International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

"Journey of Finding the Best Query": Understanding the User Experience of AI Image Generation System

Soomin Kim, Jinsu Eun, Changhoon Oh & Joonhwan Lee

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2024.2307670

"Journey of Finding the Best Query": Understanding the User Experience of AI Image Generation System

Soomin Kim[a*], Jinsu Eun[a*], Changhoon Oh[b], and Joonhwan Lee[a]

[a]Seoul National University, Seoul, Korea; [b]Yonsei University, Seoul, Korea

ABSTRACT
With the advancement of AI, even people without professional experience can create artworks using AI-based image generation systems like DALL-E 2. However, little is known about how users interact with these new AI algorithms, much less how AI-infused systems can be designed. We explore the user experience of these new technologies and their potential to foster creativity. A user study was carried out where 13 participants executed tasks of creating artworks using DALL-E 2 alongside in-depth interviews related to their experience. The results showed that users had ambivalent opinions regarding the algorithm's performance. When users were informed of the system's capabilities, they subsequently utilized more specific prompts to generate the intended output. Users also optimized their prompts (the queries they entered to create artworks) based on how algorithms worked to achieve their desired outcome. The users wanted a two-way interaction where AI explained the outcome and accepted feedback rather than simply accepting unilateral instructions. We discuss the implications for designing interfaces that maximize creativity while providing comfort for the users.

KEYWORDS
Generative AI; human-AI interaction; text-to-image generation; prompt engineering

1. Introduction

With the advancements in machine learning (ML) algorithms, Artificial Intelligence (AI) is being applied to a wide range of applications, extending the capabilities of humans and supporting their daily lives. Moreover, AI is making inroads into what was previously thought of as exclusively human prerogatives, such as drawing and writing, that are currently being researched (Colton et al., 2012). In recent times, AI art has received a lot more attention, not only because of the question of whether AI algorithms can create art but also because of its socio-cultural and political implications (Coeckelbergh, 2017; Daniele & Song, 2019; Miller, 2019).

Several advanced AI algorithms have been developed in recent years, including GAN (Cetinic & She, 2022; Goodfellow et al., 2020). These algorithms have enabled AI systems using these models to be developed by anyone, making AI art a reality (Gamage et al., 2022). As AI-created art entered the realm of reality, it sparked social and cultural debates. In September 2022, an image created by an AI image generation tool took first place in an official art exhibition, sparking a public debate over whether the use of the tool should be considered creative art or cheating in the art contest.[1] The image was created using "Midjourney"[2] for the 'Digital-Manipulated Photography' category of the Colorado State Fair's annual art contest. Consequently, a controversy arises over AI-generated art and its implications for human creativity. OpenAI has launched DALL-E 2,[3] together with the release of Midjourney, an artificially intelligent algorithm capable of generating high-quality images as a result of text queries placed to the algorithm. Compared to DALL-E 1, which was released in 2021, DALL-E 2 is capable of producing much higher-quality images.

Even though these technologies are important and receive a great deal of attention, in addition to having the potential to empower untrained users who do not have to possess technical skills, little is known about how people perceive these technologies from the point of view of human-computer interaction. Previous research on AI-infused systems in the creative field focused on co-creation with AI agents (Fan et al., 2019; Koch et al., 2020; C. Lee et al., 2020; Lin et al., 2020; Oh et al., 2020, 2018; Shi et al., 2020; Walsh & Wronsky, 2019) and the users' perceptions of AI-generated art (Daniele et al., 2021; Elgammal et al., 2017; Mikalonytė & Kneer, 2022; Ragot et al., 2020; Wu et al., 2020). However, the ML algorithms or AI-infused interfaces used in the studies are not accessible to people outside of academia, and even so, these models are still in their experimental stages, designed as research probes. They cannot address all the scenarios that users can face in-situ with these technologies. In addition, despite every study advancing our understanding of effective design, there is a lack of understanding regarding how and why such systems are used and which factors influence the acceptance and success

CONTACT Joonhwan Lee joonhwan@snu.ac.kr Seoul National University, Seoul, Korea
*The authors contributed equally to this article.
This article was originally published with errors, which have now been corrected in the online version. Please see Correction (http://dx.doi.org/10.1080/10447318.2024.2327167)

of such systems. In order to make AI production more inclusive, it is important to lower entry barriers and enable a wider range of stakeholders to take part in the process.

Taking this into account, we investigated the user experience of an AI image generation system that has actually been launched in the public domain to fill in these gaps by exploring user experiences. Based on this background, we focus on the following research questions:

- RQ1: How do users interact with a text-to-image generation system? How does the interaction pattern change before and after knowing the system's capabilities?
- RQ2: What design considerations should be taken into account for AI-infused systems for creative works?

We conducted a semi-structured in-depth interview with 13 participants, along with an image creation task. We used DALL-E 2, an ML model developed by OpenAI that generates digital images from natural language descriptions. With DALL-E 2, the participants generated images and reasoned about the algorithm through the think-aloud method and interview. They performed the task twice, once before and once after receiving information about the capabilities of the system. The following are the main findings from the interview:

- Users have an ambivalent attitude toward AI and expect AI to create output beyond their expectations.
- Knowing the system's capabilities enabled the users to enter more specific queries to obtain their desired results.
- With the ultimate goal of creating the best results, the users gradually adapted to the system by finding the best query.
- Instead of giving one-way instructions, users wanted to collaborate with the system. Their expectation was reciprocal communication with the system that would improve the outcome.

Based on these findings, we discuss the design implications for intelligent user interfaces that can empower people without expertise to create high-quality works.

2. Related work

The research we conduct falls into two distinct areas. Our first step is to draw inspiration from AI applications in creative fields. The second part of this paper explores how humans perceive AI-generated content.

2.1. Co-creation with AI

With AI advances, simple and repetitive tasks are being replaced, as well as creative ones such as writing, composing, and drawing. Once only humans were considered as having access to these domains inherently, but AI presents us with exciting possibilities. When it comes to writing, various language models can approach the quality of human-level writing in a variety of ways. As well as assisting human writers with editing and providing ideas, they interact with users and actively engage with their writing (Biermann et al., 2022; Gero & Chilton, 2019; Osone et al., 2021). AI also has the promising potential to express creativity through music. It includes studies that propose systems that produce music jointly with users of various spectrums, not just those that support general compositions (J.-W. Hong et al., 2022; Huang et al., 2020; Louie et al., 2020).

The field of AI-based drawing is also actively being explored. In the HCI field, research has focused primarily on ML-based drawing tools and understanding user collaboration with AI (Fan et al., 2019; Koch et al., 2020; C. Lee et al., 2020; Lin et al., 2020; Oh et al., 2020, 2018; Shi et al., 2020; Walsh & Wronsky, 2019). Working collaboratively with AI led to more creative results than working alone. Using DuetDraw, Oh et al. (2018) demonstrated how users and AI agents could collaborate on drawing pictures and explored the design implications of AI-infused systems for creative works. With ImageCascade, users can work in individual and shared workspaces as well as collaborate with various intelligent agents (Koch et al., 2020). Similarly, iterative design can be facilitated using GUIcomp by providing real-time, comprehensive feedback on the user's current design (C. Lee et al., 2020). AI could further enhance the results of user-created artwork by making them more prosperous and emotional. Using EmoG, users can draw expressive characters based on their strokes, increasing the expressiveness of their user stories (Shi et al., 2020). Collabdraw facilitates the collaborative sketching of everyday visual concepts using recurrent neural networks (Fan et al., 2019). Furthermore, AI has been used in co-design processes to assist multiple users beyond individuals. Cobbie enables designers to brainstorm creative and diverse ideas, stimulate exploration, and spark unexpected solutions (Lin et al., 2020). AI was also used to create more inclusive co-design experiences for marginalized groups (Walsh & Wronsky, 2019). Recent research on AI for creativity has explored how AI-based image generation systems can help visual artists in the field (Ko et al., 2022). The authors suggest a number of ways to support artists' creation processes (e.g., automating the creation process, facilitating or arbitrating communication). Recent work has also been done on assisting users in text-to-image generation systems. Wang et al. (2023) introduced RePrompt, a system that refines text prompts into more accurate expressions, specifically for the representation of emotions.

Following the threads of these studies, this study explores questions related to user experience and interactions with AI-produced creative results. As most of the above studies have incorporated AI only at an experimental level, we want to focus on higher-level, professional-quality creations. Specifically, we focus on the representative image generation algorithm, DALL-E 2 (Ramesh et al., 2021). DALL-E 2 is an ML model developed by OpenAI to generate digital images from natural language descriptions. DALL-E 2 generates highly detailed and semantically meaningful images by combining CLIP and diffusion models (Ramesh et al., 2022). Although it has gained considerable attention, little is known about its user experience. Throughout this study, we aim to investigate what users think of the technology, how they interact with it, and what challenges they encounter. We seek to gain a deeper understanding of AI in the creative space from the user's perspective.

2.2. User perception of AI-generated image

AI is increasingly being used in areas of creativity that used to be regarded as human spheres. Besides writing and music, art also actively explores AI applications. The users' perceptions of AI-generated artworks have been the subject of recent studies in art and painting. One belief holds that machines cannot surpass humans in art creation due to their lack of intelligence, autonomy, and emotions (Hertzmann, 2018, 2020). There has been a recent research effort aimed at empirically verifying this human superiority bias.

Several studies have revealed inconsistent findings about people's evaluations of AI-generated artworks compared to human-generated artworks. The results of several studies have revealed that people have negative attitudes toward AI-generated artworks. Human-created art is perceived as more beautiful and new than AI-generated art, even for identical images (Ragot et al., 2020). Similarly, people give higher ratings to paintings created by humans than those created by AI. A higher rating was given to those of human authorship regarding the spatial presence, empathy, and competence (Wu et al., 2020). As opposed to this result, J.-W. Hong and Curran discovered that the identity of the painter (human vs. AI) did not affect how people perceived the value of the artwork. Also, people regard both robot and human paintings as art on an equal footing (Mikalonytė & Kneer, 2022). Further, human evaluators could not differentiate between AI-generated art and contemporary artworks from top art fairs (Elgammal et al., 2017).

While such conflicting results have been found in the literature, evaluator characteristics may influence the perception of AI-generated artwork. Art experts' preferences and purchase intentions are affected by an artwork's creator's identity (Gu & Li, 2022). The effect was not observed in people without art expertise (Gu & Li, 2022). People with quick judgment could also distinguish between human-made strokes and AI-generated ones faster than those without (Daniele et al., 2021). Additionally, AI-generated art is perceived differently depending on the user's experience. Observing computer artists in action has reverted users' negative aesthetic bias toward computer-generated art (Chamberlain et al., 2018).

Previous studies indicate that perceptions of AI as an art creator are becoming increasingly important. This body of work has shed light on the initial understanding of AI-based image creation algorithms and systems by focusing on users' perception of AI-generated images. However, the user experience of AI-based image generation systems remains a research gap that needs further exploration. In particular, very little work has focused on the user as an image creator using AI image generation algorithms. The user has been viewed as an observer in previous studies rather than as a creator (Elgammal et al., 2017; Gu & Li, 2022; J.-W. Hong & Curran, 2019; Ragot et al., 2020). Having insight into how users create images by interacting with AI systems can help design user-friendly AI image generation systems. In this study, we investigate user experience from the viewpoint of a creator who creates artwork using a large-scale AI system that produces high-quality results.

