



## Article

# Communication in Human–AI Co-Creation: Perceptual Analysis of Paintings Generated by Text-to-Image System

Yanru Lyu <sup>1,2</sup> , Xinxin Wang <sup>3</sup>, Rungtai Lin <sup>4</sup>  and Jun Wu <sup>5,\*</sup>

<sup>1</sup> Department of Digital Media Arts, School of Media and Design, Beijing Technology and Business University, Beijing 102488, China

<sup>2</sup> Key Lab of Encyclopedia Knowledge Fusion Innovation Publishing Project, Beijing 100037, China

<sup>3</sup> Art Teaching and Research Section, Beijing International Studies University, Beijing 100024, China

<sup>4</sup> Graduate School of Creative Industry Design, National Taiwan University of Arts, New Taipei 220307, Taiwan

<sup>5</sup> Department of Digital Media Arts, School of Art and Design, Shenzhen University, Shenzhen 518061, China

\* Correspondence: junwu2006@hotmail.com

**Abstract:** In recent years, art creation using artificial intelligence (AI) has started to become a mainstream phenomenon. One of the latest applications of AI is to generate visual artwork from natural language descriptions where anyone can interact with it to create thousands of artistic images with minimal effort, which provokes the questions: what is the essence of artistic creation, and who can create art in this era? Considering that, in this study, the theoretical communication framework was adopted to investigate the difference in the interaction with the text-to-image system between artists and nonartists. In this experiment, ten artists and ten nonartists were invited to co-create with Midjourney. Their actions and reflections were recorded, and two sets of generated images were collected for the visual question-answering task, with a painting created by the artist as a reference sample. A total of forty-two subjects with artistic backgrounds participated in the evaluated experiment. The results indicated differences between the two groups in their creation actions and their attitude toward AI, while the technology blurred the difference in the perception of the results caused by the creator's artistic experience. In addition, attention should be paid to communication on the effectiveness level for a better perception of the artistic value.

**Keywords:** AI painting; human–AI interaction; artistic perception; creativity; text-to-image; prompt



**Citation:** Lyu, Y.; Wang, X.; Lin, R.; Wu, J. Communication in Human–AI Co-Creation: Perceptual Analysis of Paintings Generated by Text-to-Image System. *Appl. Sci.* **2022**, *12*, 11312. <https://doi.org/10.3390/app122211312>

Academic Editor: Agostino Forestiero

Received: 30 September 2022

Accepted: 4 November 2022

Published: 8 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

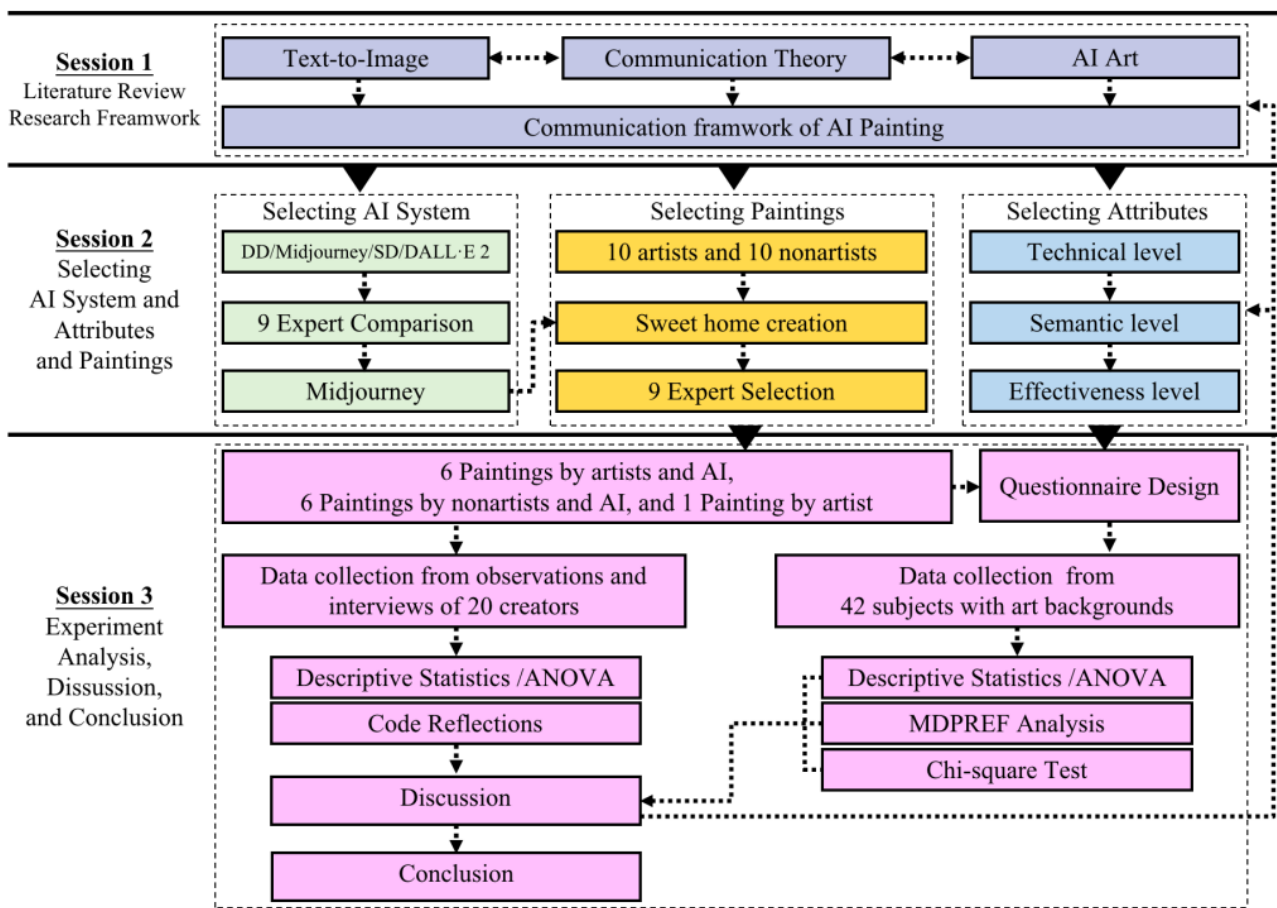
## 1. Introduction

In the last decade, the growing implementation of artificial intelligence (AI) technology in the field of art has triggered a fierce discussion on AI art. Since the generative adversarial network (GAN) portrait painting titled “Edmond de Belamy” was constructed in 2018, AI art has already entered the public's vision. One of the latest applications of AI is the generation of images based on natural language descriptions, which enhances the efficiency and effect of the transformation from creativity to visuality to a great extent. In the past, whether in traditional or digital painting creation, the author needed to be skilled in using tools and to have rich technical experience to accurately map the brain's imagination to the visual layer. However, in co-creation with text-to-image AI generators, both artists and nonartists can input the text description to produce many high-quality images. During traditional painting creation, artists and nonartists in a painting task indicated quantitative and qualitative differences in some studies, such as artists spending more time on planning their painting, having more control over their creative processes, having more specific skills, and having more efficiency than nonartists [1,2]. Whether such differences still exist in the new human–AI interaction mode and what new changes arise are worth discussing.

A series of text-to-image AI systems, such as Disco Diffusion [3], Midjourney [4], Stable Diffusion [5], OpenAI's DALL-E 2 [6], and Google's Imagen [7], is making a big splash. The generation mechanism is to use a language–vision model to understand the

“prompt” input by users, and then the generator is guided to produce high-quality images. They are capable of synthesizing images with any style and content based on a prompt. Besides, users can control the system to iterate more variations. With the rise of AI art, many artists have also started to use AI to assist in creation. According to the Colorado State Fair competition’s website [8], the art piece “Théâtre D’opéra Spatial,” which was generated by Midjourney, won first place in the digital art category. As the formation of generators using natural language text to create various styles of creative images occurs, the question that arises immediately is: what is the essence of artistic creation, and what is the core capability of artists? Though everyone thought art was one thing robots could never do, maybe we will face the challenges of emerging AI technology.

