images [16]. This development is expected to revolutionize art design education and the visual arts field [17-18]. However, creating artistic images with unique aesthetic taste and expression, similar to human-created images, remains highly challenging for AI [19]. Consequently, several critical questions have yet to be thoroughly investigated:

RQ1: Can participants distinguish between images cre-

ated by humans and those created by AI?

RQ2: What strategies do people use to differentiate between human-created and AI-generated images?

RQ3: What is the success rate of these strategies in distinguishing between human-created and AI-generated images?

We aim to address these questions through the logic of the Turing test. The famous Turing test, initially proposed by Alan Turing in 1950 and known as the "imitation game" [20], is designed to test the indistinguishability between machines and humans. With the development of large language models (LLMs) [21-22], these models have shown excellent performance in zero-shot and few-shot tasks [23]. Additionally, OpenAI's LLM, ChatGPT, is capable of generating human-level text across various domains [24-27]. It is becoming increasingly difficult for people to distinguish between human and machine outputs. Consequently, some researchers have begun studying these emerging AI tools from the perspective of the Turing test. For instance, some researchers have designed AI chatbot conversation games and surveys to investigate whether participants can identify if their conversation partner is a machine or a human [28-29]. Others have compared academic articles,

Manuscript received May 24, 2024; Revised August 12, 2024; Accepted August 14, 2024. (ID No. 24M-06-020)

Corresponding Author (*): Changsheng Wang, zhangshawang@gmail.com

¹Department of Performing Arts, Film, and Animation, Sejong University, Seoul, Korea, zhangshawang@gmail.com

medical texts, and abstracts generated by ChatGPT with those written by human authors [23,30-31]. In the field of visual art evaluation, a study conducted by Chamberlain indicates that individuals tend to attribute abstract art images to computers or AI, and representational art images to humans [6]. Gangadharbatla's research further corroborates Chamberlain and his colleagues' findings [32]. However, both studies have certain limitations. First, their research did not utilize the latest AI painting models, which may result in experimental outcomes not reflecting the true capabilities of current AI technology. Second, Gangadharbatla's study did not establish a rigorous control group for the experiment nor design experiments with specific art styles. Finally, in Chamberlain's experiments, the image materials used were lacking in aesthetic quality and information content. These factors might have led to biased experimental results.

To address the limitations of the aforementioned studies, our experimental "Human or Algorithm?" has implemented several improvements. First, we utilized the most advanced AI painting models available, which are capable of generating high-quality artworks that reflect the latest technological advancements. Second, the experiment employed a double-blind design, eliminating any potential bias from participants and analysts, thus ensuring the fairness and validity of data analysis. Finally, the study featured images of high aesthetic quality and diverse artistic styles, created by individuals with expertise in AI painting. These included 20 images across various styles such as photography, oil painting, watercolor, and abstract art, providing a comprehensive evaluation of artistic styles. We invited participants from diverse backgrounds to identify and evaluate multiple sets of images created by both AI and humans. Our experimental results not only reveal the progress of AI in artistic creation but also offer new insights into understanding AI's potential in this field.

II. EXPERIMENT DESIGN AND METHODS

2.1. Participants

The data collection for this study was conducted in two stages. The first stage involved an online questionnaire survey from May to June 2023, which yielded 218 valid responses. Upon completion of the first survey, we conducted in-depth interviews with a subset of respondents to obtain richer qualitative data and gain a deeper understanding of participants' perceptions and judgment criteria regarding AI-generated images. However, considering the insufficient sample size of experimental images used in the first survey, we increased the number of experimental images and conducted a second questionnaire survey from May 22 to May 27, 2024. Participants for the second survey were recruited

Table 1. Participant information (N=197).

Demographic		Frequency	%	Cumulative (%)
Gender	Male	113	57.36	57.36
	Female	84	42.64	100.00
Age (years)	<20	108	54.82	54.82
	20-30	71	36.04	90.86
	>30	18	9.14	100.00
Education level	Undergraduate	126	63.96	63.96
	Master	46	23.35	87.31
	Doctor	25	12.69	100.00
Country	China	152	68.02	68.02
	South Korea	45	31.98	100.00
Total		197		100

through major universities and social media platforms, including undergraduate and graduate students from China and South Korea, as well as individuals from various professions and age groups. All participants voluntarily joined the study upon viewing the questionnaire, ensuring the diversity and representativeness of the sample. The survey questionnaire, titled "Try it out, can you tell which image is AI-generated?", was administered on the online survey platform "Wenjuanxing" in mainland China. Ultimately, we received 217 responses, and after eliminating invalid questionnaires, we obtained 197 valid responses (Table 1).

The survey comprised two parts: the first part collected basic information from participants, and the second part included 20 images, asking respondents whether these images were AI-generated. To validate the survey's effectiveness, a pretest was conducted with 28 participants to assess the survey content and provide feedback on whether the content reflected characteristics that were difficult to distinguish.

2.2. Visual Turing Test Experiment Design

Our visual turing test experiment faced a core challenge: ensuring that AI-generated images were not readily distinguishable from human-created images. To address this, we invited five "mentors" from the Midjourney community, an AI platform, to generate the AI images. These mentors had generated an average of 16,000 images on Midjourney and possess over a year of experience in AI image generation. Their extensive usage and high volume of generated images had equipped them with significant AI creation expertise and techniques. For this experiment, the mentors used the Midjourney V5.2 model to generate the images, ensuring that the experimental results reflected the current advanced level of AI technology.