2.3. Text-to image generation systems and DALL-E 2

A variety of text-to-image generation systems based on deep generative models are being used to create digital images (Crowson et al., 2022; Ramesh et al., 2021). In response to a natural language prompt, these generative systems produce high-quality digital images. These systems allow users to create images simply by writing prompts in natural language without requiring technical programming expertise. A new field called prompt engineering has emerged as a result of this practice (P. Liu et al., 2021), which is also known as prompt programming (Reynolds & McDonell, 2021) and prompt design (N. Zhang et al., 2021).

Among diverse text-to-image generation systems, our research used DALL-E 2 to explore user experience with AI image generation systems. Due to its public availability, performance, and impact, we chose DALL-E 2 as our research probe. DALL-E is publicly available and has been used in a number of research projects and applications. 120k Reddit members make up this community, which ranks in the top 1% by size. With its ability to generate high-quality images from text descriptions, DALL-E has had a significant impact on the field of computer vision (Croitoru et al., 2022). In addition, DALL-E uses an advanced generative model that can handle a variety of text inputs and generate a variety of images, which makes it an ideal tool for studying user interaction with text-to-image systems (Ramesh et al., 2022).

DALL-E 2 generates images in two stages. Initially, CLIP (Contrastive Language-Image Pre-training) produces an image embedding using the provided text as input (Radford et al., 2021). Subsequently, decoders, specifically diffusion models, construct the actual image using this embedding (Ramesh et al., 2022). CLIP model uses natural language supervision to learn visual concepts, efficiently connecting textual and visual semantics. With CLIP, DALL-E can analyze millions of digital images and text captions that describe each image, analyzing patterns. By doing so, it learns to recognize associations between words and images. After that, a neural network called a diffusion model is used to generate an image satisfying the CLIP. Diffusion is a method of training a generative model by learning to undo steps of a fixed corruption (adding noise to an image). After training, the diffusion model can generate data by simply passing randomly sampled noise through the learned denoising process. It takes a random pixel and distorts it, then uses CLIP to convert that distorted image into a completely new one. Using this method, it creates a new image from scratch based on the text entered. DALL-E 2 is capable of generating creative, high-resolution images at high speeds using CLIP and diffusion models.

3. Methodology

We conducted semi-structured interviews (Bryman, 2016) to examine the experience of users experiences interacting with an AI-based image generation system. We used DALL-E 2 as our research probe. An image creation task was performed by 13 participants using the think-aloud method. In addition to the think-aloud protocol, we conducted interviews before, during, and after the task (Figure 1).

3.1. Participants

A total of 13 users were interviewed (6 female and 7 male), ranging in age from 26–42. The inclusion criteria included having no experience with DALL-E 2. They were recruited through an open call via social media. To ensure the diversity of the participants, we recruited them from various backgrounds. For this reason, we recruited people with diverse occupations and arts expertise. As a result, we recruited three data scientists, two fine artists, a graphic designer, a content creator, a UX designer, a UX researcher/amateur photographer, a product manager, a business developer, a strategic manager, and a student (Table 1).

3.2. Think-aloud and interview procedures

First, a pre-interview was conducted to assess users' perceptions of AI-generated systems. Next, the study participants created images with DALL-E 2, reasoned about the output using the think-aloud method, and then participated in interviews. For the think-aloud segment, participants were instructed to vocalize their thoughts while engaging in the image generation task, following the guidance: "As you work on the image generation task, please say out loud what comes to your mind" (Van Someren et al., 1994). This allowed for a real-time capture of their cognitive processes and experiences. In addition to exploring the overall user experience of the AI image generation system, we also explored the difference between the experience before and after knowing the system's capabilities. In order to inform users about the capabilities of the system, we used a guide. Participants used the system twice, once before and once after the guide was provided. An interview was conducted for each use of the system (before and after the guide). At the end of the user research, we conducted a wrap-up interview to explore the user's overall experience with the system and his/her attitude toward AI art. During the mid-interviews (before and after the guide), we asked participants to rate their overall experience with the system on a five-point scale. This evaluation was not intended to quantify the system's usability. It was conducted to allow users to reflect on the system's strengths and weaknesses, and to evaluate whether their experience had improved or deteriorated after the guide was provided. Table 2 listed specific question topics for each session. An average of 120 minutes was spent on each experiment. Each participant received a $15 gift voucher as a reward.

3.3. Tasks

Participants executed image creation tasks by interacting with DALL-E 2. Using DALL-E 2, they entered queries and checked the results. Based on the results generated, the original query may be modified or a new one written. When creating the image, neither the subject nor the method were restricted. The system could be used as users wished. Participant interaction trials were not limited, but at least three queries had to be entered.

We also tried to observe the differences in user experience and behavior according to the provision of explanations for the system's capabilities. Providing appropriate explanations can positively affect user experience, especially for AI systems (Amershi et al., 2019). Based on the Human-AI Interaction guidelines (Amershi et al., 2019), we define two major goals for the system guide: 1) Make clear what the system can do, and 2) Make clear how well the system

Table 1. Age, gender, and occupation of the participants and changes in user evaluation.

P#	Age	Gender	Occupation	Score (Before guide)	Score (After guide)
P1	36	M	Content Creator	4.5	5.0
P2	31	M	Data Scientist	3.0	4.5
P3	33	F	Product Manager	4.0	2.0
P4	28	F	Graphic Designer	3.5	4.0
P5	26	M	UX Researcher/Amateur Photographer	4.0	4.5
P6	29	M	Data Scientist	3.5	4.0
P7	32	F	Fine Artist	2.5	1.0
P8	37	M	Business Developer	2.0	3.0
P9	34	F	Strategic Manager	3.0	4.0
P10	41	M	UX Designer	2.0	3.0
P11	36	M	Data Scientist	3.0	4.0
P12	27	F	Student	3.5	4.5
P13	42	F	Fine Artist	2.0	3.0

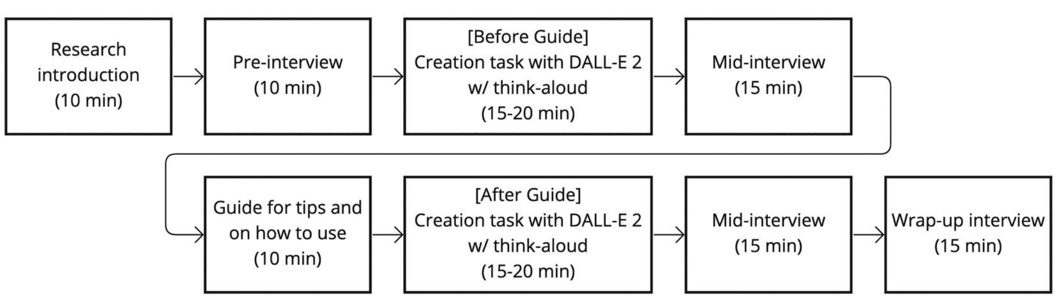

Figure 1. Overall research process. Participants performed image generation tasks with the think-aloud method and interview. They performed the task twice, once before and once after receiving information about the system's capabilities (before guide and after guide).

Table 2. Topics of the questions.

Session	Topics of the questions
Pre-interview	Demographic information User perception on AI-based image generation systems (open-ended & keywords)
Mid-interview (before guide)	General impression of the system (open-ended & keywords) 5-point rating of the system (with reasons) The best and worst aspects of using it Suggestions for adding or improving features
Mid-interview (After guide)	General impression of the system (open-ended & keywords) 5-point rating of the system (with reasons) The best and worst aspects of using it Suggestions for adding or improving features Effectiveness of the guide
Wrap-up interview	Overall impression of the system Perception of the interaction method (human or machine-like) Idea for improving the system Future applications

can do what it can do. To deliver this guide goal, we delivered quick information to the participants on how to use the system and example cases based on DALL-E 2 prompt book[4] and DALL-E official Instagram.[5]

To gain a deeper and more detailed understanding of users' thoughts, we conducted a qualitative study using a think-aloud method and semi-structured interviews (Bryman, 2016). During the tasks, participants were free to express their thoughts in real-time. All think-aloud sessions were videotaped. Following the completion of all tasks, we conducted semi-structured interviews. The interviews focused on the participants' overall impressions and attitudes toward DALL-E and their experience with it. Furthermore, we asked the users about the pros and cons of the system, its future usage, and their views on AI art. All the interviews were audio recorded.

3.4. Analysis

A thematic analysis was used to analyze the data. Thematic analysis is used in qualitative research to identify and analyze themes (Braun & Clarke, 2006). Initially, the research team, comprising three researchers, collaboratively reviewed the interview transcriptions, facilitating discussion on primary observations. This process was to get familiarized with the data. Subsequently, we scrutinized the recorded data to identify meaningful statements that indicate the user experience of the AI image generation system. This step was revisited thrice, refining our insights each time.

Next, we employed Reframer, a qualitative research software, to perform keyword tagging and identify themes. We analyzed 354 statements from the initial responses. While reviewing the statements, we annotated statements with one or more keywords/key-phrases, aiming for these terms to encapsulate the statement's essence. This process generated 732 distinct tags (keywords/key-phrases). Afterward, we consolidated these tags to initiate the theme generation process, yielding 18 unique sub-themes. Notably, sub-themes echoed by a lone participant were excluded, reducing our list to 12. Finally, we refined, interconnected, and consolidated these themes into four primary categories. Reviewing themes was repeated five times, continuing until we reached theme saturation. This process enriched our comprehension of user experiences with the AI-based image generation system. The themes and their example quotes (statements) can be found in Table 3.

To ascertain the impact of the guide on user prompt creation, we analyzed changes in two aspects: 1) prompt length and 2) prompt quality. We utilized word count as an indicator of prompt length, providing insights into the level of detail participants incorporated. On the other hand, prompt quality was assessed based on the criteria from the DALL-E 2 prompt book and components defined by Liu & Chilton (V. Liu & Chilton, 2022), focusing on the inclusion of subject, explanation, and style in the prompts. Two researchers carried out the tagging independently, achieving 100% consistency in their results.The details of this analysis can be found in Section 4.2.3.

4. Findings

4.1. Users' ambivalent attitudes and expectations toward AI

The interview revealed that participants possess both positive and negative attitudes toward the AI art generation system. In addition, users perceived AI as creative when it produced unexpected results which were beyond their expectations. Our study also found that initial experiences could influence ongoing usage. Specifically, initial negative experiences prevented users from using the system continuously.

4.1.1. Positive attitude toward the system

Many participants described their experience as positive after using the system. Some keywords people used to describe DALL-E 2 included "interesting," "fun," "creative," "plausible," "entertaining," "well-made," "amazing," "practical," "novel," "revolutionary," and "inflection point" (Table 4). According to P4, AI is skillful at processing data since it refines the output based on enormous data, instead of being skilled at artistry. Using AI as a metaphor for "pay dirt" and "bonanza," P1 praised its endless possibilities while pointing out human contribution to it: "AI is like a bonanza you can only mine with a lot of effort. AI's possibilities are endless, but it doesn't give you everything at once." DALL-E 2 surprised P5 with its unexpected results. "In spite of my low expectations before using it, I was very impressed and surprised that when the AI generated the four images, each one had a unique style, and even some had unexpected results. I liked how AI made unexpected results. It was a novel and new experience. That's really AI that does art."

Additionally, AI enabled users to materialize and articulate their imagination into verbal expressions. Participants experienced AI performing traditional artistic techniques on their behalf despite the absence of traditional artistic techniques, which generally require dexterity and skill. They only needed imagination and the ability to articulate their

Table 3. Themes, tags, example statements, and corresponding participants from the qualitative analysis.