This research aimed to analyze and understand how text-to-image technology affects art creation and appreciation. Additionally, the main discussion focused on the difference in activities and results between artists and nonartists from the perspective of art communication. Figure 1 shows that this study could be divided into three sections. In Section 1, a literature review was made to explore the research framework of the generation mechanism of visual art collaboration with AI. In Section 2, nine experts with artistic and/or aesthetic backgrounds were invited to select a suitable AI system and painting samples according to their art appreciation. In Section 3, the data were collected from the creators of the samples and from the subjects participating in the questionnaire for analysis and discussion. Finally, the conclusions of this study were given.



**Figure 1.** The procedures for this study: the horizontal line divides three sessions, and the arrows indicate the direction of functions and processes. The original name of “DD” is “Disco Diffusion”, while that of “SD” is “Stable Diffusion”.

## 2. Literature Review

### 2.1. Text-to-Image Systems

With the successful application of transformer-based architectures in neural language processing (NLP), text-to-image systems based on deep generative models have become popular means for computer vision tasks [9,10]. They generate creative images combining concepts, attributes, and styles from expressive text descriptions [11]. The primary generation mechanism is that a language–vision model (i.e., CLIP) is adopted to guide the generator to produce high-quality images.

When OpenAI released CLIP in 2021 [12], it spurred immense technical progress in text-to-image generation. CLIP is a pre-trained language–vision model that enables zero-shot image manipulation guided by text prompts. Unlike traditional representation learning that is based mostly on discretized labels, the vision–language model aligns images and texts in a common feature space, allowing zero-shot transfer to a downstream task via prompting [13]. CLIP guides the generator to synthesize digital images when used as a discriminator in a generative system. Using its joint text–image representation space, we can control the synthesis process with natural language. At present, most programs use CLIP for text encodings, such as DALL-E 2 and Stable Diffusion. Differently, Google’s Imagen uses the T5-XXL language model to encode the text and then generate images directly without learning the prior model [7]. The text input, known as the prompt, plays a crucial role in downstream datasets. It is an important aspect for improving the quality and changing the aesthetics of images, which entails the practice and capabilities of interacting with the system. The term prompt engineering knows the practice and skill of writing prompts due to its iterative and experimental nature [14]. However, identifying the right prompt is a nontrivial task which often takes a significant amount of time for word tuning—a slight change in wording could make a huge difference in performance [13].

Currently, text-to-image generation models can be divided into two designs: sequence-to-sequence modeling and diffusion-based modeling [15]. The main idea of the sequence-to-sequence modeling design is to turn images into discrete image tokens via leveraging transformer-based image tokenizers and to employ the sequence-to-sequence architectures to learn the relationship between textual input and visual output from a large collection of text–image pairs, such as Vector Quantized Variational Autoencoder (VQ-VAE) and Vector Quantized Generative Adversarial Networks (VQ-GAN). VQ-VAE discretely incorporates ideas from vector quantization and encoder network outputs. Then, by pairing these representations with an autoregressive prior, the model with a PixelCNN decoder can generate high-quality images [16]. This model is used by the first vision of DALL-E [17]. More like a variant, VQ-GAN represents a variety of modalities with discrete latent representations by building a codebook vocabulary with a finite set of learned embeddings and using Transformer instead of the PixelCNN in VQ-VAE [10]. Anyway, the PatchGAN discriminator is used to add anti-loss in the training process. The representative work of this modeling is Parti [18]. Different from the above idea, the diffusion-based models, which are built from a hierarchy of denoising autoencoders, start from random noise and gradually denoise them, conditioned on textual descriptions, until images matching the conditional information are generated [19]. Based on the power of diffusion models in high-fidelity image synthesis, the text-to-image system is significantly pushed forward by the recent effort of Disco Diffusion [3], Midjourney [4], Stable Diffusion [5], DALL-E 2 [6], and Imagen [7].

At present, the programs that use diffusion models for a better generation effect, Disco Diffusion, Midjourney, Stable Diffusion, and DALL-E 2, are open to the public, but the programs of Imagen are not. Disco Diffusion is a clip-guided diffusion model that is good at generating pretty abstract art, which can be run in Google Colab now [3]. Midjourney was created by an independent research lab with the same name. It is currently in open beta and is accessible on Discord, where users type in the textual prompt in the chat, and then the artwork is generated by the AI system [4]. Stable diffusion was released by Stability AI in 2022, which uses a latent diffusion mode trained on  $512 \times 512$  images from a subset

of the LAION-5B database. Similar to Google's Imagen, this model uses a frozen CLIP ViT-L/14 text encoder to condition the model to text prompts [20]. Furthermore, it has a better balance between speed and quality and can generate images within seconds [5]. The main novelty of DALL-E 2 seems to be an extra layer of indirection with the prior network, which predicts an image embedding based on the text embedding from CLIP. Specifically, this repository will only build out the diffusion prior network, as it is the best-performing variant [6].

With the emergence of such open-source implementations, the use of advanced text-to-image synthesis for generating images is becoming more widespread, which represents a relevant trend in the AI Art community [21].

## 2.2. Communication between Artists and Audiences

Artistic creation is a process for artists to explore and express ideas and concepts. A great painting has much more below the surface than is first seen on the surface. Therefore, it must access the mind as well as the senses [22]. Similar to how humans do not really know how they breathe, artists do not truly know how they create: while they may rely on a set of fundamental principles, such as how to arrange elements, light, colors, and other components, most of their creative decisions happen intuitively [23]. The experimental result of Eindhoven and Vinacke demonstrated that artists have more control over their creative activities and produce better results than nonartists in the creative process of painting [1]. Kay also found that nonartists, semiprofessional artists, and professional artists differed on certain process-related variables [2].

The interplay between the internal (cognitive) representation and the external (physical) representation is a fascinating problem in cognitive psychology, art, science, and philosophy [24]. The various painting attributes, such as colors, shapes, and boundaries, are selectively redistributed to the brain for processing. For example, color may be experienced as warm or cold or as cheerful or somber [25]. Audiences can also perceive the painter's actions by observing the brushstroke of the painting [26]. Apart from that, from a psychological viewpoint, Kozbelt examined various experiments on artists' perception and depiction skills and showed evidence suggesting possible perceptual differences between artists and nonartists [27,28]. Aesthetic appreciation is an active process influenced by several objective features: external and subjective factors that engage both bottom-up and top-down processes [29]. In the series of studies on experimental aesthetics by Lyu et al. [30–32], the perception of artistic style was affected by individual attributes such as knowledge background and gender. Thus, the perception of art is a complex interaction process between the top and bottom levels, which is affected by various subjective and objective factors.

According to communication theory, the process of artist expression is called encoding, and the way the artwork is perceived by the audience is regarded as decoding [33,34]. Jakobson proposed six constitutive factors with six functions in communication: the addresser, addressee, context, message, contact, and code [34]. For example, an artist (addresser) sends a message to an audience (addressee) through his/her painting. The artist's work, as the message with a story (context), plays a role in the connection between himself/herself and the audience (contact). Finally, his/her message must be based on a shared meaning system (code) by which his/her work is structured [22]. There are three levels of problems, namely technical, semantic, and effectiveness levels, that were identified in the study on the communication of paintings [31,35]. Among them, the technical level focuses on letting the addressee receive a message through visual attraction, and the semantic level requires that the addressee is allowed to understand the message's meaning without misinterpreting it. The effectiveness level concerns the effect of the audience's feelings. During the creative process of AI art, the artists choose AI algorithms according to their intentions for creating the artwork, and audience acceptance is a critical defining step in deciding whether it is "art" [36]. Studying the process of art perception can help build a bridge between artists and the audience [37,38].



### 2.3. Artworks Generated by Human–AI Co-Creation

Artworks are increasingly being created by machines through algorithms with little or no input from humans. At the Christie's auction in 2018, the portrait "Edmond de Belamy", generated by generative adversarial networks (GAN), was auctioned for \$432,500, which indicates that AI has begun to enter our field of vision at a rapid speed [39]. Recent works have addressed a variety of tasks such as classification, object detection, similarity retrieval, multimodal representations, and computational aesthetics, among others [21]. The neural style transfer in which AI technology first intervened in the field of art has been widely used in the platforms such as Prisma, Deep Dream Generator, and other art content production platforms. In 2022, text-to-image AI art generators are much more popular and have been applied to creating conceptual scenes, creative designs, and fictional illustrations. In this case, it can be seen that the processes in various art creations are changing. Meanwhile, some new jobs have also been immediately emerging, such as prompt sale [40].