To ensure the fairness and validity of the experiment, we employed a double-blind design. This design eliminates any potential biases or preconceived notions during data analy-

sis and requires participants to rely solely on their observations and analyses to make judgments, without depending on any prior knowledge or expectations. To make the identification process more challenging, we selected ten groups of artistic images, including photography, oil painting, watercolor, abstract painting, sketching, board painting, digital sculpture, Chinese painting, pastel, and marker art. This diversity in artistic styles provided a more comprehensive understanding of the effects of AI-generated images. We also invited MJ mentors to create the relevant AI images. During communication with the mentors, we emphasized that the images generated for the experimental group should correspond in content to those in the control group, aiding in the accuracy of the evaluation. Additionally, we required MJ mentors to implement measures to simulate human artwork in AI-generated images. For instance, we instructed the mentors to mimic the texture characteristics of human artworks in abstract paintings, watercolor, and oil painting during AI generation. This ensured that AI-generated content also possessed similar texture features. Furthermore, considering the impact of content differences within the same theme, we asked MJ mentors to maintain similarities in composition, color, and other characteristics between the experimental images and the control images, while ensuring texture elements were present. This design aimed to replicate the characteristics of genuine human artworks, as human artists' creations also include such elements.

In addition, we employed in-depth interview research methods for qualitative analysis to gain a deeper understanding of the strategies participants used to differentiate between AI-generated and human-created images. The purpose of this study was to explore the cognitive processes and the success rate of these strategies among participants in the visual Turing test. The qualitative part of the study involved participants selected based on their practical experience in AI art and interest in the experiment. A total of 15 participants with various professional backgrounds and educational levels were recruited. The design and implementation of this quantitative and qualitative study strictly adhered to internationally recognized research ethics guidelines. These included, but were not limited to, respecting participants' autonomy, ensuring fairness, optimizing the benefit-risk ratio, and respecting participants' privacy and confidentiality. Participants provided written informed consent (Datasheet) to participate in this study.

III. RESULTS AND ANALYSIS

This study involved 197 participants, each tasked with distinguishing whether images were AI-generated (Fig. 1). During the experiment, we randomized the order of the images to prevent participants from detecting patterns. For

clearer presentation in the analysis, we re-encoded the material images. The experiment included 20 images, 10 of which were AI-generated (re-encoded as images 1, 3, 5, 7, 9, 11, 13, 15, 17, and 19), and the other 10 were human-created (re-encoded as images 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20). The pairs of images 1 and 2, 3 and 4, 5 and 6, 7 and 8, 9 and 10, 11 and 12, 13 and 14, 15 and 16, 17 and 18, and 19 and 20 correspond to the ten artistic styles we designed (photography, oil painting, watercolor, abstract painting, sketching, digital painting, digital sculpture, Chinese painting, pastel, and marker drawings).

In our AI painting visual Turing test study, we noted important participant groups, including those proficient at identifying algorithm-generated images, those skilled at recognizing human-created images, those with a full grasp of patterns in AI and human creations, and those completely unable to identify AI-generated images. These groups and their intersections provided rich insights for our overall analysis. Through in-depth interviews with participants, we explored how people attempt to distinguish between AI and human creations. We observed that participants used a range of strategies, showcasing human thinking's flexibility and attention to detail. These strategies stemmed from individuals' accumulated experiences in artwork creation and AI image generation processes.

3.1. Analysis of Experimental Results

Through data analysis of the experimental results, we found that the average accuracy rate for participants in correctly identifying AI-generated images was 61.67% (Fig. 2). This indicates that approximately 38.33% of participants could not accurately distinguish AI-generated images from human-created images. This result is significantly higher than the original Turing test benchmark of 30%, suggesting that the realism of AI in the field of visual arts has reached a level that makes it difficult for humans to discern. For human-created images, participants' recognition accuracy was relatively high, reaching 70.71%, reflecting a stronger ability to identify traditional visual art works. Specifically, among the AI-generated images, image 13 had the highest recognition accuracy at 77.66%, while image 7 had the lowest at 46.70%. These fluctuations could be influenced by differences in image generation technology, the complexity of image styles, or the intuitiveness of image content. For human-created images, images 10 and 2 had the highest recognition success rates at 81.22% and 81.73%, respectively, showing that participants could better identify the human artistic characteristics in these images. In contrast, image 14 had the lowest recognition success rate at 47.72%, possibly because its artistic style was similar to that of AI-generated images, increasing the difficulty of recognition.

