I Lead, You Help But Only with Enough Details: Understanding the User Experience of Co-Creation with Artificial Intelligence

Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, Bongwon Suh
Human Centered Computing Lab
Graduate School of Convergence Science and Technology, Seoul National University
{yurial, jwsong0617, jinhanchoi, kim.seonghyeon, lsw0504, bongwon}@snu.ac.kr

ABSTRACT

Recent advances in artificial intelligence (AI) have increased the opportunities for users to interact with the technology. Now, users can even collaborate with AI in creative activities such as art. To understand the user experience in this new user–AI collaboration, we designed a prototype, *DuetDraw*, an AI interface that allows users and the AI agent to draw pictures collaboratively. We conducted a user study employing both quantitative and qualitative methods. Thirty participants performed a series of drawing tasks with the think-aloud method, followed by post-hoc surveys and interviews. Our findings are as follows: (1) Users were significantly more content with DuetDraw when the tool gave detailed instructions. (2) While users always wanted to lead the task, they also wanted the AI to explain its intentions but only when the users wanted it to do so. (3) Although users rated the AI relatively low in predictability, controllability, and comprehensibility, they enjoyed their interactions with it during the task. Based on these findings, we discuss implications for user interfaces where users can collaborate with AI in creative works.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

Artificial intelligence; human computer collaboration; human-AI interaction

INTRODUCTION

It is the age of artificial intelligence (AI), and recent advances in deep learning have yielded AI with capabilities comparable to those of humans in various fields [20, 30, 39]. Many interactions have been introduced based on this technology, such as voice user interfaces and autopilots of self-driving cars.
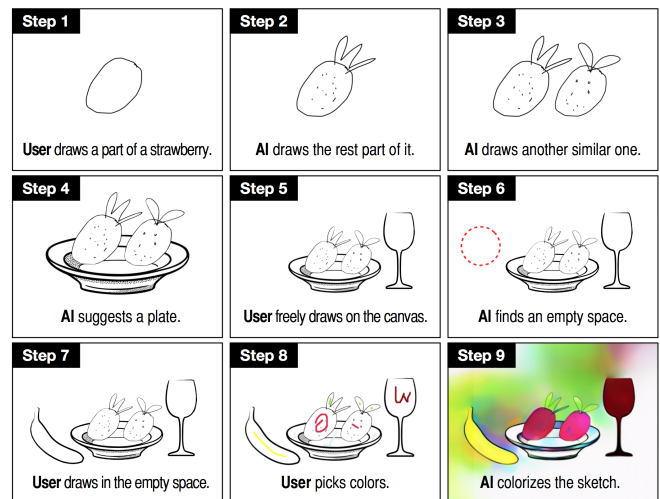
Figure 1. Drawing using DuetDraw in the Lead mode. With DuetDraw, users and AI can collaboratively draw pictures.

AI is expected to become increasingly prevalent in numerous areas [3, 6, 16, 59]. It will not only assist humans in repetitive and arduous tasks but also take on complex and elaborative works [32, 38, 49]. Moreover, while humans can guide AI, AI can also guide humans [26, 55]. They can even work together to produce reasonable results in various creative tasks, including writing, drawing, and problem solving [4, 12, 43].

As users and AI are now interacting in these novel ways, understanding the user experience with these intelligent interfaces has become a critical issue in the human–computer interaction community [11, 25, 44, 56]. Many HCI researchers have conducted user studies on various AI interfaces [28, 42, 58], and the concept of *algorithmic experience* has been suggested as a new perspective from which to view the user experience of AI interfaces [41]. In light of this, understanding this new user experience and designing better AI interfaces will require consideration of the following: How do users and AI communicate in creative contexts? Would users like to take the initiative or let AI take it when they cooperate? What factors are associated with the various experiences in this process?

To explore user–AI collaboration, we designed a prototype, *DuetDraw*, with which AI and users can draw pictures in a

collaborative manner. DuetDraw contains a variety of AI-based functions. Using state-of-the-art AI techniques, the tool can help users perform drawing tasks, such as completing the rest of the object that the user was drawing, drawing the same object in a different style, suggesting an object that matches the picture, finding an empty space on the canvas, and automatically colorizing the sketches (Figure 1).