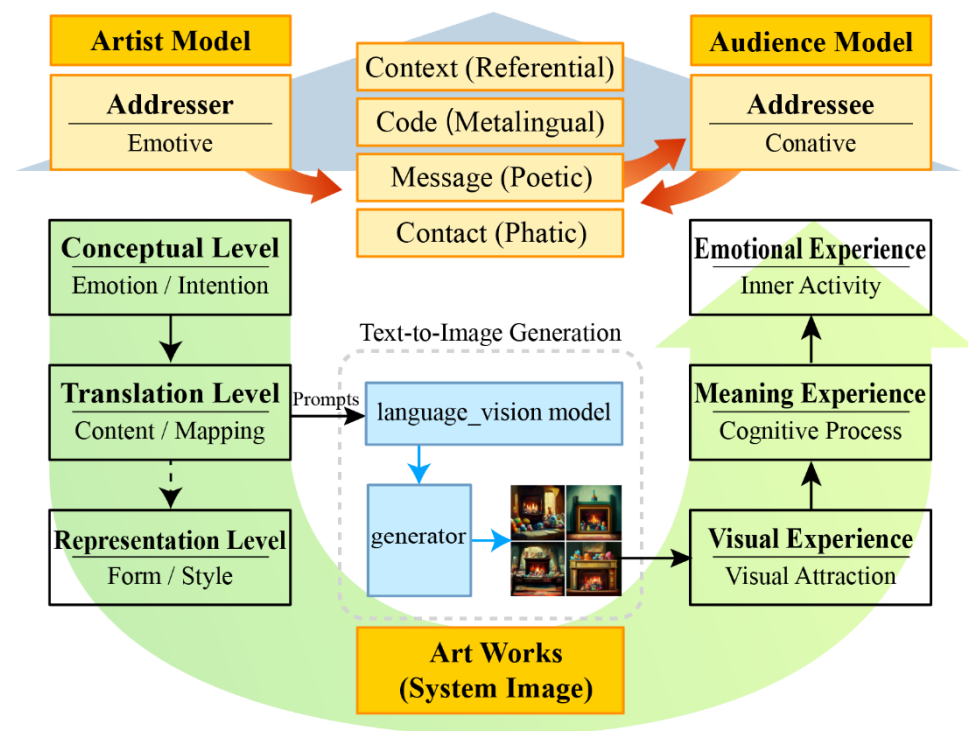
With the explosion of AI-related technologies and their continuous application in the field of art, there is a growing body of research initiatives and creative applications arising at the intersection of AI and art. Artistic creation is embedded with cultural, historical, and institutional frameworks that directly interact with the artist's own creative process [23]. Lacking human consciousness, AI does not understand what it is doing and is merely a suite of statistical models calculating favorable odds through enormous variations. Considering that, AI cannot create art, but it can create patterns that an audience will likely perceive as art [41]. The human artist, as the author, is always the mastermind behind the work, and the computer is a tool [42]. However, AI technology is not like traditional tools. Its randomness changes the way humans control it. As a sparking trigger of inspiration, artists collaborate with AI agencies to augment the artistic process [41].

As for text-based generative art, it is also argued that creativity does not lie in the final artifact but rather in the interaction with the AI and the practices that may arise from the human–AI interaction [43]. It is not hard to imagine a future where text prompts could be generated by language models, thereby completely dehumanizing the creative artistic process and severely distorting the human perception of the meaning behind an image [44]. Most studies reported that visual artworks can be recognized to some extent by humans, especially by experts of a specific art field [45,46], but other experimental results showed that individuals are unable to accurately identify AI-generated artwork [32,47]. Based on our previous research, the deep learning model, trained by large amounts of data on paintings, can simulate human painting skills on the technical level. In contrast, people prefer paintings connecting the semantic and emotional levels [31].

## 3. Materials and Methods

### 3.1. Research Framework

Based on the literature review, in this study, the research framework of communication in the AI painting generated by the text-to-image system was constructed, as shown in Figure 2. In the process of communication between the artist (Addresser) and the audience (Addressee), there is the artist model and the audience model, which construct the complex processing from creation to perception. Different from the traditional coding process, artists translated their intention and emotion into prompts instead of representing them by directly using form. However, existing paintings were taken as the data for training the AI model, which means that the creation path was changed by adding the interaction between human and AI. As for the side in the perception of artworks generated by the AI model, there were still three stages: visual experience, meaning experience, and emotional experience. Ideally, audiences could still contact the artist by receiving the message through decoding and feeling poetic in a referential context.



**Figure 2.** The communication research framework of AI paintings generated by the text-to-image system: The left part is the artist encoding model, and the right is the audience decoding model. The AI generator in the middle is regarded as the communication interface between the artist and the audience.









As the AI generator replaces the represented action of humans, what role do professional art knowledge and experience play in this human–computer interaction process? In the age of AI, what is the critical capability of artists? Instead of fear replacement, it is more important to explore the irreplaceable value of human beings. Therefore, this experiment was designed to discuss process coding and visual perception by comparing the differences in the human–AI interaction between artists and nonartists. The theme “sweet home” was used as the creative theme of the painting, and artists and nonartists were invited to map their inner feeling in visual form by inputting descriptive prompts. Additionally, AI paintings were generated as experiment stimuli by interacting with the text-to-image system. In addition to the analysis of the observations on creative action and open coding from the creator’s self-report, the evaluation items of perception for the stimulus were designed from the visual attributes (technical level), the semantic matching (semantic level), and the emotional experience (effectiveness level). Based on the framework of communication, the study was meant to explore the essence of artistic creation and artists’ unique capabilities by comparing the difference between the two groups in the interaction with the text-to-image system and in the perception of generations.

### 3.2. Stimuli

In the text-to-image system selection stage, an artist and a nonartist were invited to interact with four public text-to-image systems, namely Disco Diffusion, Midjourney, Stable Diffusion, and DALL·E 2. The theme sweet home was chosen as the theme of creation because a person’s home is unique and full of individual imagination and interpretation. They were asked to co-create an oil painting with AI by inputting a prompt to describe the theme. It was suggested that the structure of the prompt should start with “an oil painting of” and should refer to the cases in the community to establish the experience of the relationship between text description and visual generation. In order to eliminate the interference of artistic style, artists’ names and art schools were prohibited. Based on

prompt 1 and prompt 2 provided by the artist and the nonartist, respectively, a comparison was made, which is shown in Table 1. Then, nine art and/or aesthetic background experts were encouraged to select which method was more suitable for generating oil paintings of a sweet home. As a result, they all agreed that the attributes of the generated samples by Midjourney were more similar to those of oil paintings, and the concord of color could express the feeling of a sweet home on the effectiveness level [25]. Additionally, its degree of matching with text descriptions was much higher than that of the other two systems. Among them, Disco Diffusion confused the structure of elements and the canvas layout, while Stable Diffusion had an adequate understanding close to Midjourney but missed the artistic oil-painting style. Beyond that, DALL·E 2 better understood the feeding text, whereas its unity of tone was slightly weaker than Midjourney. Therefore, Midjourney was picked as the AI tool to collaborate with two group creators to generate paintings as experimental samples.

**Table 1.** The results generated by four text-to-image systems: Each system was set to generate four images. Result 1 was generated by Prompt 1, and Result 2 was generated by Prompt 2.