The results of the paired t-test analysis indicate that the

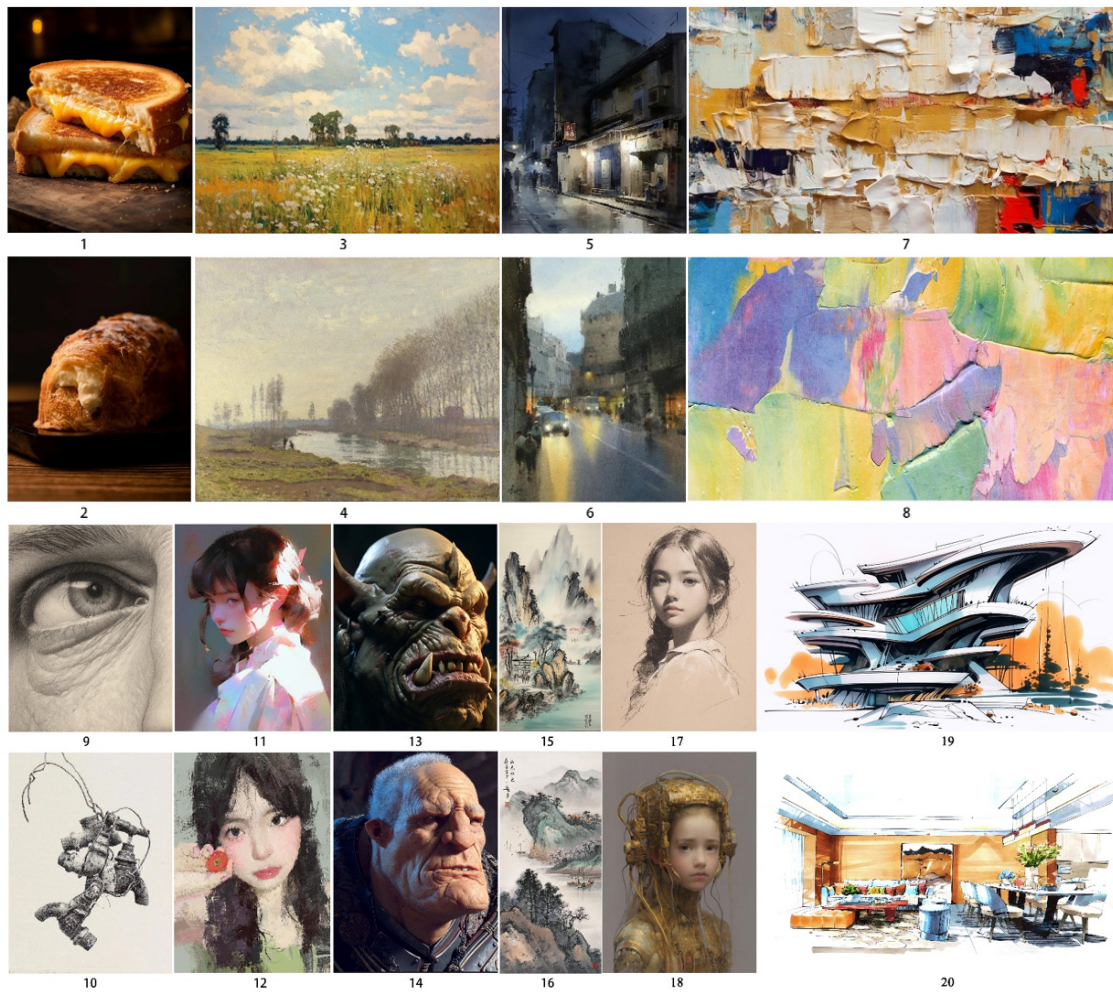


Fig. 1. Images used in the experiment (AI-generated images are 1, 3, 5, 7, 9, 11, 13, 15, 17, 19; Human-created images are 2, 4, 6, 8, 10, 12, 14, 16, 18, 20). AI-generated images were created by invited experts and authorized for use. Human-created images: Image 2, 8, 10 sourced from online searches; Image 4 sourced from Monet's "The Small Arm of the Seine at Argenteuil, 1872"; Image 6 created by Jian Zhongwei from Taiwan; Image 12 created by Wang Changsheng from China; Image 14 sourced from ArtStation by occultart; Image 16 created by Huang Huanwu from China (1906–1985); Image 18 created by Jia Wei from China; Image 20 created by Bai Taotao from China.



Fig. 2. shows the recognition accuracy of AI-generated and human-created images in different art styles.

differences in recognition accuracy between AI-generated images and human-created images were statistically significant for most image pairs. For example, the mean recognition accuracy for Pair 1 and Pair 2 were 1.35 (SD=0.48) and 1.80 (SD=0.40), respectively, with a difference of -0.45 ($t=-7.193$, $p<0.001$), indicating a significant difference. Similarly, for Pair 3 and Pair 4, the mean recognition accuracies were 1.41 (SD=0.49) and 1.84 (SD=0.37), respectively, with a difference of -0.43 ($t=-7.081$, $p<0.001$), indicating a significant difference. These results suggest that, except for Pair 7 and Pair 8, the differences in recognition accuracy for the other pairs were statistically significant, reflecting a significant difference in participants' ability to recognize AI-generated images versus human-created images. Furthermore, the overall t-test results were as follows: $t=-1.934$, $p=0.0690$. Although the paired t-tests for individual image pairs showed significant differences in recognition accuracy for some pairs, the overall t-test results did not show statistical significance (Table 2).