Theme		Tags	Example statement	P#
(1) Attitude and Expectations	·Positive attitude	Positive attitude, Creative, Imagination	*"AI's creative prowess captures my imagination. It offers the canvas and colors."*	1-13
	·Negative attitude	Negative attitude, Authenticity, Skepticism on AI creativity	*"It may replicate patterns and styles, but can it ever truly capture the soul of art? DALL-E 2 still seems to miss the heart of what art really is."*	3,5,7,8,11
	·Unexpectedness	Unexpectedness, Unpredictability	*"It wasn't what I envisioned, but the unpredictability caught my interest."*	2,5,7,8,10,13
	·Initial Experience	Initial experience, Frustration, Reluctance to reuse	*"Honestly, my first try with the system was a letdown. If that's the AI's best foot forward, I'm not sure I want to give it another go."*	3,7,8,12
(2) Pros & Cons of System Use Guidance	Detailed prompt	Detail, Step-by-Step Queries, Query Customization	*"The guide showed me how to play around with styles, moods, and more. After diving into it, I tweaked my query and boom, got the output I wanted."*	1-6,8-13
	·Reference (prompt)	Reference, Step-by-Step Queries, Detail, Directional Consistency	*"The guide is really helpful because I'm lost with the search bar at the beginning."*	2,5,6,9,12,13
	·Reference (image)	Inspiration, Aha Moments	*"Seeing others' images and questions really opened my eyes and sparked new ideas I hadn't considered before."*	8,9*****
	·Restriction	Bounded creativity, Over-reliance	*"Sometimes, sticking too closely to guides or examples can box in our imagination. With AI, the magic isn't just in searching – it's in creating"*	5,7,10*****
(3) Algorithmic Reasoning and Adaptation	·Reasoning&Sensemaking	Reasoning, Stereotypes, Data bias, Western Centricity	*"When I created images about SAT scores, mostly Asian girls appeared. It underscores how the dataset is centered on West and its stereotypes."*	1,4,6,7,11
	·Reasoning&Sensemaking	Sensemaking, Keyword prioritization, Algorithmic processing	*"It's like the weight of one keyword overshadowed the other."*	5,6,7
	·Iterative refinement	Iterative interaction, Refinement, Inference	*"Only by inferring and understanding how AI works could I get better results. (...) Multiple interactions with AI are important."*	1-13
(4) Human-like versus Machine-like Interaction	·Two-Way interaction	Machine-like interaction, Search-like interaction, Keyword optimization	*"I felt as if I was optimizing search keywords so AI could create good images.""*	2-5,7-13
	·Degrees of Freedom	High Degree of Freedom, Challenge, Navigational complexity	*"While the limitless possibilities of DALL-E interface are intriguing, it paradoxically becomes a canvas so vast that it's daunting"*	1,2,3,5,7,9,10, 12

imagination in a way that AI could understand. According to P2, "I enjoyed speaking more precisely and clearly when expressing my thoughts. Words and descriptions were better when the image was embodied in the head. AI makes exactly what I say in my head, so I know something like isn't enough, and this part should be more descriptive."

4.1.2. Negative attitude toward the system

Based on the research finding that humans differentiate AI (Oh et al., 2017), a number of participants expressed sarcastic views regarding AI art. People have a negative attitude toward AI art since they believe that art is the exclusive property of humans. P5, for instance, characterized AI that creates art as a 'heresy that challenges inviolability.' P5 said, "Humans have always been considered to be the only species capable of being creative and producing aesthetic results. My opinion is that art holds a unique position in relation to other disciplines. Suddenly, I hear that AI is doing art, not humans. Obviously, I have expectations, but I also feel resistance and resentment, 'Is this right?', as if rejecting heresy." He also expressed concern about the 'laziness' of art and creation. In addition, P7 defined the system as a tool rather than an artist or co-creator: "I don't see AI as an artist. Basically, it's a tool for sketching ideas. There are many sources of inspiration for artists, and this system, what we call AI, is one of them. I was expecting something new since the style generator came out a long time ago. But it feels like DALL-E 2 is focused only on accuracy and output quality, so it hasn't evolved artistically."

4.1.3. Users' expectations on unexpected results

It was found that users expect more than AI to generate images precisely according to their instructions. Users regarded AI as creative when they encountered results that exceeded their expectations. We were able to infer that the accuracy of the algorithm is crucial, but the possibility of

Table 4. Keywords for the overall perception and experience of AI image generation system defined by users.

P#	Before Usage	Before Guide	After Guide
P1	Magician, Extremely skilled, Genius	Surprising, Fun, Diligent	Pay dirt, Human-like, Guess Who I am Game
P2	Mimic, Inspirational, Magical	Surrealistic, Abstract, Disappointing	Customizable, Inductive, Trial and error
P3	Goosebumps, bizarre, curious	Interesting, Prompt, Detailed	Stupid, Disappointing, Bizarre
P4	Curious, Skilled, Regrettable	Cute, Expectation, Fun	Entertaining, Well-made, Domain-professional
P5	Impossible, Heresy, Human prerogative	Fun, Unexpectedness, Creative	Amazing, Practical, Lazy art
P6	Keyword-based, Plausible, Imitating style	Plausible, Imperfact, Novel	My secret artist, Somewhat explainable
P7	Coincidence, Unintentional, Boundary	Disappointing, Ellipsis, Boring	Tool, Restrictive, Passive
P8	Robot, Unfinished, Infinite	Letdown,Realistic, Noob	Avatar, Far way to go, Nimble-footed lackey
P9	Synthesize, Just a beginning, Controversy	Fun, Popularization of art, Addictive	Inflection point, Emergence, Revolutionary
P10	Pessimistic, How dare, Curious	Against expectations, Fun, Creative	Infinite, Versatile, New genre
P11	Novel, Worrisome, Ethical Issues	7\10, Defiant, Trigger&Flare	Artist, American, Prototype
P12	Curios, Le Penseur, Unexpected quality	User-dependent, Convenient, Clumsy	Full of surprises, Shy Inspirational
P13	Inflection point, Market changer, Push the boundaries	Artificial, Mechanical, Only so much	Endeavorer, Efficient, Take at his word

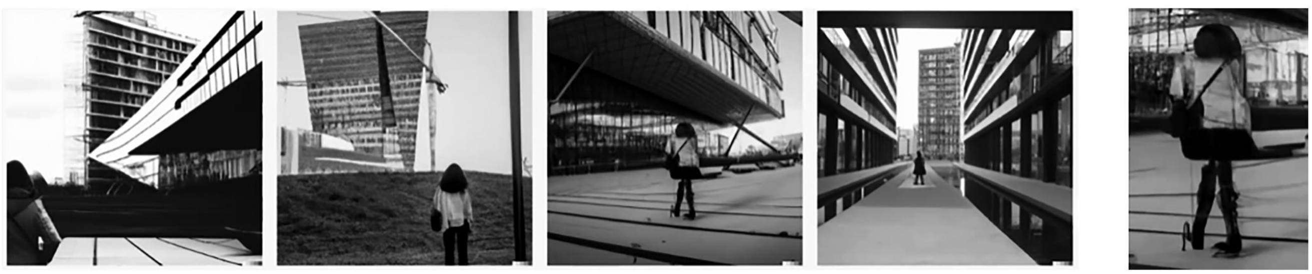

Figure 2. Images generated by DALL-E 2 for the query of 'a black and white scenery with modern architecture in the background and a woman in the center." the image on the right is an enlarged view of the woman in the third picture.

generating emergent and unpredictable outcomes cannot be completely ruled out, even at the expense of accuracy.

Those who witnessed the picture of AI that exceeded their expectations described it as creative. As a result of AI's imperfect and unpredictable outcomes, participants found opportunities for creativity. By observing the component in the picture that looks like a human at a distance, but not at a close distance (Figure 2), P5 recognized that AI is creative. He said: "I recognized the potential of AI art, since it did more than copy and paste a human shape, it expressed it artistically. I wonder why AI only mimics the human figure." Due to an unexpected result generated by the AI, P5 raised the evaluation score by 0.5 points.

On the other hand, participants who received only their expected results expressed disappointment and did not perceive AI as creative. In spite of the high-quality output, several participants expressed regret that AI's results are still within the boundaries of the user's query. After using the system, P8 concluded that AI is still "in human hands." "AI is only capable of generating my input. There is nothing it can create on its own. Sensitivity and creativity are what we expect from the artwork. AI doesn't meet that standard. How is it different from Photoshop's more advanced version?" P7 was also disappointed that the image generated for 'frogs in the pools singing at night' was what she exactly expected without a hitch. "It was like searching an encyclopedia," P7 said. In addition, P13 expressed disappointment that the system only superficially interpreted her query. In her description, the AI's image of 'love' looked like a flyer (Figure 3). In comparison with Midjourney, P10 pointed out the monotony and typicality of DALL-E 2: "The overall style of midjourney is just so interesting because it always gives me unexpected results. Usually, Dall-E is more coherent, but also much more boring." This unexpected novel emergency was also emphasized by P8: "Even if it doesn't follow my exact instructions, I'd like random elements to be added automatically. What I want to see is beyond my imagination. "I get inspired by something missing my prediction."

4.1.4. Initial experience affects ongoing usage

Another important finding from the user study is that users' initial impressions also play a crucial role in shaping their overall experience. Negative initial experiences affected continuous usage intentions negatively. P3 described the overall use experience as "annoying" and "stressful" since she did not get the expected results despite learning the guide: "There was irritation in me because it seemed as if AI would implement what I wanted, but it didn't. I feel like AI tried to wind me up! This system isn't something I'd like to use

Figure 3. Images generated by DALL-E 2 for the query of 'love'.

in the future." In a similar vein, P7 stated that she would not use the system in the future since it did not meet her expectations. This finding is consistent with prior research, which demonstrates that an individual's experience and trust in AI significantly influence their intent to utilize such technologies (Choung et al., 2023).

Moreover, the user's initial experience also influenced the overall evaluation of the system. In response to P8's failure to obtain the expected results for 'Billionaires making a positive impact on the world', P8 commented, "It doesn't seem any better than Google search." Additionally, P8 expressed doubts about whether images retrieved from Google are presented randomly to the system and entered the same query into Google and compared the results. Prior to knowing about the system's capabilities, P8 referred to it as an ignorant "noob" who had just entered the art industry.

Meanwhile, users also adjusted their expectations toward AI by simplifying their queries. For instance, P3 changed "battlefield of gigantic horrible ghosts without faces being attacked by a powerful wizard, from the lord of rings(2001)" to "thousands of gigantic monsters being attacked by a powerful wizard, from the lord of rings(2001)." Despite this, when P3 failed to get the desired output, P3 modified "the Lord of Rings (2001)" to "digital art," which P3 thinks AI could handle. In this regard, most participants suggest that the system should adjust the user's expectations in advance. Not only should it display the best cherry-picked results, but it should also provide preliminary information about dominant and inferior domains (P3, P4, P9, P10, P12).