Methods	Disco Diffusion	Midjourney	Stable Diffusion	DALL·E 2
Source	<a href="https://github.com/alembics/disco-diffusion">https://github.com/alembics/disco-diffusion</a> (accessed on 10 June 2022)	<a href="http://www.midjourney.com">www.midjourney.com</a> (accessed on 25 August 2022)	<a href="https://beta.dreamstudio.ai/dream">https://beta.dreamstudio.ai/dream</a> (accessed on 2 September 2022)	<a href="https://labs.openai.com">https://labs.openai.com</a> (accessed on 29 September 2022)
Prompt 1	An oil painting of a room full of toys by the fireplace.			
Result 1				
Prompt 2	An oil painting of a father reading a newspaper in front of the computer, a mother cooking in the kitchen, a little son sitting on the sofa watching the cartoon named tom and jerry, and a big daughter just bringing a golden retriever into the room.			
Result 2				

In the experimental sample-generation phase, a total of ten artists and ten nonartists participated in theme painting creation by interacting with Midjourney, whose information is displayed in Table 2. In selecting creators, the following criteria were used to distinguish artists from nonartists: An artist should have painting experience and should have to derive some income from their pictures. A nonartist was any subject who had not engaged in this type of creative activity. Before the experiment, they had never used similar tools to assist in painting and had only heard of the power of AI.

**Table 2.** The information of the artists and nonartists: the age and painting years of the artists and the age of the nonartists were listed. The label “AP01” represents the artist that created the P01 painting, while “NH01” is the nonartist who created the H01 painting.

Artists	AP01	AP02	AP03	AP04	AP05	AP06	AP07	AP08	AP09	AP10
Age (Years)	39	41	22	22	40	23	23	35	41	43
Painting Experience (Years)	9	15	12	12	24	7	4	13	20	23
Nonartists	NH01	NH02	NH03	NH04	NH05	NH06	NH07	NH08	NH09	NH10
Age (Years)	38	67	40	42	49	23	22	26	25	42

They were asked to write a prompt describing a visual form that could express their imagination of a sweet home. The basal commands in Midjourney were to use the V1, V2, V3, or V4 buttons to create variations of their chosen image and then to click the U1, U2, U3, or U4 buttons to add details to the chosen image. To avoid interference due to unfamiliarity with the tools, the researcher observed and supported the whole process but did not affect the participants' writing and selection. Individual differences were so great as to suggest that each person attained their final product in their own way. Finally, the nine experts mentioned above filtered through six samples from each group by excluding samples that were similar. Twelve paintings (P01–P06 by the artists, and H01–H06 by the nonartists) are listed in Table 3. In addition, a painting created in the 1980s by artist Yong Wang on the topic of a sweet home also was chosen as the thirteenth stimulus, functioning as the reference sample. This painting recorded his poor kitchen environment at a time when his wife was busy cooking for the whole family. The limited living environment and his wife's busyness form an artistic conflict, highlighting that the inner sweetness is the critical value of a home. Furthermore, the stimuli for this experiment were classified into three types according to the research purpose.

**Table 3.** The thirteen paintings and prompts: there are three groups including Midjourney + artist paintings (P01–P06), Midjourney + nonartist paintings (H01–H06), and artist painting.














Type		Midjourney + Artist Paintings		
No.		P01	P02	P03
Prompt		An oil painting of a room full of toys by the fireplace.	An oil painting of love harbor full of laughter and warmth.	An oil painting of parents happily walking in the park hand in hand, and an active dog is chasing me.
Painting				
No.		P04	P05	P06
Prompt		A warm tone oil painting of a little pink bear holding a honey jar to enjoy the cool under the shade of the big tree in front of the yellow wooden house, and beautiful flowers and grass, and gurgling streams beside the wooden house on a bright summer day.	A warm tone oil painting of mother toasting bread for her daughter in a Europe style room.	An oil painting of a Samoyed dog with a space helmet and a space suit floating in outer space.
Painting				



Table 3. Cont.

Type		Midjourney + Nonartist Paintings		
No.		H01	H02	H03
Prompt		An oil painting of one family, balloons, toys and food in amusement park.	An oil painting of a family playing in the yard of a house, also including trees, sun, birds.	An oil painting of kids playing, cat napping, and parents cooking while chatting.
Painting				
No.		H04	H05	H06
Prompt		An oil painting of a two-and-a-half floor house with red roofs and gray walls, surrounding with a beautiful garden full of plants and flowers, and a crystal-clear stream flowing through the garden.	An oil painting of a family having dinner and a fish in the center of the table.	An oil painting of a father reading a newspaper in front of the computer, a mother cooking in the kitchen, a little son sitting on the sofa watching the cartoon named tom and jerry, and a big daughter just bringing a golden retriever into the room.
Painting				
Type		Artist Painting		
Painting			Description	Smoke curls up from the kitchen, roosters look for food, and the simple open-air kitchen emits the smell of cooking.

### 3.3. Experiment Procedures

During the creative process, the observer recorded the cost time of each creator, the number of adjustments to the statements, and the number of times the U button was clicked for variations. After the creators submitted the paintings co-created with Midjourney, they had a one-on-one interview to self-report their experience and to think about the interaction process and results. Then, the recordings were coded to discuss the difference in the process of human–AI interaction.

As for the perceptual evaluation of stimuli, forty-two participants with artistic backgrounds were recruited into the questionnaire survey. A PDF file containing a QR code link to the online questionnaire and to the thirteen samples was emailed to them. In addition, the requirements that each slide should be viewed on a computer screen no less than 14 inches for more details and that the online questionnaire should be filled in after scanning the QR code on the mobile phone were highlighted. A painting was displayed randomly on each slide with its prompt for rating followed by all of the paintings being displayed for ranking. Finally, 42 valid data were received for statistical analysis.



### 3.4. Questionnaire Participants


A total of 42 subjects (15 males and 27 females) participated in the experiment. About 47% were 20~30 years old; 17% were 31~40 years old; 14% were 41~50 years old; 17% were 51~60 years old; and 5% were over 61 years old, indicating a relatively even distribution of age groups apart from the youngest group. In terms of professions, they all had experience in painting or art research, so the questionnaire data can be featured with the reliability.

The participants were asked to rate the paintings' degree of each attribute and to rank them according to their subjective aesthetic experience. The procedure is described below in detail.














### 3.5. Questionnaire Design

The questionnaire comprised two parts. Part one was a rating test in which the participants should provide subjective ratings for the thirteen paintings on nine visual attributes, as described in Table 4. The evaluation attributes belonged to three levels: the technical level (f1–f3), the semantic level (f4–f6), and the effectiveness level (f7–f9). The items explored the perceptive degree of the attributes in the paintings, and subjects scored the responses using a 5-point Likert scale from 1 (“Very low”) to 5 (“Very high”). In part two (ranking test), the subjects were asked about their most preferred painting and attribute (see Table 5).

**Table 4.** Part one: questionnaire for subjective ratings of paintings on the nine attributes.

Painting	Attributes	1 2 3 4 5
 <p>P01: An oil painting of a room full of toys by the fireplace.</p> <p>Please subjectively rate each painting according to visual attributes, with a maximum of 5 points and a minimum of 1 point.</p>	f1. Color harmony	
	f2. Element accuracy	
	f3. Layout coordination	
	f4. Tone matching	
	f5. Content matching	
	f6. Scene matching	
	f7. Sweetness	
	f8. Creativity	
	f9. Preference	

**Table 5.** Part two: questionnaire for subjective rankings of paintings.

Please Select One Painting	Paintings
Which one is the most professional?	    
Which one is the sweetest?	    
Which one is the most creative?	  
Which ones are the creations of artists?	

### 3.6. Statistical Analysis

Based on the observation data, the time spent by artists and nonartists and the number of interactions was recorded. For the reflections obtained from the interview, the grounded theory method was used to code the opening data. For the rating data in the questionnaire, descriptive statistics and ANOVA were firstly adapted to test whether there was a significant difference between the three types of paintings. For items reaching the significance level, we used the Duncan multiple comparative methods to test whether there was a significant difference among the three averages. In addition, multidimensional preference

analysis (MDPREF) was performed to determine the relationships between stimulus and attributes. Finally, percent statistics and Chi-square were used to analyze the raking data.

## 4. Results

### 4.1. Coding of Reflections in Human–AI Co-Creation

According to the results of the variation analysis in Table 6, after the two groups of creators co-created with Midjourney, they displayed a significant difference in their behavior during the time spent, the number of modified prompts, and the number of clicked U buttons. The average time spent by artists was 22 min, which was significantly higher than the 14 min spent by nonartists. Apart from that, artists tried more than 6 times to modify the prompts and averaged about 4 U-button clicks for repeated attempts, which was far higher than the activity frequency of nonartists. Apparently, there were still obvious differences between the two groups in the co-creation of AI.