3.2. Strategies for Distinguishing Between Human-created images and AI-Generated Images

We found through in-depth interviews that participants

employed a range of strategies to identify the source of images, and we recorded the frequency and success rate of these strategies. We calculated the usage frequency and accuracy of recognition strategies and provided formulas and example results (Tables 3 and Table 4). Additionally, based on the results of the independent sample t-tests, we analyzed the differences in the success rates of identifying AI-generated images and human-created images among the three recognition strategies (details and logic, aesthetic experience, Human-like characteristics and material properties). The differences between the "details and logic" strategy and the "Human-like characteristics and material properties" strategy were significant ($t=5.722$, $p<0.001$), as were the differences between the "aesthetic experience" strategy and the "Human-like characteristics and material properties" strategy ($t=4.267$, $p=0.013$). However, the differences between the "details and logic" strategy and the "aesthetic experience" strategy were not significant ($t=0.761$, $p=0.487$). This indicates that the "details and logic" strategy and the "aesthetic experience" strategy have similar effectiveness in recognizing AI-generated images versus human-created images, both being significantly more effective than the "Human-like characteristics and material properties"

Table 2. Results of paired t-test analysis.

Pair name	Paired (mean±S.D)		Difference (Pair 1 - Pair 2)	<i>t</i>	<i>p</i>
	Pair 1	Pair 2			
1 Pair 2	1.35±0.48	1.80±0.40	-0.45	-7.193	0.000**
3 Pair 4	1.41±0.49	1.84±0.37	-0.43	-7.081	0.000**
5 Pair 6	1.49±0.50	1.77±0.42	-0.28	-3.858	0.000**
7 Pair 8	1.63±0.49	1.68±0.47	-0.05	-0.761	0.449
9 Pair 10	1.33±0.47	1.90±0.30	-0.57	-9.947	0.000**
11 Pair 12	1.24±0.43	1.65±0.48	-0.41	-6.435	0.000**
13 Pair 14	1.13±0.34	1.33±0.47	-0.20	-3.761	0.000**
15 Pair 16	1.15±0.36	1.63±0.49	-0.48	-8.319	0.000**
17 Pair 18	1.40±0.49	1.72±0.45	-0.32	-4.707	0.000**
19 Pair 20	1.51±0.50	1.73±0.45	-0.22	-3.591	0.001**

* $p<0.05$, ** $p<0.01$.

Table 3. Formulas for calculating usage frequency and success rate of recognition strategies.

Calculation item	Formula
Total number of recognitions	Total number of recognitions = Usage frequency × Number of images judged per person
Correct recognitions of AI-generated images	Correct recognitions of AI-generated images = Usage frequency × Number of AI-generated images judged per person × Success rate for identifying AI-generated images
Correct recognitions of human-created images	Correct recognitions of human-created images = Usage frequency × Number of human-created images judged per person × Success rate for identifying human-created images
Total correct recognitions	Total correct recognitions = Correct recognitions of AI-generated images + Correct recognitions of human-created images
Overall success rate	Overall success rate = (Total correct recognitions / Total number of recognitions) × 100%

Table 4. Usage frequency and success rate of recognition strategies.

Recognition strategy	Usage frequency (n)	Accuracy for identifying AI-generated images (%)	Accuracy for identifying human-created images (%)	Overall success rate (%)
Detail and logic	13	71.1	80.3	75.7
Aesthetic experience	10	69.3	76.8	73.05
Human-like characteristics and material properties	8	61.2	67.8	64.5

strategy.

3.2.1. Details and Logic

A total of 13 respondents used the "details and logic" strategy, with an accuracy rate of 71.1% for identifying AI-generated images and 80.3% for identifying human-created images, resulting in an overall success rate of 75.7%. This indicates that the "details and logic" strategy is the most widely used and effective strategy. Participants judged the source of the images by examining details such as signatures, text processing, facial features, and limbs. AI-generated images often have noticeable errors or simplifications in these details, making this strategy highly successful in distinguishing the source of the images.

The first type of detail mentioned by interviewees was the signature and text processing in the image. Some participants judged the source of the image by checking the signatures and text processing. This strategy was quite effective in identifying AI-generated images. Human creators typically sign their works or use text correctly, whereas AI-generated images often contain errors in these details. A university student in art design from Beijing, Li (pseudonym), noted, "AI-generated works cannot correctly sign and handle text, and careful observation reveals that the text is all wrong" (Female, 22 years old, Beijing, China). Conversely, a graduate student from Xiamen, Wu (pseudonym), expressed a different view: "In the Midjourney V4 version, many AI-generated works with signatures appeared, so signatures cannot be used as a strategy to determine whether an image is AI-generated. However, AI still lacks the ability to handle text correctly" (Male, 25 years old, Xiamen, China).

The second detail is the depiction of facial features and limbs. When it comes to images depicting people, participants paid special attention to the details of facial features and limbs. They noticed that AI-generated images often had obvious structural errors or were overly simplified in these parts, whereas human-created images paid more attention to the details and individuality of these parts. A university student from Hangzhou, Zhao (pseudonym), stated, "When observing AI-generated portraits, I found that it often has problems when generating facial features and hands and feet. For example, AI may generate overly refined and ho-

mogenized facial features, ignoring the uniqueness of each person's appearance" (Female, 20 years old, Hangzhou, China). A university student from Nanjing, Li (pseudonym), emphasized, "I noticed that AI tends to be clumsy when dealing with hands and feet. For example, it might produce unnatural proportions or obvious structural errors when depicting fingers and palms. In human-created works, I can see the artist's careful consideration and meticulous handling of these details" (Male, 22 years old, Nanjing, China).