To understand the user experience of user–AI collaboration, we conducted a user study of DuetDraw with both quantitative and qualitative approaches. We focused on the effects of *communication* (Detailed/Basic) and *initiative* (Lead/Assist) on the user experience. By combining the two factors, we designed four experimental conditions (*Lead/Assist*)×(*Detailed/Basic*) and one control condition (*no-AI*). We recruited 30 participants and asked them to conduct a series of drawing tasks with five conditions. We gathered users' feedback during the tasks with the think-aloud method. We also conducted post-hoc surveys and semi-structured interviews. The results of the study indicated the following:

- Users prefer detailed instructions to basic instructions when communicating with AI.

- Users always want to take the initiative, and they want AI to provide detailed explanations about its process but only when they want it to do so.

- AI can provide users with fun as well as useful, effective, and efficient experiences.

- AI can lower users' perceived predictability, comprehensibility, and controllability of the drawing tasks, while detailed instructions can offset these adverse effects. Moreover, low predictability can even increase users' enjoyment.

Based on these findings, we discuss the design implications for user interfaces with which users and AI can closely cooperate on creative work.

The main contributions of this work to the HCI community are as follows:

- We designed and created an interface based on neural network technology, thus pioneering the UX of AI-embedded interfaces.

- Through both quantitative and qualitative approaches, we closely observed the interaction between AI and users and discovered new aspects of this interaction.

- Finally, we discussed implications for interfaces with which users and AI closely communicate and cooperate for creative work.

RELATED WORK
We reviewed related works on (1) AI, deep learning, and new UX in creative works and (2) communication and leadership between humans and computers.

AI, Deep Learning, and New UX in Creative Works
The rise of AI in recent years is largely due to the development of deep learning. It has introduced not only technological innovation [31, 46, 50] but also new interfaces [15, 36], providing

users with experiences that they have never experienced before. It is also being applied in creative areas considered unique to humans, such as writing [4, 51], musical composition [7, 22], and drawing [5].

In drawing, several new interfaces using deep learning have been introduced. *Quick, Draw!* [24] is an online game that challenges players to draw a picture of an object, and then AI guesses what the drawings represent using a neural network. *AutoDraw* [29] recognizes a hand-drawn doodle and suggests its clean clip art replacement. Davis et al. developed *Drawing Apprentice*, which can collaborate with users by analyzing their drawn input and responding in real time in improvisational interactions [9, 10].

Various deep learning algorithms that can support these drawing interfaces have been developed. *Sketch-RNN* [17] is a recurrent neural network for constructing stroke-based drawings of common objects; it can mimic human figures and draw pictures. When a user starts drawing a shape, it automatically completes the drawing. It can also generate similar but unique objects. *PaintsChainer*, a CNN-based line drawing colorizer, can automatically paint any sketch [60]. Gatys et al. also developed a neural network for blending the content of one image and the style of another image [13]. It can transform any image into a classical painter's style.

Although these new interfaces and algorithms are still in the experimental stages, they have opened up the possibility for humans and AI to work together to produce creative outcomes. This thus calls for a new research agenda: understanding users' perceptions of these new technologies and developing design guidelines to improve UX [25, 44]. In this respect, by combining AI algorithms and perspectives of prior studies, we designed a prototype with which humans and AI can produce complex and creative output such as drawn pictures.

Communication and Leadership among Users and AI
How humans and computers should communicate and who should take the initiative in their interaction has been studied as a primary subject in HCI. In the case of communication, there has been discussion about whether providing users with detailed instruction is beneficial [45]. Detailed instructions, such as dialogue, modal windows, and alerts, can help users to complete tasks and reach their goals more quickly and easily and direct users' attention to the tasks. However, they can frustrate users when they are wrong or when they interrupt users' performance [2, 45]. Given these advantages and disadvantages, it is important to design an appropriate communication style in accordance with the characteristics of each interface.