**Table 6.** The distribution of time spent, number of modified prompts, and number of U-button clicks during painting in Midjourney for ten artists and ten nonartists.

	Two Groups of Creators		Significance
	Artists ( <i>n</i> = 10)	Nonartists ( <i>n</i> = 10)	
Time spent (Min.)	22 ± 4.25	14 ± 4.25	***
Number of modified prompts	6 ± 2.16	4 ± 2.46	*
Number of U-button clicks	10 ± 2.95	3 ± 1.52	***

\*  $p < 0.05$ ; \*\*\*  $p < 0.001$ .

Reflections obtained from the two groups of unstructured interviews were coded with grounded theory methods in three steps: (a) initial open coding, (b) intermediate coding, and (c) advanced coding [48]. First, the essence of the interview recordings was synthesized during the initial coding step. Next, new codes focused on similarities and differences were formulated, and selective codes were developed. Finally, the codes were intermediated into six core categories, as can be seen in Table 7.

All of the creators in this experiment used Midjourney to generate paintings for the first time. The coding results showed that creators with artistic backgrounds paid more attention to such core categories, such as visual performance, semantic matching, subject control in the interaction mode, and creative stimulation in creation experience, whereas the nonartists focused on the semantic matching and culture cognition. In the category of technological ethics, there were some different thoughts.

### 4.2. Descriptive Statistics and ANOVA Analysis of Rating Data

The purpose of this study was to find out whether there are any differences in the perception of co-creation paintings with AI between creators with and without artistic backgrounds. According to the results of the variation analysis in Table 8, after the subjects viewed the three types of paintings, no significant difference was shown on the technical level (i.e., “Color harmony”, “Element accuracy”, and “Layout coordination”), the semantic level (i.e., “Element accuracy”, “Content matching”, and “Scene matching”), or the effectiveness level (i.e., “Creativity” and “Preference”), which demonstrated that the perception effect of painting technology, semantic matching, artistic creativity and preference were similar among three types of paintings. It is worth noting that there were significant differences in the option of “Sweetness” ( $p < 0.001$ ). The scores of AI generation by artists (3.43 points) and nonartists (3.45 points) were significantly higher than that by the artist Yong Wang (2.68 points), which related to how subjects communicate with paintings.

**Table 7.** The codes used to code the creators' reflections on co-creation with AI: ▲ was used to mark the feedback from the artists; ■ indicates the feedback from the nonartists; and ● represents the feedback from both of the two groups.

Core Category	Selective Coding	Open Coding
Visual performance	Artistic style	▲ Paintings generated by AI can be identified because of high standardization (AP05, AP09–10). ▲ The visual style lacks a uniqueness (AP05, AP18, AP10).
	Techniques	▲ The color is very harmonious (AP01, AP06–07). ▲ The strokes are rich and vivid (AP04).
Semantic matching	Element accuracy	■ Did not generate elements accurately based on the prompt (NH04). ● There are some mistakes in element positions (AP01, AP06–07, AP09, AH06).
	Space attributes	● When prompts are complex, some elements are usually lost (AP09, NH02, NH06). ● The generated space layout has deviation (AP04, AP09, NH06).
	Expression characters	▲ In some results, the animal state was a little decadent, which did not meet the prompt (AP06).
Human–AI interaction	Prompt restrictions	● Some naughty words are banned (AP03, NH06). ▲ Unlike traditional brushes and paints, they can help you realize that what you think is what you get, and it is completely under your control (AP01, AP05–07).
	Subject control	▲ More iterations can make the results closer to inner thoughts (AP03–05). ● Prompting rules are related to the final generated effect to a great extent (AP02–04, AP06, NH02–04).
	Prompt grammar rules	● Any small difference in prompts would cause disparate generation (AP01–08, NH01–10).
Creation experience	Creation assistance	▲ There are still differences in using language to express emotions instead of painting, even though all the elements described are generated (AP07–09). ■ Generated some fantastic images that I just imagine but cannot draw (NH01/, NH04–08, NH095).
	Creative generation	▲ Compared with the result of matching with prompts, some unexpected surprise is preferred (AP02, AP04). ▲ It is like Pandora's Box. If it is not a surprise, it may be a shock (AP01, AP06).
Culture cognition	Cross cultural differences	■ The originally generated image is full of Indian style home decorations with cultural differences (NH03).
Technological ethics	Work displacement	▲ AI cannot generate my unique styles and cannot replace senior painters (AP10). ▲ A little confused about own core competitiveness (AP06–07). ● Maybe some work related to painting will be impacted by AI (AP06, NH05).
	Copyright issues	▲ Due to the mixture and collage of painting styles, the ownership of copyright is a complex issue (AP01, AP03–07).

**Table 8.** Results of descriptive statistics and ANOVA analysis. They compare whether there are perceptual differences among the three types of paintings.














Subjective Questionnaire (1–5 Points)	Sweet Home Paintings			Significance
	Midjourney + Creator with Art Background	Midjourney + Creator without Art Background	Artist	
F1. Color harmony	3.96	4.00	3.79	
F2. Element accuracy	3.76	3.71	3.89	
F3. Layout coordination	3.69	3.71	3.66	
F4. Tone matching	3.83	3.83	3.95	
F5. Content matching	3.64	3.74	3.97	
F6. Scene matching	3.66	3.80	4.05	
F7. Sweetness	3.43 <sup>a</sup>	3.54 <sup>a</sup>	2.68 <sup>b</sup>	***
F8. Creativity	3.38	3.38	3.32	
F9. Preference	3.36	3.36	3.03	

\*\*\*  $p < 0.001$ ; <sup>a,b</sup> are Duncan ex-post test grouping results.

#### 4.3. MDPREF Analysis of Rating Data in Attribute Vectors

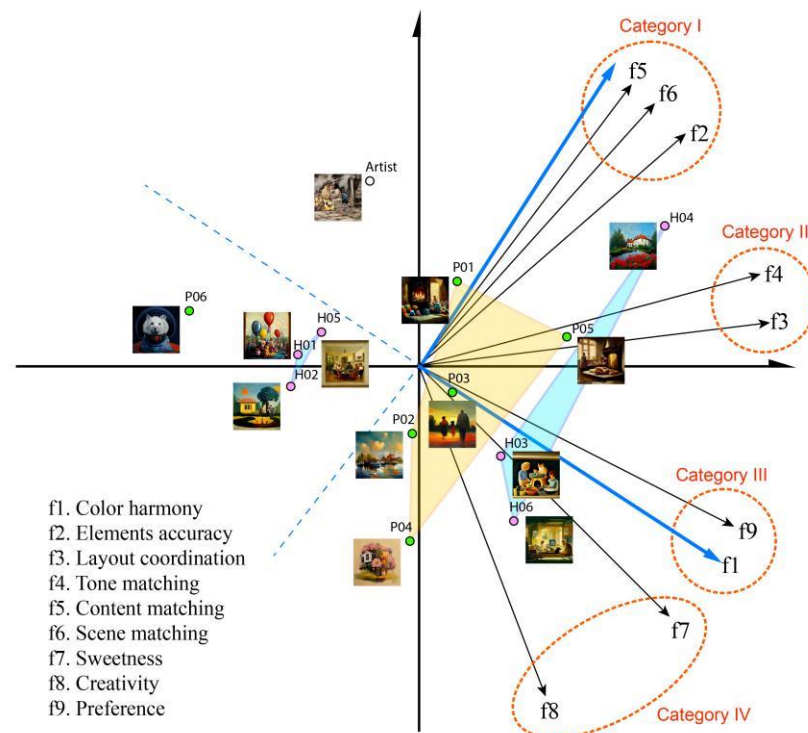
The cognitive space was set up by conducting a multidimensional preference analysis (MDPREF) which expressed the relationship between the stimuli and their attributes. A matrix was created from the raw data to illustrate the mean scores of the nine fundamental relations in each of the thirteen paintings, as shown in Table 9. The matrix allowed SPSS statistics software to compute MDS and generate a two-dimensional (2D) spatial plot demonstrating the relationship between two crucial correspondence indications. Kruskal's stress was 0.14589, which was less than 0.2, and the determination coefficient (RSQ) was 0.92544, which was close to 1.0, revealing that the spatial relationships between the thirteen paintings and nine attributes could be appropriately represented in 2D. Moreover, the stress index indicated that the 2D plot and the original data exhibited a satisfactory fit, while the RSQ denoted that the 2D plot could explain 90.92% of the variance [49]. The cognitive matrix is shown in Figure 3.