AI also tends to make mistakes in logical coherence. Participants showed impressive insight when evaluating whether images conformed to physical and real-world principles. For instance, they looked at whether the image correctly depicted lighting, shadows, and proportions of objects. Human-created images usually followed these principles better, while AI-generated images often displayed unnatural aspects. A university student from Hangzhou, Wang (pseudonym), shared his observations: "I found that AI often handles light and shadow unnaturally. For example, sometimes it draws too much light in a shadowed area or produces unreasonable dark parts in bright areas. This makes me feel that AI does not understand how light exists in three-dimensional space" (Male, 21 years old, Hangzhou, China). Another undergraduate student from Xinjiang, Lin, remarked, "When distinguishing AI-generated images, I found that AI sometimes makes obvious logical errors. For example, in the experiment, the depiction of butter spreading on a sandwich showed a physically impossible shape. This immediately led me to identify it as AI-generated. However, the experiment's images were very deceptive, and I also misidentified another human-created work as AI-generated" (Male, 21 years old, Xinjiang, China).

3.2.2. Aesthetic Experience

Some interviewees reported that they relied on aesthetic experience to identify whether an image was AI-generated. A total of 10 respondents used the aesthetic experience strategy, with a success rate of 69.3% for identifying AI-generated images and 76.8% for identifying human-created images, resulting in an overall success rate of 73.05%. This indicates that participants have a relatively high success rate when distinguishing the source of images through their aesthetic judgment, though there is still a certain rate of

misjudgment.

A university student from Shanghai, Liu (pseudonym), shared his observations, "I found that human creators are thoughtful and interesting when dealing with the use of negative space and positive and negative shapes. In the sketch works from the experiment, I judged it to be a human-created images based on the relationship and composition between the faucet and the wire" (Male, 22 years old, Shanghai, China). Zhao (pseudonym), a graduate student in architectural design from Beijing, expressed a similar view, "For those familiar with painting, observing the use of negative space and positive and negative shapes can effectively identify the source of an image. AI tends to fill the image too much when drawing, as it lacks an understanding of overall compositional harmony" (Female, 25 years old, Beijing, China). Another identification strategy is the sense of unreality. Some participants mentioned that through extensive artistic training, they rely on intuition to make judgments. AI-generated images often appear too realistic, creating a sense of artificiality.

A Ph.D. student from Hangzhou, Zhao (pseudonym), stated, "I feel that AI-generated images, in dealing with the issue of unreality, tend to be too realistic; therefore, in the experiment, I judged both digital sculptures as AI-generated content. However, one of them was made by humans, making me realize that relying solely on the presence of realism, though it can filter out AI content, also leads to misjudgment of human digital works" (Male, 29, Hangzhou, China). A graduate student from Beijing, Zhang (pseudonym), also agreed with his view, "In my opinion, the finesse of AI-generated content is enough to confuse people now, almost indistinguishable from human digital sculptures or modeling works" (Female, 25, Beijing, China).

3.2.3. Human-Like Features and Material Properties

When distinguishing between AI-generated images and human-created images, some interviewees mentioned that closely observing the brushstrokes is an effective strategy. A total of 8 respondents used the human-like features and material properties strategy, with a success rate of 61.2% for identifying AI-generated images and 67.8% for identifying human-created images, resulting in an overall success rate of 64.5%. This indicates that while this strategy depends on observing human-like features and material properties in the images, the overall success rate is relatively low, possibly because AI-generated images are increasingly approaching the level of human creations in these details.

Although AI-generated images exhibit brushstroke characteristics similar to those of humans, they differ significantly from the irregular qualities of human brushstrokes, especially in thick-painting styles. This observation is supported by Han (pseudonym), a doctoral student in the oil

painting department, who noted, "The brushstrokes in AI-generated images usually appear stiff. For those of us who have been engaged in artistic creation for a long time, this is a very clear sign" (Female, 31 years old, Guangzhou, China). A student from Nanjing, Ding (pseudonym), agreed with this view and further stated, "Even beginners can distinguish works created by human artists by carefully observing the variations in digital painting brushstrokes" (Male, 19 years old, Nanjing, China). In the category of Chinese painting, we found that interviewees could distinguish AI and human creations by observing the brushstroke techniques specific to Chinese painting, known as "cunfa." They noticed that AI often lacks the liveliness and naturalness when simulating this complex technique, whereas human creators can use it more flexibly. An undergraduate student from Changsha, Ding (pseudonym), shared his experience, "I found that in images using Chinese painting brushstroke techniques, AI-generated images do not pay much attention to the brushwork. From my learning experience, I understand that the brushstroke techniques in Chinese painting are very subtle and full of charm" (Male, 20 years old, Changsha, China). Conversely, a graduate student from Beijing, Wang (pseudonym), expressed a different view, "As someone without art training, I could hardly distinguish which Chinese paintings were AI-generated in this experiment. It mimicked the Chinese painting style very well" (Female, 24 years old, Beijing, China).