4.2. Pros and cons of being informed of the system's Capability

A study was conducted to examine the experiences and behaviors of users before and after they were informed of the capabilities of the system. In line with the Human-AI Interaction guideline of 'making clear what the system can do' (Amershi et al., 2019), we would like to observe how a user's perception and behavior are altered when the user understands what an AI system is capable of as opposed to when they do not. Due to the complexity and black box nature of AI systems, providing information regarding their strengths can be valuable to the user (Liao et al., 2020). Should it not be useful, we attempted to discern why. In addition, it is possible to derive insights into improving the search-like interface that many image generation systems, such as DALL-E, Midjourney, and Craiyon, have applied. Based on the analysis, we found that being informed about system capabilities had both positive and negative effects on user experience.

4.2.1. Positive aspects of the guide

Users were provided with information about the capabilities of the system in the form of guides. First, the guide enabled participants to write a more detailed and step-by-step query by introducing them to AI's ability to manipulate styles, looks, aesthetics, mood boards, etc. Following the learning of the guide, P4 was able to produce the expected output by adding the specific option "extra long shot, 1980s black and white film" to the original query. P8 also explored various styles of images by entering queries such as "in the style of David Hockney," "modern art style," "oil painting," and "digital art." The participants described their original prompts in more detail along with trying different styles. In order to create a more desired mood in the image, P13 added a background of 'Eiffel Tower in the background' and a modifier 'stunning' to her original query.

It was found that participants perceived a high level of satisfaction when carrying out guided creative activities by referring to the instructions rather than creating the activities from scratch. P2 said, "The guide is really helpful because I'm lost with the search bar at the beginning. There is a great deal of value in actual trial and error and tweaking little by little. After learning a guide, giving input and checking the output becomes much easier, as well as reinforcing that input and verifying the modified result." According to P6, when P6 entered a more detailed description after learning from the guide and examples, AI reflected the instruction and provided a higher-quality result, which increased P6's satisfaction. In addition, P6 said the system has become more explainable: "By giving the AI a controllable and structured input, I'm more likely to get expected results. As the system became more understandable, I raised my score by 0.5."

Aside from entering more detailed queries and varied styles, participants were inspired to come up with new ideas by observing the image output and queries created by others. For example, based on the photo of "A photo of Michelangelo's sculpture of David wearing headphones DJing" uploaded to OpenAI Instagram, P8 created a new query "Michelangelo's sculpture of muscular David running a sprint in NY Olympic," and tested it in numerous ways.

Table 5. Examples of changes in user queries before and after the guide.

P#	Queries (before guide)	Queries (after guide)
P2	The soldiers are marching with the flower guns in 1920 Three baby pandas are suffering from the hell A jiu jitsu player wins a taekwondo player	The man is standing on the street holding an avocado with the two hands, wearing golden sunglasses, the film style of sin city The man is standing on the street holding an avocado with the two hands, wearing golden sunglasses, the film style of sin city, extremely high contrast The entire appearance of the teleport machine that can be ridden by a person, steampunk, photo, 8k
P4	High school students are smiling in front of their SAT scores Corgi is walking on Chanel runway show In the renaissance era, a panda is drawing a portrait of its owner	Corgi with its owner is walking on a high-end fashion brand runway show, extra long shot, 1980s black and white film 120 Black cats resting on IKEA furniture, realistic, extra-long shot
P6	A fresh ph.d. just thrown out in the industry A group of researchers celebrating paper acceptance	A group of researchers opening champagne after the paper acceptance notification, anime style in the style of Johannes Vermeer A young athlete licking the cheek of his step sister in a pink bed A young athlete, softly touching the neck of his step sister in a pink bed, Japanese Manga
P7	Frogs in the pools singing at night Remember me when no more day by day Life, soul, breath, night, sun, cloud, aloud, rain Yeeees ta-dah hooray oh :) :(	Man, Hockney style Man, Picasso Oriental painting :), oil painting
P8	A heaven with peace and no pain A happy man who has peace in mind Billionaires making a positive impact on the world	Michelangelo's sculpture of muscular David running a sprint in NY Olympic Michelangelo's sculpture of muscular David running a sprint in Athene Olympic in front of audience, digital art Walking gentle male model in Paris in the rain with peace, in the style of David Hockney
P9	Flying puppy The most funniest picture in the world A luxury building made with iPhone	A white puppy flying in the clouds, digital art A white puppy with wings flying in the sky surrounded by baby angels, rococo style painting A white puppy flying in the peaceful sky surrounded by baby angels, rococo style painting, in the ceiling of the cathedral
P10	A cat smiling at the rise of bitcoin Airplane flying among pigeons covering the whole sky A portrait of a millionaire who became a dog	A maltese is surfing on Santa Monica beach, cyberpunk style, close-up image, 8k Polaroid photo of of a squirrel in sunglasses and a white puppy taking selfie Airplane flying among pigeons covering the whole sky, realistic photo, 4k
P11	Show me some delicious Vietnamese foods Show me some delicious local Vietnamese foods in Saigon Show me some delicious local Korean foods	Show me some delicious traditional sweet Korean summer deserts in cyberpunk style Show me some delicious traditional sweet Italian summer deserts in cyberpunk style Show me some delicious traditional sweet American summer deserts in cyberpunk style
P12	cat, snow, cartoon cute cat, snow, cartoon Zoom meeting, pajamas, woman	Artemis shooting with an arrow on the moon, realistic, photo Artemis shooting with an arrow on the moon, digital art 3D rendering of game character, Timo walking in the style of League of Legends 3D rendering of game character, Teemo walking in the style of League of Legends
P13	Love Love, Love, Love The most precious love on earth	Stunning fireworks finale at night, with the Eiffel Tower in the background, digital art Stunning fireworks finale at night, with the Eiffel Tower in the background, Aeon Flux cartoon style A cyberpunk illustration of stunning fireworks finale at night, the Eiffel Tower in the background

Regarding creative query generation, P9 added: "There is no doubt that prompt-engineering will become an increasingly developed skill, and that people should be allowed to charge for their time, effort, and expertise."

4.2.2. Negative aspects of the guide

However, referring to guides or examples can also restrict a user's imagination. P5 noted that AI's advantage lies in the generation, not search and that previewing all the best cases offsets its own distinctive strengths that set it apart from Instagram and Pinterest. In addition, P10 remarks "As soon as I entered the website, I was drawn to the examples below." I think I would have subconsciously referenced the sample images if not for the interview situation." The guide also countered P7's perception that AI creates art: "Instructions like 'Please enter this. Please be specific' lead me to believe that this is a machine that faithfully performs specific input, rather than an artist." These contradictory results suggest that algorithm information or level of freedom should be adjusted according to the user's perception of AI and the purpose of utilization (Maxwell et al., 2020).

4.2.3. Change of user queries before and after the guide

Table 5 provides examples illustrating the differences in user queries before and after the introduction of the guide. We assessed changes in user prompts in terms of prompt length and prompt quality. There was a significant difference in the length of prompts before and after the introduction of the guide ($t\ (126) = 5.73$, $p = 0.000$). As participants learned and familiarized themselves with the system using the guide, their prompts became more detailed ($M_{Before} = 8.14$, $SD_{Before} = 3.91$; $M_{After} = 13.11$, $SD_{After} = 5.50$).

In terms of quality, our evaluation revolved around the inclusion of subject-specific details and stylistic elements within the prompts V. Liu and Chilton (2022). Table 6 illustrates the changes in prompt quality before and after the guide. Upon receiving the guidance, it was found that participants incorporated a greater emphasis on detailed description and distinct styles in their prompts. Specifically, after introducing the guide, the prompts with detailed subject descriptions saw a 21.6% rise, while those incorporating style registered a 79.0% increase. This implies that participants engaged with the AI-based image creation system in a more nuanced manner, emphasizing both style and description.

4.3. Algorithmic reasoning and adaptation

One of the overarching themes throughout our findings was the user experiences of algorithmic reasoning and adaptation. For close alignment between judgments and AI outputs, users developed and tested hypotheses and inferred the algorithm's working principles. In addition, users gradually refined their queries to come closer to the desired results.

4.3.1. Users' reasoning and sensemaking toward AI

We found that participants developed their own hypotheses and gradually refined their queries to verify them. The following are examples of users' hypothesis: there are areas where AI excels and areas where it fails (P1-13); AI does not have the capability of describing facial features in detail (P1, P2, P3, P4, P6, P8, P9); AI is capable of portraying objects better than humans (P1, P2, P3, P8, P12); AI is capable of portraying animals better than humans (P4, P9, P13); due to the poor ability of AI to represent human faces, there are many high-quality images of astronauts (P9); the data set contains a significant number of images that are centered on the West (P1, P4, P6, P7, P11); instead of capturing the exact meaning of a sentence, AI combines various elements in it (P2, P3, P8, P10, P12); AI interprets combinations of elements in the query in a variety of ways (P5, P7); concepts that are abstract are not well expressed (p7, P8, P13); AI is unable to reflect the number (P4); as long as similar elements are included, AI is unable to differentiate between them and express them appropriately (P10); AI automatically selects the style that is most appropriate (or with the most data) for the query (P6, P8); When AI represents an element it is not familiar with in its existing data set, it synthesizes elements that do not match exactly but are related (P11) etc. By using these reasonings, they narrowed the gap between their initial thoughts and those of AI. This conceptual reasoning could help users better understand AI and achieve their goals.

Table 6. Changes of prompt quality before and after the guide.

	Num. of Prompt (All)	Num. of Prompt (incl. subject)	Num. of Prompt (incl. description)	Num. of Prompt (incl. style)
Before guide	56	56 (100%)	40 (71.4%)	7 (12.5%)
After guide	71	71 (100%)	66 (93.0%)	65 (91.5%)

Along with building conceptual hypotheses, users inferred how algorithms worked and the characteristics of the data. First, users inferred how the ML model works. Instead of accepting algorithmic results, users deduced and interpreted the black box model. P5 inferred that the algorithm transformed the input natural language into a vector, analyzed the most important keywords and intents, and then extracted the most relevant output among the image candidates with high coherence with major intents. Regarding the output emphasizing the "money" from "A researcher just received his ph.d. and entered in the company to earn money," P6 reasoned that among the elements comprising the query, those with high frequency in the training data are most likely to be saliently represented, while those with low frequency are more likely to be ignored. Furthermore, P6 pointed out that algorithms do not seem to be able to capture abstract concepts well. Observing that in the result of "man, Hockney style," the algorithm brought the structure of David Hockney's work, rather than the painting style, P7 inferred that the algorithm might have given more weight to "man" than to "Hockney." P7 also deduced that the algorithm appears to combine and classify queries in a variety of ways after observing the variations generated by "man, Picasso" (e.g., an artist man, a man of Picasso's painting style, Picasso as a painter and adult male) (Figure 4). P1, who found that the "Man and computer playing on Mars from the animation Adventure Time" results did not reflect the animation style well, claims that picking a theme that fits the animation style is necessary to implement a high-quality style. It is consistent with previous research that people interpret and infer AI results based on their expertise (Oh et al., 2020).

Figure 4. Images generated by DALL-E 2 for the query of "man, Picasso.".