**Table 9.** Average score rating in nine perceptual attributes: the highest score of each attribute was marked in a red color, and the lowest score was marked in blue a color.

													
	P01	P02	P03	P04	P05	P06	H01	H02	H03	H04	H05	H06	Artist
F1	3.91	4.10	3.76	4.12	4.21	3.74	3.86	3.55	4.19	4.07	3.93	4.38	3.83
F2	3.52	3.71	3.83	3.62	4.05	3.69	3.69	3.45	3.55	4.43	3.50	3.52	3.88
F3	3.76	3.93	3.76	3.60	3.79	3.31	3.41	3.33	3.81	4.31	3.55	3.83	3.67
F4	3.91	3.69	3.91	3.86	4.05	3.60	3.55	3.64	3.86	4.29	3.76	3.95	3.91
F5	4.02	3.45	3.69	3.17	3.95	3.45	3.55	3.62	3.76	4.36	3.55	3.43	4.00
F6	3.98	3.50	3.64	3.31	4.02	3.41	3.60	3.67	3.76	4.21	3.76	3.69	4.05
F7	3.45	3.45	3.67	3.76	3.67	2.48	3.31	3.55	3.48	3.55	3.29	3.83	2.64
F8	3.14	3.24	3.41	3.71	3.31	3.52	3.14	3.17	3.88	3.29	3.02	3.79	3.29
F9	3.26	3.36	3.31	3.48	3.62	3.02	3.00	3.14	3.64	3.64	3.00	3.71	3.00

According to the distribution of visual vectors in Figure 3, the nine visual attributes can be grouped into four categories: category I included the visual attributes of “Element accuracy (f2)”, “Content matching (f5)”, and “Scene matching (f6)”; “Layout coordination (f3)” and “Tone matching (f4)” belonged to group II; and “Color harmony (f1)” and “Preference (f9)” were in group III; while in group IV, “Sweetness (f7)” and “Creativity (f8)” were individually separated. The vector of attribute f7 (Sweetness) intersected with category I at nearly 90°. Based on the MDPREF analysis, the attribute vectors of semantic matching were irrelevant to sweetness and creativity.

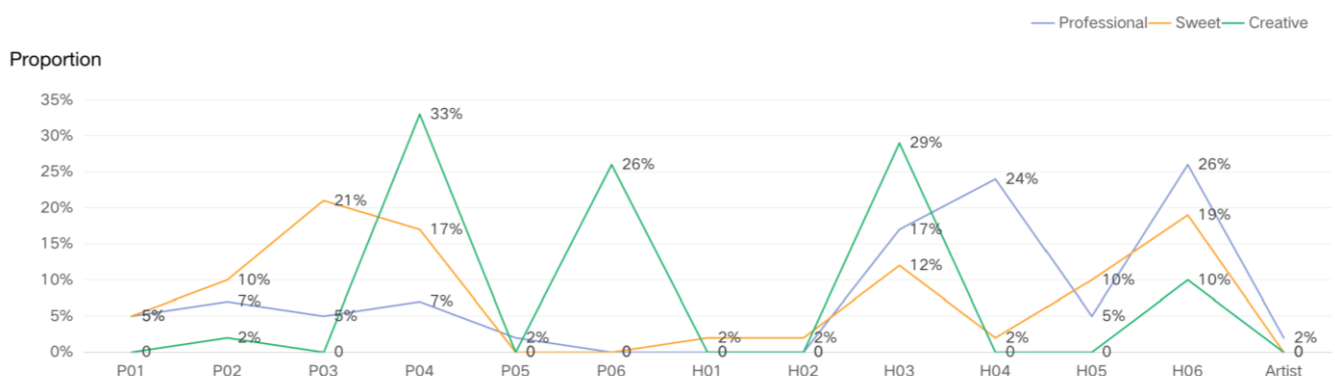
The thirteen paintings were presented in the cognitive space of preferences in the form of point coordinates. The locations of stimulus paintings that were grouped together represented that they had a similar rating, while the locations of stimulus paintings that were separated represented that the paintings held different attributes. Each painting could be projected onto every attribute vector. According to the distribution of paintings in Figure 3, the most generations interacted with by AI and creators with artistic backgrounds (P01–P05) could be projected onto the positive pole of most attribute vectors, whereas P06 was far away from others of the same type and had more negative perceptions. Furthermore, the paintings co-created by AI and nonartists were located in three clusters. H03 and H06 had higher perceptions of high-level attributes, and H04 was better on low-level attributes. In contrast, H01, H02, and H05 gathered and projected onto the negative pole of all the attribute vectors. As for the reference sample, the paintings by artists performed better on semantic matching.



**Figure 3.** Perceptual matrix of nine visual attributes and thirteen paintings. The points in the space stand for stimuli, while their distance indicates their difference. The attribute vectors were labelled as f1–f9.

#### 4.4. Analysis of Subjective Ranking

To further determine whether there were perceptual differences among the three types of paintings, in this study, subjects were invited to choose what they considered to be the most professional, sweet, and creative painting. Finally, the work that they thought was most like human paintings was picked. Figure 4 shows the proportion of people selecting the most professional, sweet, and creative painting among all the subjects. As for the professional aspect, the top three paintings were H06 (26%), H04 (24%), and H03 (17%); while considering the sweet aspect, the order was P03 (21%), H06 (19%), and P04 (17%); and in the creativity aspect, the top three were P04 (33%), H03 (29%), and P06 (26%).



**Figure 4.** Proportion of each painting being selected as the most professional, sweet, and creative one: the x-axis represents three groups of samples, and the y-axis shows the percentage of votes.




A Chi-Square test was conducted to analyze the differences in the subjective ranking of professional, sweet, and creative aspects and to analyze the selection of the human painting according to age, gender, and education. Only female and male subjects had significant



differences in the selection of which one was the human painting. Since the number of some samples selected was less than five people, the exact probability method was adopted to calculate the Chi-Square value  $\chi^2 = 18.891$ ,  $p < 0.05$ . The proportion of female subjects choosing P03 and P04 was obviously higher than the average of 64.29%, while males preferred to choose H04 and H06, which was higher than the average of 35.71%.

Table 10 shows the top three paintings that the subjects thought were most like those created by humans. The order was H04 (21%), P03 (13%), and Artist (13%). In a combined interview with the participants, the clues that affected their judgment included various details, such as the stroke and texture in H04 and P03, as well as a structure and tone style similar to the textbook in the artist's painting.

**Table 10.** Proportion of top three paintings thought to be most like a creation by humans: from left to right, the number of votes is from high to low.