IV. DISCUSSION

This study aims to explore whether participants can distinguish between human and AI-generated images using the Turing test method. The results indicate that the realism of AI in the field of visual arts has reached a level where it is difficult to discern. Participants' accuracy in identifying AI-generated images was 61.67%, although higher than the traditional Turing test benchmark of 30%, still, 38.33% of participants could not accurately distinguish between AI-generated and human-created images. This result is consistent with the findings of Chamberlain and Gangadharbatla, who noted difficulties participants have in distinguishing between abstract and figurative art [6,32]. Additionally, the results of the paired t-test analysis indicate that the differences in recognition accuracy between AI-generated images and human-created images were statistically significant for most image pairs. This suggests that participants showed a significant difference in their ability to recognize AI-generated versus human-created images, with people being better at recognizing human-created images. This also highlights that human-created images still hold an advantage, particularly those created by top human artists, which far exceed the quality of AI-generated images. How-

ever, despite the significant differences in recognition accuracy for some individual image pairs, the overall t-test results did not show statistical significance. Specifically, the t-statistic was -1.934 and the p-value was 0.0690 , meaning that there is no significant difference in the overall recognition accuracy between AI-generated images and human-created images. Especially in the abstract painting style, participants' accuracy in identifying AI-generated images was below 50%. This may reflect that with technological advancements, AI-generated images have approached the level of human-created images in terms of visual expressiveness and complexity. Particularly across different artistic styles, AI can simulate works that closely resemble those of human artists. These findings reveal the immense potential of AI in the field of artistic creation and suggest that the boundaries between human and AI art are becoming increasingly blurred. In the long term, the growing prevalence of AI-generated art may redefine the concept of artistic creation. The art community might need to reconsider the definitions of originality and creativity, as AI can simulate and even surpass the skills of certain human artists. This shift could lead to changes in art education, moving from traditional skills to a greater focus on fostering creativity and critical thinking. Moreover, the collaboration between AI and human artists holds significant promise. Artists can use AI as a tool to expand their creative possibilities, producing more complex and diverse works. AI can also help artists achieve new artistic styles and modes of expression, driving innovation and development in art forms. Collaborative creation with AI might become a major trend in future artistic practices [2].

Furthermore, this study delves into the strategies participants use to distinguish between human and AI-generated images. Interview results indicate that participants employed three main strategies: attention to detail and logic, aesthetic experience, and recognition of human-like characteristics and material properties. These findings confirm previous research on Turing tests with ChatGPT, which mentioned strategies involving details and human-like characteristics (such as identifying grammatical errors and responses to emotional questions to determine if AI is involved) [29]. The application of these strategies reveals the complex cognitive processes humans engage in when identifying AI-generated artworks and underscores the importance of detailed observation [33], logical reasoning, and aesthetic experience. Despite AI's significant progress in simulating human artistic creation, humans still possess unique observational and judgment abilities. As AI technology continues to advance, these strategies may evolve and adapt to new challenges, potentially impacting artists [34]. However, this does not signify the "end of art," but rather reshapes the roles and practices of creators and alters the aesthetics of contemporary media [35].

Lastly, our study also revealed the frequency, success rate, and differences of the strategies participants used to distinguish between human and AI-generated images. The detail and logic strategy was the most frequently used and the most effective, showing significant differences compared to the "human-like characteristics and material properties" strategy. Participants judged the origin of images by examining details such as signatures, text processing, facial features, hands and feet, as well as the logical consistency of lighting, shadows, and object proportions. The high success rate of this strategy reflects human strengths in detailed observation and logical reasoning, further confirming previous literature on AI's lack of subtle details and depth found in human artworks [13]. The aesthetic experience strategy was also widely adopted and showed significant differences compared to the "human-like characteristics and material properties" strategy. This strategy relies on participants' art training and intuition to determine the origin of the images. The high success rate indicates the importance of art training and aesthetic intuition in recognizing the authenticity and artistic quality of images. However, the success rate for identifying AI-generated images using this strategy was only 69.3%. This suggests that even trained artistic eyes face challenges when dealing with images generated by evolving AI painting tools. The success rate of the "human-like characteristics and material properties" strategy was relatively lower, with significant differences compared to the other two strategies. This indicates that AI has made significant progress in mimicking human artistic characteristics and material representations, gradually improving the realistic simulation of texture and brushstroke effects. Although AI-generated images still appear somewhat unnatural in the material representation of certain styles, they have achieved a high level of realism. Therefore, participants had a lower success rate when using this strategy to identify images.

Moreover, our study uncovered some unexpected findings. Specifically, images classified as figurative photography (Image 1) had an success rate of 65.99%, while images classified as abstract painting (Image 7) had an success rate of 46.7%. This is contrary to the findings of Chamberlain and Gangadharbatla [6,32]. In our experiment, people did not attribute abstract art images to AI or figurative art images to humans. This may be due to advancements in AI painting models that have diminished the gap between figurative and abstract images. However, our study has some limitations. Although it included ten groups of common art image types, it may not have fully covered all artistic styles that AI can generate. Additionally, participants' awareness of the experiment's purpose and environment might have influenced their identification strategies. This influence could lead to judgments that differ from those they would make in everyday contexts. Finally, due to the

two-stage data collection process, the participants in the in-depth interviews and the second questionnaire survey were not the same group of people. This unique situation may have had a certain impact on the research results.