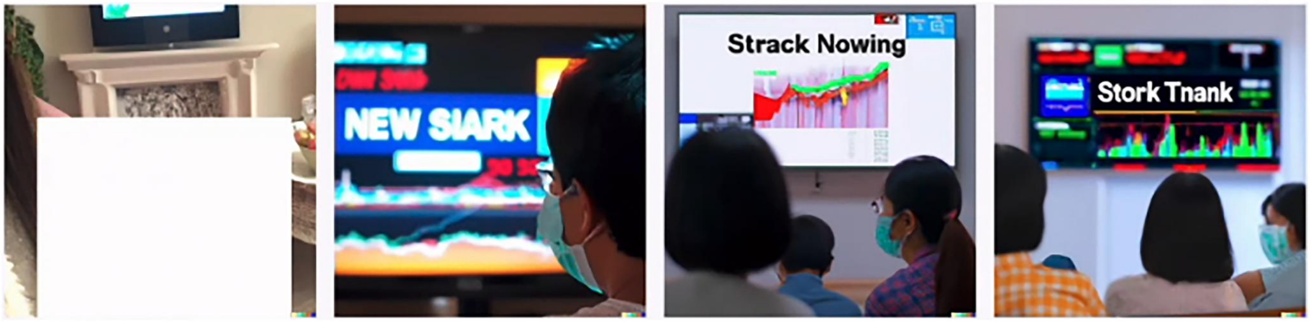

Figure 5. Images generated by DALL-E 2 for the query of "people are wearing masks while watching stock market news on the television, long shot, realistic, 2021.".

Additionally, participants presented various inferences about the characteristics of the trained data based on the algorithm's output. P4's query "People are wearing masks while watching stock market news on the television, long shot, realistic, 2021" returned all Asian figures (Figure 5). The algorithm appears to be biased towards recognizing the subject wearing the mask as Asian based on this result, according to P4. Similarly, P11 raised the issue of the AI's Western-biased data set, as the results showed American desserts were better represented than those from Southeast Asia or other eastern regions. During the interview with P10, 'A cat in the gentleman's hat' and 'A dog in the hat' were presented as realistic pictures, while 'A cat in the hat and a dog in the hat' was presented as an illustration. In regards to this result, P10 stated that images containing both dogs and cats tend to be illustrations, while images containing dogs and cats respectively are mostly photographs. Furthermore, users made inferences about the mapping between images and text. In the case of P6, when the instruction of 'in the style of Johannes Vermeer, anime style' was added to the original query 'a group of researchers opening champagne after the paper acceptance notification,' a female researcher appears more frequently compared to the original image. Based on this result, P6 predicted that Johannes Vermeer and anime would be highly correlated with data containing women.

4.3.2. Gradual finding of the best query for the best results

In order to create desired output, users gradually refine their queries based on their reasoning. Their imaginations were articulated in more specific verbal expressions. Thanks to the guide explaining the system's capabilities, users were able to express their queries more in detail.

Using the instruction that the system could imitate a specific movie style ("the film style of Sin City"), P2 entered the query of "The man is standing on the street holding an avocado with two hands, wearing golden sunglasses, the film style of Sin City." Before pressing the generation button, P2 expected that "Firstly, it will be a 'Sin City' style, it'll be black and white, with a very dark background and a point color. On a black background, on a white road, a very dark man holds an avocado in his hand. The point color appears to be yellow gold-rimmed sunglasses and greenish avocado." Following the generated result (Figure 6 (up)), P2 said, "I think it's similar, but… AI is so good at describing my query exactly, so I should add a little more detail. My language was inadequate to express certain elements. My query was so accurately described by AI that I don't have anything to say. Not AI, but I should do more." P2 then added 'extremely high contrast' and commented on the output (Figure 6 (down)): "Yes, this is the feeling! It was this feeling I had in mind! This is exactly what I was thinking. It captures high contrast well. The photo ate the contrast well. It's a little disappointing that golden sunglasses don't have point color, except for the second photo. I think I would mull it over too. There are two point colors, yellow and green, if you add a point color to the sunglasses. As AI did, I think I'd pick one."

All participants except P7 gradually adjusted their queries by observing generated output and reasoning algorithms to create the best results. For example, P5 included details of detailed photography styles (close-up portrait, wide-open aperture, dim lighting, and low saturation) and could create the expected output by specifying "beautifully designed architecture" as "modern architecture." After the original query failed to yield the desired results, P4 modified it more precisely (i.e., from "120 Black cats resting on IKEA furniture, realistic, extra-long shot" to "Realistic photo of hundreds of black cats resting on chair and sofa in the IKEA store, 8k, extreme wide shot"). In addition, P8 modified "a sprint in NY Olympic" into "a sprint in Athene Olympic" and added "in front of the audience, digital art" to embody his imagination on a web canvas. P9 also modified the lighting, lens, and framing of photos to achieve the desired effect, as well as changing the background and motion (Figure 7).

As for the experience of inferring algorithms and adapting to the system, P2 said the following, "The results used to feel a bit magical, but now that I understand the algorithm, I can get good results. The algorithm seems to have a category of styles, and I got used to it. In addition to AI trying to understand what I'm saying, I also throw queries in an AI-friendly way." P1 also said, "Even though AI is good at creating, it doesn't do it all at once. I've got to explain, explain, and explain again… Only by inferring and understanding how AI works could I get better results. It's not

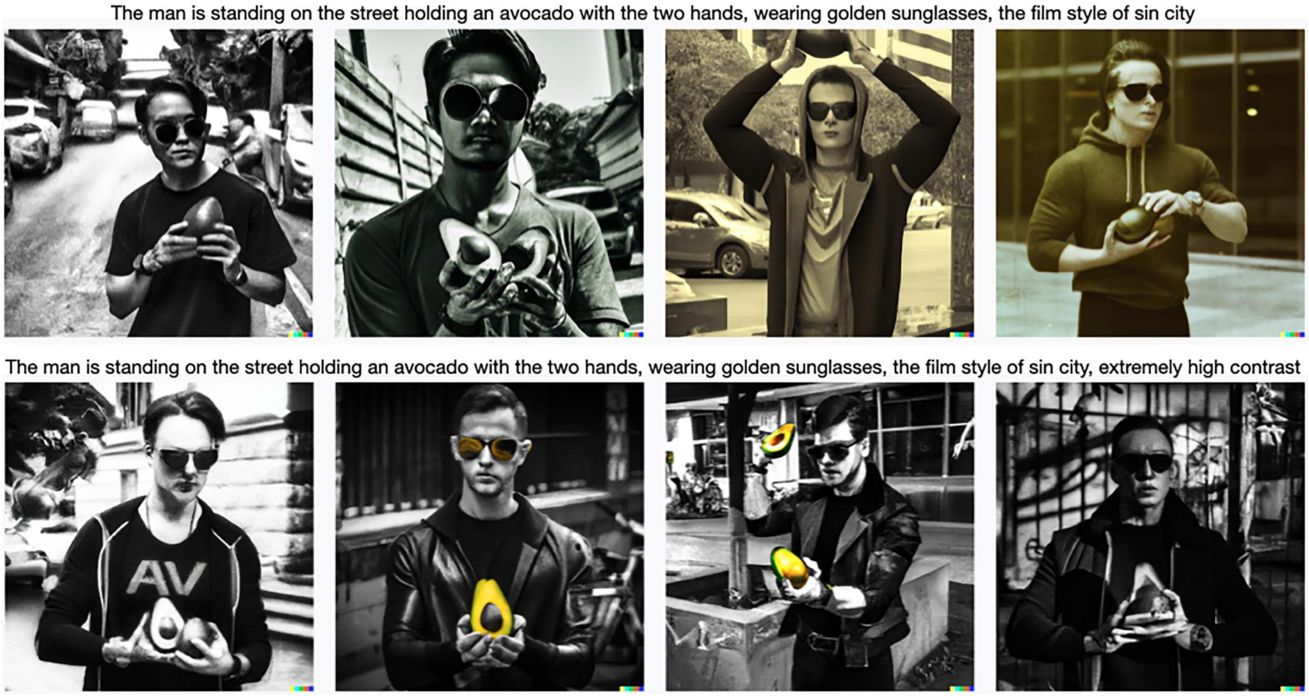

Figure 6. Images generated by DALL-E 2 for the query of "the man is standing on the street holding an avocado with the two hands, wearing golden sunglasses, the film style of sin city" (up) and after adding" extremely high contrast" (down).

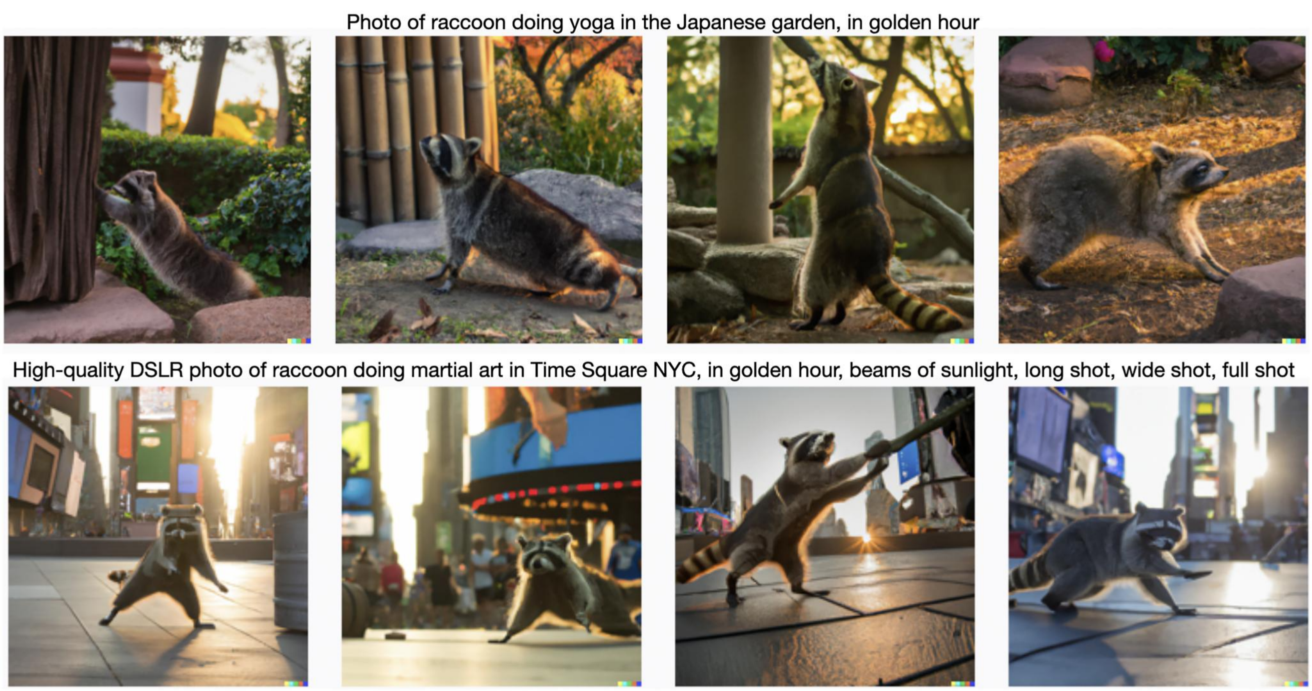

Figure 7. Images generated by DALL-E 2 for the query variations of " … photo of raccoon doing … ."detailed descriptions are provided in the image below.

like AI understands me, even when I ramble gibberish. Multiple interactions with AI are important. This way, I can maximize AI's abilities while also explaining what I want more clearly. It wasn't just about painting. This is about finding an optimal point where AI excels and where it fails." To borrow the expression of P12 "AI is hard in long-shot but infinite when seen in close-up with a little bit of learning."