Question	The Top Three		
Which one is the creation by an artist?			
	H04 (21%)	P03 (13%)	Artist (13%)

## 5. Discussion

### 5.1. Differences of Coding in Co-Creation with AI

According to the action observation data, the artists still kept their behavior characteristics, which differed from nonartists, in the creative process [1,2], such as more control over tools and repeated actions. Even in the process of interaction with AI, actions different from those of nonartists still existed. However, it can be seen from the interview data that artists were not satisfied with the control effect of AI, and they even felt a little out of control. The artists' attitude towards technology was related to their experience. The artists (AP05, AP09–10) with more painting experience claimed that they could identify the paintings generated by AI due to some similarities and firmly believed that they would not be replaced. However, the creators with relatively little painting experience had contradictory attitudes toward AI. On the one hand, they affirmed the professionalism of AI paintings in terms of color and brush strokes, and they felt that the paintings could generate some surprise even though they were not being very obedient. On the other hand, they considered the possibility of potential competition and had some confusion about core ability. Additionally, some artists were surprised by accidents and thought that they had control of their creativity, although their paintings were different from the descriptive text, such as sample P04, while others (AP01, AP05–07) felt a loss of control of the AI compared with traditional tools. Based on the analysis of the prompt, more artists used metaphors instead of direct descriptions of real-life scenes and constantly sought the vision that they wanted by iteration. For example, AP03 imagined home as a harbor of love, and the P06 creator compared herself to a Samoyed dog and stated that floating in the endless space was the sweetest destination. It can be seen that metaphors, as the basic mechanism of art, were still widely used in the coding process of artists and artificial intelligence. Generally, in the process of interaction with AI, artists still kept the original parts during creation. However, unlike traditional tools, the loss of control may bring surprise or fright [23,36]. Moreover, due to their different experiences and skills, they had different attitudes toward AI.

As for most nonartists, their creative process was simple and direct, and they were generally excited about a series of excellent results. They preferred to depict certain people in a scene based on their memory or hope. The work of H06, for instance, restored the

author's childhood memory of watching the cartoon *Tom and Jerry*, and H02 depicted the author's expectation of their grandson's arrival in the future. AI as an interface helped the crowd of people without painting skills to visualize their imagination (NH01, NH04–08, NH09). Considering that, there is an example of this point. NH06 generated an Indian painting, but as a Chinese man, it was difficult for him to resonate and feel any sweetness. Instead of focusing on artistic techniques and creativity, they were more focused on semantic matching and cultural consistency.

To sum up, there were differences in actions between artists and nonartists as well as differences in their attitudes and concerns that were influenced by personal knowledge. Ultimately, the text-to-image system has introduced a new human–AI interaction mode as a transformation interface from internal imagination to visual form. Due to the randomness and variation of AI generation, artists gradually lose confidence in the ability to control tools like before.

### 5.2. Differences in Decoding in Communication with Creators

Except for the perception of sweetness, most attributes had no significant difference in scores, which showed that co-creation with the text-to-image system really reduced the function of painting ability in artworks. The assistance of AI not only made the perception of human–AI co-creation with and without artistic background converge, but also blurred the difference between AI generation and human painting. It is worth noting that there were significant differences in the perception of sweetness and that the score of the artist's painting was much lower than that of the AI generator. It seems that, as the audience could not decode the effectiveness level without the Yong Wang's life experience in the countryside in the 1980s, they could not feel the sweetness of the painting.

Combining the rating score and the distribution of the thirteen paintings in the perceptual matrix, more samples (P01–05) were created by the collaboration of Midjourney and the creators with artistic backgrounds projected onto the positive direction of most attribute vectors, whereas generations without artistic backgrounds were divided into two extremes. Additionally, the result indicated that, owing to art expertise, the communication between the artist and the audience could be more stable unless the coding with a strong personal thinking or experience system was too difficult to understand and could not resonate with audiences, such as the space dog in P06 and the outdoor cooking in the artist Yong Wang's painting. As for the nine attributes, the cognition for the accuracy of element shaping in painting was closely correlated with content and scene matching with prompts, which demonstrated the process from shape to meaning. In addition, the perception of color harmony grouped with sweetness and preference did not relate to semantic matching. Color could express feeling on the effectiveness level [25] even though the paintings failed in structure and significance. This was also the reason to select the Midjourney instead of the other systems. Thus, color perception was an important channel for feeling the degree of sweetness and affecting the preference. Apart from that, semantic matching did not seem to be closely related to high-level perception. As the prompt of the artist sample was obtained based on the painting description, of course, the score of the attributes at the semantic level was higher, but the perception at a high level was still lower. On the contrary, although P04 failed in semantic matching, the special combination could still impress the audience with its sweetness and creativity. Furthermore, the audience model was an active process influenced by several subjective features [29,35]. Subjects usually used their cognitive system to decode the meaning of the painting so that the results generated based on text did not affect their perception of high-level features because of the high semantic matching. The fitness degree of prompts affected the artists' perception of the AI control ability.

The ranking result demonstrated that more subjects considered the AI productions as more professional than the painting by the artist Yong Wang, and even the samples created by nonartists obtained the most votes. AI technology was able to imitate artistic presentation techniques very well, although it only relied on the features' statistics without knowing the image's intention [31]. P03, as the sweetest painting, showed the artist's skill

in transmitting emotion through visual information. Additionally, creativity could still be handled by the group with artistic backgrounds. Although the rating score of P04 and P06 in the nine terms was not high, their unique representation, different from ordinary thinking, improved the perception of creativity. However, to enable the audience to decode and communicate with artists successfully, it is not enough to rely solely on creativity, and links in culture, experience, and other aspects are also required [41]. As for the differences of gender in the selection of artists' paintings, although there was evidence showing gender differences in style perception [31], considering the small sample size of this experiment, it is appropriate to discuss it in future general test research.

AI algorithms have simulated excellent visual patterns, similar to traces of drawings by humans. Through interaction with technology such as text-to-image systems, nonartists can express their creativity by breaking the limitations of their drawing skills. Artists must face the narrowing distance in technical skills with people featuring nonartistic backgrounds. Therefore, a high level of communication with the audience should be paid more attention.

## 6. Conclusions

Understanding how humans collaborate with AI and perceive the generated results is complex and necessary in the age of machine learning. From the perspective of art communication, this study explored the difference in coding in co-creation and decoding in perception with a text-to-image system between artists and the nonartists. Furthermore, the overall conclusion of the present research can fall into two parts: Firstly, the actions and reflections of the creators supported the view that the action characteristics of artists were still different from those nonartists as well as that their attitudes and concerns were related to their knowledge. Secondly, AI blurred the differences in painting techniques enhanced through professional training, whereas stable performance in art action was strictly tied to experience in creation. Additionally, the evidence of the perception of human–AI co-creation suggested that it is necessary to pay attention to emotional communication above the form of formal features and semantic matching in the interaction with AI technology.

This study had several limitations. Firstly, the painting samples in this study were all displayed on a digital screen, which was different from the feeling of watching an offline exhibition. However, with the development of the metaverse concept and the significant impact of COVID-19, virtual reality space will be a new trend for showing paintings in the future. Secondly, since there was not a wide range of age involved in this study, the results were more applicable to 20 to 30 years old adults. In this case, in the future, the research team will balance the age distribution and cover various professional backgrounds to further understand the differences in the perception of AI art between different subjects. Thirdly, considering that there were only 42 subjects in each experiment in this study, a more general conclusion could be obtained if the number of subjects is increased.

**Author Contributions:** Conceptualization, Y.L.; formal analysis, Y.L.; original draft, Y.L.; editing investigation, Y.L.; resources, Y.L.; methodology, X.W.; writing—review, X.W., R.L. and J.W.; writing—editing, J.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Beijing Municipal Education Commission, NO. SM202110011005.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data sharing not applicable.