V. CONTRIBUTION

This study makes three major contributions to the exploration of the identification of AI-generated images versus human-created images. First, in the design of the AI visual Turing test method, we adopted an innovative double-blind experimental approach to avoid any potential biases from participants and analysts, ensuring the fairness and validity of data analysis. By using the latest AI painting model, Midjourney V5.2, to generate high-quality images in various artistic styles, the experimental results are more representative and broadly applicable, addressing the shortcomings of previous research. Secondly, we proposed strategies for identifying AI-generated images and explored in depth the three main strategies participants use to distinguish between human and AI-generated images: detail and logic, aesthetic experience, and human-like characteristics and material properties. These strategies enhance our understanding of the cognitive processes humans engage in when encountering AI-generated artworks. Finally, our study analyzed the frequency, success rate, and differences in the practical application of these strategies. The results indicate that the detail and logic strategy is the most frequently used and most effective, with participants successfully identifying AI-generated images by examining details such as signatures, text processing, facial features, and the logical consistency of lighting and shadows. The aesthetic experience strategy, which relies on participants' art training and intuition, also achieved significant success. The human-like characteristics and material properties strategy had the lowest success rate, indicating that while AI has made significant progress in mimicking human artistic characteristics, it still faces challenges.

VI. CONCLUSION

This study aims to address three key questions: Can participants distinguish between images created by humans and those created by AI? What strategies do people use to differentiate between human-created and AI-generated images? What is the success rate of these strategies in distinguishing between human-created and AI-generated images? Through visual Turing tests and in-depth interviews, we explore these questions comprehensively.

Firstly, the study results show that participants' accuracy in identifying AI-generated images is 61.67%, slightly

higher than the 30% benchmark of the traditional Turing test, yet 38.33% of participants still fail to accurately distinguish between AI-generated and human-created images. This indicates that AI has achieved a level of realism in the field of visual arts that makes differentiation challenging, highlighting the blurring boundaries between human and AI artistic creation. Secondly, interview results reveal that participants primarily use three strategies to distinguish between human and AI-generated images: details and logic, aesthetic experience, and Human-like characteristics and material properties. Among these strategies, the detail and logic strategy is the most commonly used and effective, with a success rate of 75.7%. The success rate of the aesthetic experience strategy is 73.05%, while the Human-like characteristics and material properties strategy shows a success rate of 64.5%. These results reflect the advantages humans have in detailed observation, logical reasoning, and perception of aesthetic and material properties. Lastly, this study makes significant contributions in several areas: In the design of AI visual Turing test methods, we adopted an innovative double-blind experimental design to ensure the fairness and validity of data analysis; we proposed specific strategies for identifying AI-generated images, enriching our understanding of human cognitive processes when encountering AI-generated artworks; We validated the success rates of these strategies, providing important empirical evidence for research on AI artistic creation and recognition.

REFERENCES

- [1] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, and P. S. Yu, et al., "A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT," *arXiv Pre.arXiv:2303.04226*, 2023.
- [2] C. Wang, "AI-driven digital image art creation: Methods and case analysis," *Chinese Journal of Intelligent Science and Technology*, vol. 5, no. 3, pp. 406-414, Sep. 2023.
- [3] M. A. Runco and G. J. Jaeger, "The standard definition of creativity," *Creativity Research Journal*, vol. 24, no. 1, pp. 92-96, Feb. 2012.
- [4] K. Crowson, S. Biderman, D. Kornis, D. Stander, and E. Hallahan, and L. Castricato, et al., "VQGAN-CLIP: Open domain image generation and editing with natural language guidance," in *European Conference on Computer Vision*, Cham, 2022, pp. 88-105.
- [5] E. Cetinic and J. She, "Understanding and creating art with AI: Review and outlook," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 2, pp. 1-22, Feb. 2022.
- [6] R. Chamberlain, C. Mullin, B. Scheerlinck, and J. Wagemans, "Putting the art in artificial: Aesthetic res-