In the context of user–AI interfaces, especially those in which users and AI closely collaborate, the communication issue is also significant. AI algorithms are often considered black boxes [27]; that is, it is difficult to convey their operational processes and principles to users. Thus, it is important to identify the appropriate communication method to enhance the user experience of novice interface users.

Meanwhile, there has been discussion about whether users or computers should take the initiative in the interaction. The most notable debate concerns whether direct manipulation or

interface agents should be employed [11, 48]. Researchers supporting the former mainly claim that direct manipulation affords the user control and predictability in their interfaces. In contrast, researchers supporting the latter argue that users have to delegate certain tasks or certain parts of tasks to agents. Further studies have been conducted on how to take the initiative with an agent when arranging collaboration between users and computers [21, 53].

In designing interfaces in which users and AI collaborate in creative work, the initiative issue could also be a critical factor. Since creative work has been considered human-specific, it is important to understand humans' perceptions of initiative in collaborating with AI and consider them in design.

Based on this background, in this study, we focus on the communication and initiative issues and explore how these can affect the user experience of interfaces in which users and AI work together.

DUET DRAW

To understand the user experience of user–AI collaboration, we designed a research prototype, *DuetDraw*, where AI and the user draw a picture together (Figure 1). The tool runs on a Chrome browser using P5.js [37], a JavaScript library for sketchbook software. For AI-based functions, DuetDraw uses the open source code of Google's Sketch-RNN [17] and PaintsChainer [60]. Users can draw pictures using DuetDraw on a tablet PC with a stylus pen. We used an iPad Pro 12.9-inch model and Apple Pencil as an experimental apparatus.

Five AI Functions of DuetDraw

Users can create collaborative drawings with the help of the various functions of DuetDraw. Specifically, DuetDraw provides five functions based on AI technologies.

- *Drawing the rest of an object:* This function enables the AI to automatically complete an object that a user has drawn. When a user stops drawing an object, this function enables the AI to immediately draw the rest of the object (Step 2 in Figure 1). It is based on Google's Sketch-RNN [17].

- *Drawing an object similar to a previous object:* This function enables the AI to draw the same object that a user has just drawn in a slightly different form (Step 3 in Figure 1). The object is drawn to the right of the existing object and at the same scale. It is also based on Sketch-RNN [17].

- *Drawing an object that matches previous objects:* This function enables the AI to draw another object that matches the objects a user has just drawn. A clip-art-like object is drawn on the canvas considering the other objects' positions (Step 4 in Figure 1).

- *Finding an empty space on the canvas:* This function enables the AI to find and display an empty space on the canvas. We implemented this by devising an algorithm finding the space where the biggest circle can be drawn without overlapping with the drawn objects (Step 6 in Figure 1).

- *Colorizing sketches with recommended colors:* This function enables the AI to colorize sketches based on a user's
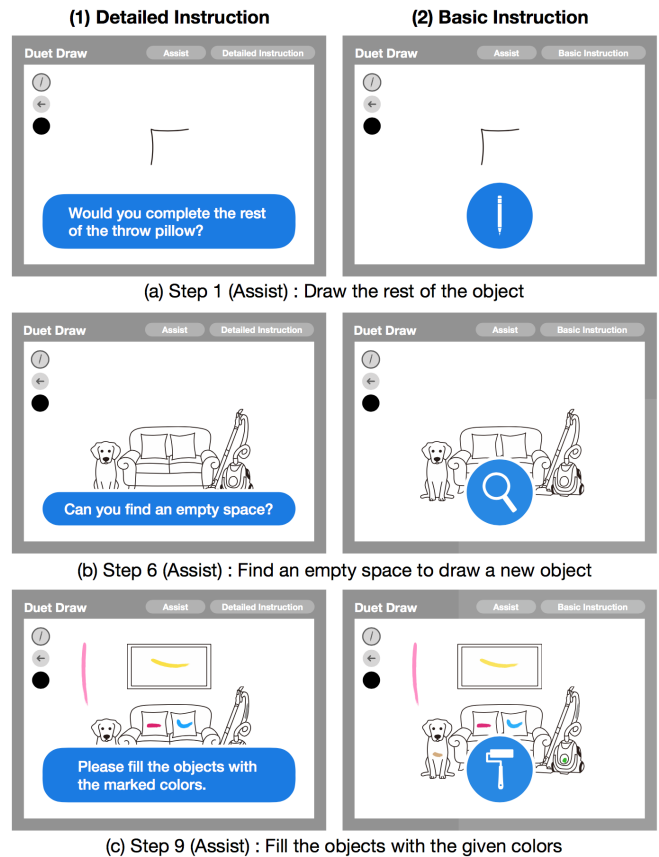