4.4. Human-like versus machine-like interaction experience

4.4.1. Users' expectations of a two-way interaction rather than a search-like experience

All participants, except for P1 and P6, stated that their experience of using the system was similar to interacting with a machine, not a human. Additionally, a number of participants said their experience of tuning queries and inferring algorithms was similar to that of Google Search. For example, P2 stated, "We adjust keywords when searching YouTube or Google to make sure we can find what we are looking for. That's how it felt to me as well. A Google search, for instance, comes out better if you know how to search for better results. It's not too far off from DALL-E 2. I felt as if I was optimizing search keywords so AI could create good images." Similarly, P10 said, "It was like a machine. Instead of thinking about the interaction itself, I focused on what kind of query I needed to enter to get the image I wanted. It was similar to searching on Google."

A search-like interface also contributed to the perception of machine-like interaction. P5 said, "From my experience, I can clearly tell that I'm using a machine or system. The process of typing a query in a search box, waiting for the algorithm to process, and then receiving the results was very similar to the existing command and search interface." P7 replied that viewing the parallel results was just like seeing the search results. P7 added, "It would be better if the method of presenting the results was more creative and artistic, such as expanding images that AI considers to be more artistic and aesthetically pleasing."

DALL-E is a one-way system in which the user inputs a prompt, and the system then generates an image in response. In this regard, we found that the search-like interface was just a surface-level feature that affects the perception of machine-like interactions. Indeed, the underlying cause was a one-way interaction without feedback. For example, P3 said, "Feedback is exchanged between people. If my illustrator brings me a draft, I'll say, 'Please edit this in this way'. This is how I communicate with human artists, but in this system such interaction does not occur. As I communicated with this system, I felt like I was communicating with a machine because it doesn't learn and develop with me, but always produces results within its capabilities." P7 also said, "This system? Passive. My command is the only one it executes. There is no two-way communication. A guideline doesn't feel like 'I'll show you something new,' but rather 'I can only do this, you must do this to me,' which limits the system's boundaries." P11 expected the system to provide feedback and suggestions. He said, "It would be great if I could get some feedback on my query from the system. By analyzing the relationship between words through NLP, DALL-E could suggest a new query."

The results of these studies suggest that the system should allow for two-way interaction since the direction of communication affects perceived interactivity (Ashktorab et al., 2021). To facilitate two-way communication, the user should be able to perform the roles of both sender and receiver (Edward & Sally, 2000). Current AI image generation systems, including DALL-E 2, only allow users to play the sender's role by inputting queries. During the discussion session, a design recommendation will be presented addressing how to support users to act as receivers.

4.4.2. Challenge of the system with high degrees of freedom

A high degree of freedom is provided by AI-based image generation systems, such as DALL-E 2 and Midjourney, that allow users to enter text queries in an empty search bar-like prompt. Despite conflicting views, seven participants stated that this interface, which gives users complete control over creations, needs to be improved. It was challenging to create from scratch, they said. For instance, P2 said, "To be honest, I don't know what to do. It's hard to come up with something to write." P10 also added "It was like buying my wife a gift for the first time at a luxury women's store without any help! The thing is, I don't even know what my wife likes … "

It was noted that some participants suggested improvement directions as well as noted discomfort. Their proposal was to present predefined options to reduce users' anxiety, and to facilitate the creation process. P1 said, "It'd be nice to have preset options for things like style and color. In a similar way as Prisma. I hope to develop my ideas one by one like stepping stones." P10 also added, "It's good to show a preview of possible styles to ease uncertainty. I have to spend too much time fine-tuning queries now. Having a few presets will allow me to make more images in a short time. Think 80's movies or 90's dramas … I don't get these styles right away. If the system displays these creative capabilities saliently in advance, they will be much more usable." Additionally, P9 suggests that users might be able to compose more creative queries if natural language technologies such as GPT-4 are used to suggest keywords. In the future, research can be conducted on how to integrate natural language processing technology with image generation technology to inspire users.

In contrast, some participants with artistic expertise (2 fine artists and 1 amateur photographer) preferred more flexible interfaces. When translating from language to image, P5 discussed the positive effects of various interpretations: "Currently, we are giving the system verbal, textual commands. Regardless of how detailed the user describes the query, since the medium has changed from words to images, the way the system understands it can be completely different. It's nice because no matter how specific you describe, you'll find something you don't expect, even if you do expect it in some way." A limitation of the interface that shows predefined options was noted by P7: "As long as all possible and good examples are shown in advance, it's no different from the photo app filters, Instagram, or Pinterest."

5. Discussion

In this section, we discuss lessons learned from the user study and their implications for designing AI-infused user interfaces that convey creative works in diverse ways. We also report the limitations of the study.

5.1. Creativity of AI & unexpected outcomes

In contrast to the previous experiments which found that people perceived AI as being less creative than humans (Chamberlain et al., 2018; Ragot et al., 2020; Wu et al., 2020), we found that people believed that AI could be as creative as humans. Specifically, users were able to recognize AI's creativity by experiencing unexpected outcomes. This is in line with the idea that creativity is based on unexpectedness and unpredictability (Boden, 1995). People are expected to be surprised by the elements in AI art, especially when AI is a co-creator.

We propose an AI-infused system that generates unexpected results by altering the user's query in the light of results that allow users to be inspired by outcomes that have subverted their expectations. The interpretation of the user's commands as-is may increase the accuracy of the system but reduce the user's satisfaction with aesthetics. With the help of NLP technology, user queries can be interpreted in a variety of ways. By adding keywords and phrases that are not directly related to the original query, but are highly relevant (or irrelevant), NLP technology can produce more interesting results (Hagtvedt et al., 2019). Additionally, the user should be informed that a new result has been generated as a result of adding a specific new keyword, thus making the system more explainable. Additionally, NLP technology can parse the user's input and provide feedback on how the prompt can be improved. It could facilitate the user's creative process. NLP technology is being used to stimulate users' creativity in AI co-writing (M. Lee et al., 2022). As a further step, NLP technologies such as GPT-3 and ChatGPT could be integrated to interpret and expand users' queries so that participants can be inspired by different modalities.

- **Design implication (D1)**: Integrate state-of-the-art ML algorithms like Natural Language Processing to provide users with a space that can inspire and motivate them to think creatively.
  The lack of a guide for writing prompts could prevent the majority of users from generating their desired output, which affects their consistency. The use of NLP can assist prompt engineering, which is the formal search for prompts aimed at producing a desired output (Sanh et al., 2021). In particular, prompt engineering for text-to-image models involves a number of schemas composed of *medium*, *style* and *subject* (V. Liu & Chilton, 2022; Xie et al., 2023). Additionally, our result has shown that users' prompts also include *look-and-feel (mood)* as well as detailed *descriptions*. These components could be key building blocks for prompt engineering. It might be helpful to provide a set of templates that users can fill out with their own information. By using these templates, users can focus on the important aspects of the prompt and ensure that all the necessary information is included. Furthermore, providing a library of pre-defined templates that users can select from, depending on what they want to accomplish with the model, can also streamline the prompt creation process.
- **Design implication (D2)**: Provide users with templates and a prompt builder that includes key components (e.g., medium, style, subject, mood, description) to streamline prompt creation.

5.2. Knowing AI's capability and freedom of the system

Research has shown that a variety of factors influence end-user expectations and experiences with technology (Olshavsky & Miller, 1972). A representative factor affecting user experience is information about the system's capabilities prior to usage (Kocielnik et al., 2019; Ribeiro et al., 2016). Especially when it comes to recent AI technologies, providing users with informative examples could increase the system's usability due to their highly complex, contextual, and sometimes internally conflicting preferences (Ekstrand et al., 2015; Lyngs et al., 2018). The user's freedom to use the system, however, can also be limited by knowing the system's capabilities. Predefining what the system can do and what it does well might prevent users from utilizing it in other creative ways. This can be a particularly significant issue in a system that emphasizes creativity rather than task-oriented work.

Despite many studies indicating that the information on a system's capabilities should be given to the users (Ekstrand et al., 2015; Lyngs et al., 2018), our results have shown that the level of freedom preferred by users varies based on their expectations and experiences. Informing the participants of the system's capabilities through guides and examples proved to be beneficial because they could get a better understanding of what the system could do. However, whether such features should be explicitly accessible depends on the user. Three participants preferred the current system with a high degree of freedom, while seven others preferred a system with less freedom and more information. It should be noted that the participants (P5, P7, P10) who prefer the current system with a high degree of freedom were experts in design and art. Ordinary users, on the other hand, wanted to directly access the system's different features through a more structured and ostensive interface. Consequently, we should allow individuals to determine their level of system freedom based on personal preferences, experience, expertise, and expectations.

- **Design implication (D3)**: Allow users to customize the freedom of the system by providing users with adequate information about its capabilities according to their preferences, experience, expertise, and expectations.

5.3. Algorithmic reasoning and Adaptation

People use sensemaking when they recognize that their current understanding of events is insufficient and attempt to fill the 'knowledge gap' (Dervin, 1992; Klein et al., 2006). They use several strategies in situations where a great deal of information is irrelevant to the problem at hand, to make sense of the information they need. In this process, a mental model is constructed, verified, and modified to account for the missing information and unrecognized features. Our study also found that sensemaking occurs when users review the output images and deduce the algorithms for achieving the desired results.

The design of AI systems should support the users' reasoning and sensemaking processes. The majority of our participants (7 out of 10) wanted a more structured interface with less freedom. These users can make sense of AI systems more easily with the interface that defines the AI's capabilities beforehand and offers them options. For example, a system can be designed to allow the user to browse and select various style options. The DALL-E prompt book, for instance, provides examples of photography, illustration, art history, and 3D artwork prompts. Users may be able to choose these features or add them to their queries. This direction aligns with the users' expectations of future system improvements based on our interview. As P1 suggested, selecting the style of painting or photo in advance could help users find the *sweet spot* between what they want and what AI can do. Also, P10 pointed out that with a system that checks specific options only the experts can know without directly typing, the public will be able to access more creative opportunities. However, presenting predefined AI capabilities to users may not be sufficient. Providing many examples might help people understand the algorithm more easily and intuitively. Providing users with completely different examples may be a potential macroscopic approach for enabling new ideas and expanding user horizons. Users can also microscopically improve their understanding by observing small but distinct examples.

- **Design implication (D4)**: Assist users to grasp the capabilities of AI, find the best query and ultimately generate the best images they want by providing a variety of examples and options. Some of these examples may be derived from existing knowledge that has already been created and shared by other users in the community.

Furthermore, DALL-E 2 could explain how an image was generated based on a user's prompt, allowing the user to understand and reason how to optimize their prompts (Samek et al., 2019). In spite of the fact that DALL-E is a black-box model, which means that its internal workings cannot be easily understood, there are a few ways that it can explain its output (Xu et al., 2023). Saliency maps, for instance, could be used by DALL-E to show which parts of the input prompt contributed most to the generated image. This can help the user understand which parts of the prompt the model is paying attention to and how they influence the output. Additionally, DALL-E could provide a confidence score on generated images, which can assist the user in understanding the model's confidence in the generated image and making informed choices.

- **Design implication (D5)**: Enhance the interpretability of text-to-image models by providing explainable results such as saliency maps or confidence scores to support the user's reasoning process.

5.4. Two-way communication for human-AI collaboration

We found that users prefer to have two-way interactions with AI rather than giving unilateral instructions (Shneiderman, 2020). The lack of feedback from the AI frustrated users after they tried various methods to understand it. Indeed, the lack of interaction made the experience of using AI more like interacting with a machine than with a human, according to all but two users. The users wanted a two-way interactions in which both AI and users could work together to improve the results. Rather than deliver results one-way, users expect AI to explain why the results were generated and reflect their feedback.