**Acknowledgments:** The authors would like to appreciate the experts and participants that took part in the experiments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Eindhoven, J.E.; Vinacke, W.E. Creative processes in painting. *J. Gen. Psychol.* **1952**, *47*, 139–164. [CrossRef]
2. Kay, S. The figural problem solving and problem finding of professional and semiprofessional artists and nonartists. *Creat. Res. J.* **1991**, *4*, 233–252. [CrossRef]
3. Disco Diffusion. Available online: <https://github.com/alembics/disco-diffusion> (accessed on 10 June 2022).
4. Midjourney. Available online: [www.midjourney.com](http://www.midjourney.com) (accessed on 25 August 2022).
5. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695.
6. Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical text-conditional image generation with clip latents. *arXiv* **2022**, arXiv:2204.06125.
7. Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.; Ghasemipour, S.K.S.; Ayan, B.K.; Mahdavi, S.S.; Lopes, R.G.; et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *arXiv* **2022**, arXiv:2205.11487, 2022.
8. State Fair’s Website. Available online: <https://coloradostatefair.com/wp-content/uploads/2022/08/2022-Fine-Arts-First-Second-Third.pdf> (accessed on 25 August 2022).
9. Gu, S.; Chen, D.; Bao, J.; Wen, F.; Zhang, B.; Chen, D.; Yuan, L.; Guo, B. Vector quantized diffusion model for text-to-image synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10696–10706.
10. Crowson, K.; Biderman, S.; Kornis, D.; Stander, D.; Hallahan, E.; Castricato, L.; Raff, E. Vqgan-clip: Open domain image generation and editing with natural language guidance. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 88–105.
11. Lee, H.; Ullah, U.; Lee, J.S.; Jeong, B.; Choi, H.C. A Brief Survey of text driven image generation and manipulation. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), Gangneung, Korea, 1–3 November 2021; pp. 1–4.
12. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G. Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning, Virtual Event, 18–24 July 2021; pp. 8748–8763.
13. Zhou, K.; Yang, J.; Loy, C.C.; Liu, Z. Learning to prompt for vision-language models. *Int. J. Comput. Vis.* **2022**, *130*, 2337–2348. [CrossRef]
14. Liu, V.; Chilton, L.B. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In Proceedings of the CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 30 April–5 May 2022; pp. 1–23.
15. Wu, Y.; Yu, N.; Li, Z.; Backes, M.; Zhang, Y. Membership Inference Attacks Against Text-to-image Generation Models. *arXiv* **2022**, arXiv:2210.00968.
16. Van Den Oord, A.; Vinyals, O. Neural discrete representation learning. In Proceedings of the Neural Information Processing Systems Annual Conference, Long Beach, CA, USA, 4–9 December 2017; pp. 1–10.
17. Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; Sutskever, I. Zero-shot text-to-image generation. In Proceedings of the International Conference on Machine Learning, Virtual Event, 18–24 July 2021; pp. 8821–8831.
18. Yu, J.; Xu, Y.; Koh, J.Y.; Luong, T.; Baid, G.; Wang, Z.; Vasudevan, V.; Ku, A.; Yang, Y.; Ayan, B.K.; et al. Scaling autoregressive models for content-rich text-to-image generation. *arXiv* **2022**, arXiv:2206.10789.
19. Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 2256–2265.
20. Stable-Diffusion. Available online: <https://github.com/CompVis/stable-diffusion> (accessed on 2 September 2022).
21. Cetinic, E.; She, J. Understanding and creating art with AI: Review and outlook. *ACM Trans. Multimed. Comput. Commun. Appl.* **2022**, *18*, 1–22. [CrossRef]
22. Lin, C.L.; Chen, J.L.; Chen, S.J.; Lin, R. The cognition of turning poetry into painting. *J. US-China Educ. Rev. B* **2015**, *5*, 471–487.
23. Audry, S. *Art in the Age of Machine Learning*; MIT Press: Cambridge, MA, USA, 2021; pp. 30, 158–165.
24. Solso, R.L. *Cognition and the Visual Arts*; MIT Press: Cambridge, MA, USA, 1996; pp. 34–36.
25. Steenberg, E. Visual Aesthetic Experience. *J. Aesthet. Educ.* **2007**, *41*, 89–94. [CrossRef]
26. Taylor, J.; Witt, J.; Grimaldi, P. Uncovering the connection between artist and audience: Viewing painted brushstrokes evokes corresponding action representations in the observer. *J. Cogn.* **2012**, *125*, 26–36. [CrossRef] [PubMed]
27. Kozbelt, A. Gombrich, Galenson, and beyond: Integrating case study and typological frameworks in the study of creative individuals. *Empir. Stud. Arts* **2008**, *26*, 51–68. [CrossRef]
28. Kozbelt, A.; Ostrofsky, J. Expertise in drawing. In *The Cambridge Handbook of Expertise and Expert Performance*; Ericsson, K.A., Hoffman, R.R., Kozbelt, A., Eds.; Cambridge University Press: Cambridge, UK, 2018; pp. 576–596.
29. Chiarella, S.G.; Torromino, G.; Gagliardi, D.M.; Rossi, D.; Babiloni, F.; Cartocci, G. Investigating the negative bias towards artificial intelligence: Effects of prior assignment of AI-authorship on the aesthetic appreciation of abstract paintings. *Comput. Hum. Behav.* **2022**, *137*, 107406. [CrossRef]
30. Lyu, Y. A Study on Perception of Artistic Style Transfer using Artificial Intelligence Technology. Unpublished Doctor’s Thesis, National Taiwan University, Taipei, Taiwan. 2022. Available online: <https://hdl.handle.net/11296/grdz93> (accessed on 23 October 2022).

31. Lyu, Y.; Lin, C.-L.; Lin, P.-H.; Lin, R. The Cognition of Audience to Artistic Style Transfer. *Appl. Sci.* **2021**, *11*, 3290. [\[CrossRef\]](#)
32. Sun, Y.; Yang, C.H.; Lyu, Y.; Lin, R. From Pigments to Pixels: A Comparison of Human and AI Painting. *Appl. Sci.* **2022**, *12*, 3724. [\[CrossRef\]](#)
33. Fiske, J. *Introduction to Communication Studies*, 3rd ed.; Routledge: London, UK, 2010; pp. 5–6.
34. Jakobson, R. *Language in literature*; Harvard University Press: Cambridge, MA, USA, 1987; pp. 100–101.
35. Lin, R.; Qian, F.; Wu, J.; Fang, W.-T.; Jin, Y. A Pilot Study of Communication Matrix for Evaluating Artworks. In Proceedings of the International Conference on Cross-Cultural Design, Vancouver, BC, Canada, 9–14 July 2017; pp. 356–368.
36. Mazzone, M.; Elgammal, A. Art, creativity, and the potential of artificial intelligence. *Arts* **2019**, *8*, 26. [\[CrossRef\]](#)
37. Gao, Y.-J.; Chen, L.-Y.; Lee, S.; Lin, R.; Jin, Y. A study of communication in turning “poetry” into “painting”. In Proceedings of the International Conference on Cross-Cultural Design, Vancouver, BC, Canada, 9–14 July 2017; pp. 37–48.
38. Gao, Y.; Wu, J.; Lee, S.; Lin, R. Communication Between Artist and Audience: A Case Study of Creation Journey. In Proceedings of the International Conference on Human-Computer Interaction, Orlando, FL, USA, 26–31 July 2019; pp. 33–44.
39. Yu, Y.; Binghong, Z.; Fei, G.; Jiaxin, T. Research on Artificial Intelligence in the Field of Art Design Under the Background of Convergence Media. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Ulaanbaatar, Mongolia, 10–13 September 2020; p. 012027.
40. Promptbase. Available online: <https://promptbase.com/> (accessed on 25 August 2022).
41. Hageback, N.; Hedblom, D. *AI FOR ARTS*; CRC Press: Boca Raton, FL, USA, 2021; p. 67.
42. Hertzmann, A. Can Computers Create Art? *Arts* **2018**, *7*, 18. [\[CrossRef\]](#)
43. Oppenlaender, J. Prompt Engineering for Text-Based Generative Art. *arXiv* **2022**, arXiv:2204.13988.
44. Ghosh, A.; Fossas, G. Can There be Art Without an Artist? *arXiv* **2022**, arXiv:2209.07667.
45. Chamberlain, R.; Mullin, C.; Scheerlinck, B.; Wagemans, J. Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychol. Aesthet. Crea.* **2018**, *12*, 177. [\[CrossRef\]](#)
46. Hong, J.-W.; Curran, N.M. Artificial intelligence, artists, and art: Attitudes toward artwork produced by humans vs. artificial intelligence. *ACM Trans. Multimed. Comput. Commun. Appl.* **2019**, *15*, 1–16. [\[CrossRef\]](#)
47. Gangadharbatla, H. The role of AI attribution knowledge in the evaluation of artwork. *Empir. Stud. Arts* **2022**, *40*, 125–142. [\[CrossRef\]](#)
48. Corbin, J.; Strauss, A. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*; Sage Publications: Newbury Park, CA, USA, 1998; pp. 172–186.
49. Lin, Z.Y. *Multivariate Analysis*; Best-Wise Publishing Co., Ltd.: Taipei, Taiwan, 2007; pp. 25–35.