- ponses to computer-generated art," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 12, no. 2, p. 177-192, May 2018.
- [7] J. W. Hong and N. M. Curran, "Artificial intelligence, artists, and art: Attitudes toward artwork produced by humans vs. artificial intelligence," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 15, no. 2s, pp. 1-16, Jul. 2019.
- [8] X. Liu, "Artistic reflection on artificial intelligence digital painting," *Journal of Physics: Conference Series*, vol. 1648, no. 3, p. 032125, 2020.
- [9] M. A. Boden, "Creativity and artificial intelligence," *Artificial Intelligence*, vol. 103, no. 1-2, pp. 347-356, Aug. 1998.
- [10] M. Zeilinger, "The politics of visual indeterminacy in Abstract AI Art," *Leonardo*, vol. 56, no. 1, pp. 76-80, Feb. 2023.
- [11] Y. Sun, Y. Lyu, P. H. Lin, and R. Lin "Comparison of cognitive differences of artworks between artist and artistic style transfer," *Applied Sciences*, vol. 12, no. 11, p. 5525, May 2022.
- [12] E. Ch'ng, "Art by computing machinery: Is machine art acceptable in the artworld?," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 15, no. 2s, pp. 1-17, Jul. 2019.
- [13] J. W. Hong and N. M. Curran, "Artificial intelligence, artists, and art: Attitudes toward artwork produced by humans vs. artificial intelligence," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 15, no. 2s, pp. 1-16, Jul. 2019.
- [14] S. S. Lee, K. Park, and Y. Kim, "Are you ready to embrace art work made by artificial intelligence?-The asymmetric effects of attitudes toward art work (art vs. art infused product) and painting agent (human vs. artificial intelligence)," in *Global Fashion Management Conference*, Paris, 2019, pp. 46-50.
- [15] M. Ragot, N. Martin, and S. Cojean, "AI-generated vs. human artworks. a perception bias towards artificial intelligence?," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, 2020, pp. 1-10.
- [16] Z. W. Wu, H. Qu, and K. Zhang, "A survey of recent practice of artificial life in visual art," *Artificial Life*, vol. 30, no. 1, pp. 106-135, Feb. 2024.
- [17] J. Ploennigs and M. Berger, "AI art in architecture," *AI in Civil Engineering*, vol. 2, no. 1, p. 8, Aug. 2023.
- [18] F. Kong, "Application of artificial intelligence in modern art teaching," *International Journal of Emerging Technologies in Learning*, vol. 15, no. 13, pp. 238-251, Jul. 2020.
- [19] I. Santos, L. Castro, N. Rodriguez-Fernandez, A. Torrente-Patino, and A. Carballal, "Artificial neural networks and deep learning in the visual arts: A review," *Neural Computing and Applications*, vol. 33, pp. 121-157, Jan. 2021.
- [20] A. M. Turing, "Computing Machinery and Intelligence," *Creative Computing*, vol. 6, no. 1, pp. 44-53, Jan. 1980.
- [21] G. I. Winata, A. Madotto, Z. Lin, R. Liu, J. Yosinski, and P. Fung, "Language models are few-shot multilingual learners," *arXiv Pre.arXiv:2109.07684*, 2021.
- [22] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, and P. Mishkin, et al., "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27730-27744, 2022.
- [23] W. Liao, Z. Liu, H. Dai, S. Xu, Z. Wu, and Y. Zhang, et al., "Differentiating ChatGPT-generated and human-written medical texts: Quantitative study," *JMIR Medical Education*, vol. 9, no. 1, p. e48904, May 2023.
- [24] T. Susnjak, "Applying BERT and ChatGPT for sentiment analysis of Lyme disease in scientific literature," in *Borrelia burgdorferi: Methods and Protocols*, New York, NY, Springer US, pp. 173-183, 2024.
- [25] X. Wei, X. Cui, N. Cheng, X. Wang, X. Zhang, and S. Huang, et al., "Zero-shot information extraction via chatting with ChatGPT," *arXiv Pre.arXiv:2302.10205*, 2023.
- [26] H. Dai, Z. Liu, W. Liao, X. Huang, Y. Cao, and Z. Wu, et al., "AugGPT: Leveraging ChatGPT for text data augmentation," *arXiv Pre.arXiv:2302.13007*, 2023.
- [27] Z. Liu, Y. Huang, X. Yu, L. Zhang, Z. Wu, and C. Cao, et al., "DeID-GPT: Zero-shot medical text de-identification by GPT-4," *arXiv Pre.arXiv: 2303.11032*, 2023.
- [28] A. Ujhelyi, F. Almosdi, and A. Fodor, "Would you pass the turing test? Influencing factors of the turing decision," *Psihologijske Teme*, vol. 31, no. 1, pp. 185-202, Apr. 2022.
- [29] D. Jannai, A. Meron, B. Lenz, Y. Levine, and Y. Shoham, "Human or not? A gamified approach to the Turing test," *arXiv Pre.arXiv:2305.20010*, 2023.
- [30] C. A. Gao, F. M. Howard, N. S. Markov, E. C. Dyer, S. Ramesh, and Y. Luo, et al., "Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers," *BioRxiv*, Dec. 2022, doi: 10.1101/2022.12.23.521610.
- [31] S. Ariyaratne, K. P. Iyengar, N. Nischal, N. Chitti Babu, and R. Botchu, "A comparison of ChatGPT-generated articles with human-written articles," *Skeletal Radiology*, vol. 52, no. 9, pp. 1755-1758, Apr. 2023.
- [32] H. Gangadharbatla, "The role of AI attribution knowledge in the evaluation of artwork," *Empirical Studies of the Arts*, vol. 40, no. 2, pp. 125-142, Feb. 2022.

- [33] C. Wang, "Art innovation or plagiarism? Chinese Students' attitudes towards AI painting technology and influencing factors," *IEEE Access*, vol. 12, pp. 85795-85805, Jul. 2024.
- [34] H. H. Jiang, L. Brown, J. Cheng, M. Khan, A. Gupta, and D. Workman, et al., "AI art and its impact on artists," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 2023, pp. 363-374.
- [35] Z. Epstein and A. Hertzmann, "Art and the science of generative AI," *Science*, vol. 380, no. 6650, pp. 1110-1111, Jun. 2023.

AUTHOR



Changsheng Wang is a Ph.D. candidate in the Animation Department at Sejong University, Seoul, Korea. His main research interests are in AI painting and art.