In continuation of the recent academic discussions, this result emphasizes the importance of viewing AI as an independent and cognitive actor rather than as a tool (Horner, 2020). The users wanted AI to serve as a collaborator, not as a tool, and two-way communication is crucial to a successful human-AI collaboration (R. Zhang et al., 2021). Indeed, explicit and implicit communication can enable humans and AI to share a common understanding and achieve the best results (Liang et al., 2019; R. Zhang et al., 2021). For example, besides asking users for feedback (explicit communication), AI can also analyze behavior data such as clicks and saves (implicit communication). The system must implement features that are capable of interacting with the user and receiving feedback from them. Furthermore, this approach also aligns with a body of research on human-AI teaming which indicates that two-way communication improves team performance (M.-T. Hong et al., 2018; Liao et al., 2019).

To enable two-way communication, AI could be able to provide explainable results. In terms of explainable AI, this direction corresponds to providing explanations for "why" (i.e., Why/How is this instance given this output?, What feature(s) of this instance leads to the system's prediction?) (Liao et al., 2020). This information allows users to ask themselves "what if" (e.g., What would the system generate if this feature of the query changes to … ?, What would the system predict for a different query?) questions and modify their queries accordingly. Using an attention map-like interface, P6 proposed highlighting the weighted elements in the query to make the results more explainable. Making results more explainable will make it easier for users to modify their queries.

Aside from providing explainable results, AI can facilitate two-way communication by actively collecting and integrating user feedback. P3 expressed the expectation that AI would collect user responses by collecting likes or dislikes for each image. P8 also said that the system should be developed in a way that the model continuously learns from the user data, thereby generating images that reflect the user's individual preferences. As a result of these findings, AI has the potential to co-create with users via two-way interactions, through which users are informed and provided with feedback on the results, as well as vice versa.

- **Design implication (D6)**: Provide two-way communication between users and AI agents so that they can feel that they are actively interacting with them. Describe the results of AI in detail and suggest a framework for combining user feedback with AI results.

5.5. Ethical issues

A number of ethical concerns have been raised regarding the misuse of the results and the bias in the data. The first concern raised was the potential misuse of AI-generated data. Specifically, as the demand and supply of virtual characters increase and transactions become more active, the likelihood of abuse is increasing. It is possible to make fictional characters appear as real as possible. The synthesis of real people can contribute to social unrest by creating fake news. In addition, P2 mentioned the possibility of abusing a low-quality model built on DALL-E 2 data. A popular image generation model such as DALL-E is easily accessible and can provide the necessary data for training deepfake models. Since deepfake models can generate more sophisticated content with larger data, the data for training deepfake models can be generated from DALL-E. Continual research should be carried out on a technology that can distinguish authentic images from fakes in order to prevent the risk of misuse. Additionally, just as OpenAI prohibits entering real people into queries, it is also crucial to filter keywords and sentences that will generate data that is likely to be used for malicious purposes.

In addition to possible malicious use of models and data, participants noted biases in the training data. Despite OpenAI's new technology for creating images of people that more accurately represent the diversity of the world's population,[6] there was still a gender or race bias in the generated images. For example, P4 perceived that the algorithm was biased in that Asian women appeared in all the results of 'high school students are smiling in front of their SAT scores." P6 also expressed concerns about biased content when all but one or two were male with the query that included 'ph.d.' (Figure 8). As a precaution against a negative user experience, he suggested adding a message in the system that the gender or age can be specified. Additionally, P9 expressed discomfort about the image of a woman that could evoke social stereotypes in response to the term 'housekeeper'. Generative art has long been known to have gender bias, and many scholars are working to reduce this bias (Srinivasan & Uchino, 2021). In order to address this problem, it is necessary to involve users, as well as researchers and AI developers, in the development of algorithms. OpenAI is currently receiving user feedback when there is an issue with the generated output (e.g.," Image contains sensitive or biased content"). A reduction in bias can be achieved by continuously collecting user feedback in this manner and incorporating that feedback into the algorithm (Xu et al., 2023).

- **Design implication (D7)**: Whenever possible, be mindful that artificial intelligence techniques can contain biases or prejudices against women, people of color, etc., and incorporate reporting tools into the interface that can help prevent reoccurrence.

5.6. Limitations and future work

Before we wrap up this paper, we would like to briefly mention the limitations of this study in addition to the research plans that we have for the future. Firstly, Dall-E is only a research probe for our user research, and cannot be said to represent all creative algorithms and interfaces. In order to provide a more generalizable perspective on AI for creativity, we will conduct a user study in the future that explores a wider range of scenarios and areas where AI may be used for creativity. The second point is that although our goal was to carry out an in-depth study with a small number of participants at an initial stage to understand their thoughts on the algorithm in a more detailed way, we still have a small number of participants at this stage. Our future study will consist of recruiting a large number of users that reflect a diverse demographic mix to conduct a user study, which will ensure a more accurate representation of different demographics in terms of demographics. Additionally, we intend to determine how different interactions are according to participants' levels of expertise or understanding of machine learning. Third, our proposed design implications may be influenced by our user study. As part of our future studies, we plan to evaluate the usefulness, effectiveness, and user experience of AI interfaces that incorporate the design considerations we proposed in our study.

6. Conclusion

In this study, we explored users' expectations and experiences of AI-based image generation systems through the case study of DALL-E 2. Our qualitative user study revealed that people have both positive and negative attitudes toward AI-generated art. They also wanted the AI's output to be something unexpected, which they considered to be the essence of the AI's creativity. A user's knowledge of a system's capabilities significantly affects their perception of the user experience in different ways depending on their level of

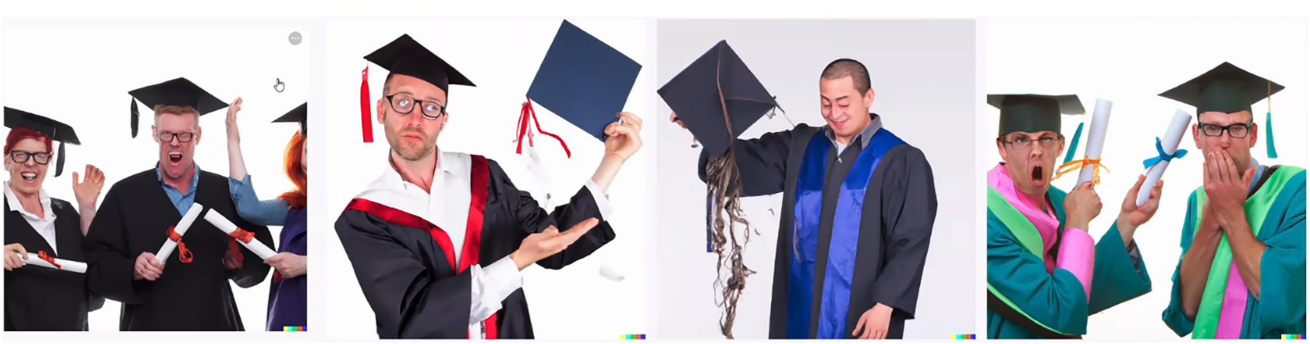

Figure 8. Images generated by DALL-E 2 for the query of "a fresh ph.d. just thrown out in the industry." .

experience and their expectations of the system. Through the continuous refinement of their queries, users were able to build and verify hypotheses about AI as they attempted to get the results they wanted. They also place a high priority on having a two-way communication system with them as well.

We hope that this paper will raise awareness about the emerging algorithmic experiences from researchers and practitioners in the creativity area. Furthermore, we hope the findings of the research will contribute to the development of new interfaces and interactions that incorporate AI.

Notes

1. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html.
2. https://www.midjourney.com/.
3. https://openai.com/dall-e-2/.
4. https://dallery.gallery/the-dalle-2-prompt-book/.
5. https://www.instagram.com/openaidalle/.
6. https://openai.com/blog/reducing-bias-and-improving-safety-in-dall-e-2/.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5B8096358).

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Jinsu Eun http://orcid.org/0000-0003-3051-7193
Changhoon Oh http://orcid.org/0000-0002-3340-551X

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., … Collisson, P., others. (2019). Guidelines for human-ai interaction. In *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems* (pp. 1–13.). https://doi.org/10.1145/3290605.3300233

Ashktorab, Z., Dugan, C., Johnson, J., Pan, Q., Zhang, W., Kumaravel, S., Campbell, M. (2021). Effects of communication directionality and ai agent differences in human-ai interaction. In *Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems* (pp. 1–15.). https://doi.org/10.1145/3411764.3445256

Biermann, O. C., Ma, N. F., Yoon, D. (2022). From tool to companion: Storywriters want ai writers to respect their personal values and writing strategies. In *Designing Interactive Systems Conference* (pp. 1209–1227.). https://doi.org/10.1145/3532106.3533506

Boden, M. (1995). Creativity and unpredictability. *Stanford Humanities Review*, *4*(2), 123–139. https://dl.acm.org/doi/abs/10.5555/212154.212171

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bryman, A. (2016). *Social research methods*. Oxford university press.

Cetinic, E., & She, J. (2022). Understanding and creating art with ai: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *18*(2), 1–22. https://doi.org/10.1145/3475799

Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(2), 177–192. https://doi.org/10.1037/aca0000136

Choung, H., David, P., & Ross, A. (2023). Trust in ai and its role in the acceptance of ai technologies. *International Journal of Human–Computer Interaction*, *39*(9), 1727–1739. https://doi.org/10.1080/10447318.2022.2050543

Coeckelbergh, M. (2017). Can machines create art? *Philosophy & Technology*, *30*(3), 285–303. https://doi.org/10.1007/s13347-016-0231-5

Colton, S., & Wiggins, G. A. (2012). Computational creativity: The final frontier? *Frontiers in Artificial Intelligence and Applications*, *242*, 21–26. https://doi.org/10.3233/978-1-61499-098-7-21

Croitoru, F.-A., Hondru, V., Ionescu, R. T., & Shah, M. (2022). Diffusion models in vision: A survey. *arXiv preprint arXiv: 2209.04747*

Crowson, K., Biderman, S., Kornis, D., Stander, D., Hallahan, E., Castricato, L., & Raff, E. (2022). *Vqgan-clip: Open domain image generation and editing with natural language guidance* [Paper presentation]. In European Conference on Computer Vision, (pp. 88–105).

Daniele, A., Bernardi Luft, C. D., & Bryan-Kinns, N. (2021). *"What is human?" a turing test for artistic creativity* [Paper presentation]. In International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of Evostar), (pp. 396–411).

Daniele, A., Song, Y.-Z. (2019). Ai + art = human. In *Proceedings of the 2019 Aaai/Acm Conference on ai, Ethics, and Society* (pp. 155–161.).

Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In J. D. Glazier & R. R. Powell (Eds.), *Qualitative research in information management* (pp. 61–84). Libraries Unlimited.

Edward, J. D., & Sally, J. (2000). Defining interactivity. *New Media Society*, *2*(2), 157. https://doi.org/10.1177/14614440022225751

Ekstrand, M. D., Kluver, D., Harper, F. M., Konstan, J. A. (2015). Letting users choose recommender algorithms: An experimental study. In *Proceedings of the 9th Acm Conference on Recommender Systems* (pp. 11–18.).

Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv preprint arXiv: 1706.07068*

Fan, J. E., Dinculescu, M., Ha, D. (2019). Collabdraw: An environment for collaborative sketching with an artificial agent. In *Proceedings of the 2019 on Creativity and Cognition* (pp. 556–561.).

Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E., & Sasahara, K. (2022). *Are deepfakes concerning? analyzing conversations of deepfakes on reddit and exploring societal implications* [Paper presentation]. Chi Conference on Human Factors in Computing Systems, In (pp. 1–19). https://doi.org/10.1145/3491102.3517446

Gero, K. I., Chilton, L. B. (2019). Metaphoria: An algorithmic companion for metaphor creation. In *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems* (pp. 1–12.).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Gu, L., & Li, Y. (2022). Who made the paintings: Artists or artificial intelligence? The effects of identity on liking and purchase intention. *Frontiers in Psychology*, *13*, 941163. https://doi.org/10.3389/fpsyg.2022.941163

Hagtvedt, L. P., Dossinger, K., Harrison, S. H., & Huang, L. (2019). Curiosity made the cat more creative: Specific curiosity as a driver of creativity. *Organizational Behavior and Human Decision Processes*, *150*, 1–13. https://doi.org/10.1016/j.obhdp.2018.10.007

Hertzmann, A. (2018). Can computers create art? In *Arts*, *7*(2), 18. https://doi.org/10.3390/arts7020018

Hertzmann, A. (2020). Computers do not make art, people do. *Communications of the ACM*, *63*(5), 45–48. https://doi.org/10.1145/3347092

Hong, J.-W., & Curran, N. M. (2019). Artificial intelligence, artists, and art: Attitudes toward artwork produced by humans vs. artificial intelligence. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *15*(2s), 1–16. https://doi.org/10.1145/3326337

Hong, J.-W., Fischer, K., Ha, Y., & Zeng, Y. (2022). Human, I wrote a song for you: An experiment testing the influence of machines' attributes on the ai-composed music evaluation. *Computers in Human Behavior*, *131*, 107239. https://doi.org/10.1016/j.chb.2022.107239

Hong, M.-T., Benjamin, J. J., & Müller-Birn, C. (2018). *Coordinating agents: Promoting shared situational awareness in collaborative sensemaking* [Paper presentation]. In Companion of the 2018 Acm Conference on Computer Supported Cooperative Work and Social Computing, (pp. 217–220).

Horner, S. (2020). Unthought: The power of the cognitive unconscious, by n. katherine hayles. *Academy of Management Learning & Education*, *19*(3), 428–431. https://doi.org/10.5465/amle.2020.0312

Huang, C.-Z A., Koops, H. V., Newton-Rex, E., Dinculescu, M., & Cai, C. J. (2020). Ai song contest: Human-ai co-creation in songwriting. *arXiv preprint arXiv:2010.05388*

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model.*IEEE Intelligent Systems*, *21*(5), 88–92. https://doi.org/10.1109/MIS.2006.100

Ko, H.-K., Park, G., Jeon, H., Jo, J., Kim, J., & Seo, J. (2022). Large-scale text-to-image generation models for visual artists' creative works. arXiv preprint arXiv:2210.08477

Koch, J., Taffin, N., Beaudouin-Lafon, M., Laine, M., Lucero, A., & Mackay, W. E. (2020). Imagesense: An intelligent collaborative ideation tool to support diverse human-computer partnerships. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW1), 1–27. https://doi.org/10.1145/3392850

Kocielnik, R., Amershi, S., Bennett, P. N. (2019). Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems* (pp. 1–14.).

Lee, C., Kim, S., Han, D., Yang, H., Park, Y.-W., Kwon, B. C., Ko, S. (2020). Guicomp: A gui design assistant with real-time, multi-faceted feedback. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–13.). https://doi.org/10.1145/3313831.3376327

Lee, M., Liang, P., & Yang, Q. (2022). *Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities* [Paper presentation]. Chi Conference on Human Factors in Computing Systems, In (pp. 1–19). https://doi.org/10.1145/3491102.3502030

Liang, C., Proft, J., Andersen, E., Knepper, R. A. (2019). Implicit communication of actionable information in human-ai teams. In *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems* (pp. 1–13.). https://doi.org/10.1145/3290605.3300325

Liao, Q. V., Gruen, D., Miller, S. (2020). Questioning the ai: Informing design practices for explainable ai user experiences. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–15.). https://doi.org/10.1145/3313831.3376590

Liao, Q. V., Wang, Y.-C., Bickmore, T., Fung, P., Grudin, J., Yu, Z., & Zhou, M. (2019). *Human-agent communication: Connecting research and development in hci and ai* [Paper presentation]. In Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, (pp. 122–126).

Lin, Y., Guo, J., Chen, Y., Yao, C., Ying, F. (2020). It is your turn: Collaborative ideation with a co-creative robot through sketch. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–14.). https://doi.org/10.1145/3313831.3376258

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*

Liu, V., & Chilton, L. B. (2022). *Design guidelines for prompt engineering text-to-image generative models* [Paper presentation]. In Chi Conference on Human Factors in Computing Systems, (pp. 1–23). https://doi.org/10.1145/3491102.3501825

Louie, R., Coenen, A., Huang, C. Z., Terry, M., Cai, C. J. (2020). Novice-ai music co-creation via ai-steering tools for deep generative models. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–13.). https://doi.org/10.1145/3313831.3376739

Lyngs, U., Binns, R., Van Kleek, M., Shadbolt, N. (2018). So, tell me what users want, what they really, really want!. In *Extended Abstracts of the 2018 Chi Conference on Human Factors in Computing Systems* (pp. 1–10.).

Maxwell, W., Beaudouin, V., Bloch, I., Bounie, D., Clémençon, S., d'Alché Buc, F., … Parekh, J. (2020). Identifying the'right'level of explanation in a given situation. In *Proceedings of the First International Workshop on New Foundations for Human-Centered ai (Nehuai), Santiago de Compostella, Spain* (p. 63.).

Mikalonytė, E. S., & Kneer, M. (2022). Can artificial intelligence make art? folk intuitions as to whether ai-driven robots can be viewed as artists and produce art. *ACM Transactions on Human-Robot Interaction*, *11*(4), 1–19. https://doi.org/10.1145/3530875

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

Oh, C., Kim, S., Choi, J., Eun, J., Kim, S., Kim, J., … Suh, B. (2020). Understanding how people reason about aesthetic evaluations of artificial intelligence. In *Proceedings of the 2020 Acm Designing Interactive Systems Conference* (pp. 1169–1181.).

Oh, C., Lee, T., Kim, Y., Park, S., Kwon, S., Suh, B. (2017). Us vs. them: Understanding artificial intelligence technophobia over the google deepmind challenge match. In *Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems* (pp. 2523–2534.).

Oh, C., Song, J., Choi, J., Kim, S., Lee, S., Suh, B. (2018). I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems* (pp. 1–13.).

Olshavsky, R. W., & Miller, J. A. (1972). Consumer expectations, product performance, and perceived product quality. *Journal of Marketing Research*, *9*(1), 19–21. https://doi.org/10.1177/002224377200900105

Osone, H., Lu, J.-L., Ochiai, Y. (2021). Buncho: Ai supported story co-creation via unsupervised multitask learning to increase writers' creativity in japanese. In *Extended Abstracts of the 2021 Chi Conference on Human Factors in Computing Systems* (pp. 1–10.).

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., … Agarwal, S., others. (2021). Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning* (pp. 8748–8763.).

Ragot, M., Martin, N., Cojean, S. (2020). 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* (pp. 1–10.).

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., … Sutskever, I. (2021). Zero-shot text-to-image generation. In *International Conference on Machine Learning* (pp. 8821–8831.).

Reynolds, L., McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 Chi Conference on Human Factors in Computing Systems* (pp. 1–7.). https://doi.org/10.1145/3411763.3451760

Ribeiro, M. T., Singh, S., Guestrin, C. (2016). "Why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144.).

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (2019). *Explainable ai: Interpreting, explaining and visualizing deep learning.* (Vol. 11700). Springer Nature.

Sanh, V., Webson, A., Raffel, C., Bach, S. H., Sutawika, L., … Alyafeai, Z., others. (2021). Multitask prompted training enables zero-shot task generalization. *arXiv preprint arXiv:2110.08207*

Shi, Y., Cao, N., Ma, X., Chen, S., Liu, P. (2020). Emog: Supporting the sketching of emotional expressions for storyboarding. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–12.). https://doi.org/10.1145/3313831.3376520

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Srinivasan, R., Uchino, K. (2021). Biases in generative art: A causal look from the lens of art history. In *Proceedings of the 2021 Acm Conference on Fairness, Accountability, and Transparency* (pp. 41–51.).

Van Someren, M., Barnard, Y. F., & Sandberg, J. (1994). *The think aloud method: A practical approach to modelling cognitive*. AcademicPress. *11*, 29–41.

Walsh, G., & Wronsky, E. (2019). *Ai + co-design: Developing a novel computer-supported approach to inclusive design* [Paper presentation]. In Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, (pp. 408–412).

Wang, Y., Shen, S., Lim, B. Y. (2023). Reprompt: Automatic prompt editing to refine ai-generative art towards precise expressions. In *Proceedings of the 2023 Chi Conference on Human Factors in Computing Systems* (pp. 1–29.).

Wu, Y., Mou, Y., Li, Z., & Xu, K. (2020). Investigating american and chinese subjects' explicit and implicit perceptions of ai-generated artistic work. *Computers in Human Behavior*, *104*, 106186. 106186. https://doi.org/10.1016/j.chb.2019.106186

Xie, Y., Pan, Z., Ma, J., Jie, L., & Mei, Q. (2023). *A prompt log analysis of text-to-image generation systems. In Proceedings of* [Paper presentation].The Acm Web Conference 2023, (pp. 3892–3902). https://doi.org/10.1145/3543507.3587430

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to human interaction with ai systems: New challenges and opportunities for hci professionals to enable human-centered ai. *International Journal of Human–Computer Interaction*, *39*(3), 494–518. https://doi.org/10.1080/10447318.2022.2041900

Zhang, N., Li, L., Chen, X., Deng, S., Bi, Z., Tan, C., … Chen, H. (2021). Differentiable prompt makes pre-trained language models better few-shot learners. *arXiv preprint arXiv:2108.13161*

Zhang, R., McNeese, N. J., Freeman, G., & Musick, G. (2021). "An ideal human" expectations of ai teammates in human-ai teaming. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3), 1–25. https://doi.org/10.1145/3432945

About the authors

Soomin Kim obtained her Ph.D. in Communication from the Seoul National University. She is currently working as a senior designer. Her research focuses on Human-AI Interaction, conversational agents, and user experience design. She integrates user experience insights with AI approaches in her studies.

Jinsu Eun, a PhD candidate at Seoul National University's Institute of Convergence Science and Technology, excels in data visualization, robotic journalism, and chatbot tech. His research focuses on innovative, user-friendly methods to present complex data, combining advanced technology with user-centered design for effective communication.

Changhoon Oh is an Assistant Professor in the Graduate School of Information at Yonsei University. He also worked as a Postdoctoral Researcher in the Human-Computer Interaction Institute at Carnegie Mellon University. He earned his Ph.D. from Seoul National University. His research encompasses Human-AI Interaction, UX design, and responsible AI.

Joonhwan Lee is a Professor in the Department of Communication at Seoul National University. He holds a Ph.D. in Human–Computer Interaction (2008) from Carnegie Mellon University. His research includes HCI, social computing, situationally appropriate user interaction, and information visualization. He directs the Human-Computer Interaction + Design Lab (hcid.snu.ac.kr).