(a) Step 1 (Assist) : Draw the rest of the object

(b) Step 6 (Assist) : Find an empty space to draw a new object

(c) Step 9 (Assist) : Fill the objects with the given colors

Figure 2. Examples of two different communication styles of DuetDraw.

color choices. When the user chooses colors from the palette and marks them on each object with a line, this function automatically paints the entire picture according to the colors. It is implemented using PaintsChainer [60], a CNN-based line drawing colorizer (Step 9 in Figure 1).

Initiative and Communication Styles of DuetDraw

In designing DuetDraw, we considered two main factors, initiative and communication, and devised two different styles for each factor.

- **Initiative:** There are two initiative styles: *Lead* and *Assist*. In the *Lead* style, users complete their pictures with the help of the AI. In this mode, users take the initiative. Users draw a major portion of the figure, and the AI then carries out secondary tasks. In contrast, in the *Assist* style, users help AI to complete the picture. In this mode, the AI takes the initiative. The AI draws the main parts of the picture and asks users to complete supplementary/subsidiary parts.

- **Communication:** There are two styles of communication: *Detailed Instruction* and *Basic Instruction*. In *Detailed Instruction*, the AI explains each step and guides the user. At the bottom of the interface, an instruction is displayed as a message, and users can confirm the message by tapping yes or no buttons. On the contrary, in *Basic Instruction*, the AI automatically proceeds to the next step with basic notifications. An instruction is displayed as an icon on the canvas (More detailed examples are given in Figure 2).

STUDY DESIGN

To assess the user experience of DuetDraw from various angles, we designed a user study consisting of a series of drawing tasks, post-hoc surveys and semi-structured interviews.

Participants

We recruited participants by posting an announcement on our institution's online community website. We recruited 30 participants (15 males and 15 females). Their mean age was 29.07, and the SD was 4.74 (M: Mean = 30.53, SD = 5.42, F: Mean = 27.6, SD = 3.54). Before the experiment, we explained the purpose and procedures to the participants. As we identified that it is important to prevent the participants from heavily weighting their first impressions of the interface through the pilot test, we devised ways to make them get used to the system. We specially prepared a separate guide document describing the functions, modes and conditions, and scenarios of DuetDraw in as much detail as possible. We also let the participants try out the system a few times. Each experiment lasted about 1 hour, and each participant received a gift certificate valued at about $10 in exchange for participating in the experiment.

Tasks and Procedures

For the experiments, we designed five conditions for using DuetDraw: four treatment conditions that combined its initiative and communication styles ((a) *Lead-Detailed*, (b) *Lead-Basic*, (c) *Assist-Detailed*, (d) *Assist-Basic*) and one control condition ((e) *no-AI*). The *no-AI* condition had the same interface but no interaction with AI so that users could complete the picture independently on an empty canvas. The experiments had a within-subjects design in which all users performed all five conditions. To reduce the bias due to the sequence of tasks, we randomized the orders of the five conditions.

Drawing Scenarios

Although users can normally draw and color any object, for the experiments, it was necessary to control the users' behaviors through assigning tasks rather than letting them perform too many different actions. Therefore, we designed user scenarios consisting of the following nine steps (Table 1) in which the AI and the user drew a picture together.

In the experiments, in the *Lead* conditions, the user is the leader and the AI is the assistant. The user performs steps 1, 5, 7, and 8, leading the drawing. The AI performs steps 2, 3, 4, 6, and 9. Conversely, in the *Assist* conditions, the user and the AI do the opposite: the AI is the leader, and the user is the assistant. In the *Detailed Instruction* conditions, the AI provides detailed information, waiting for the user's confirmation on each step. In the *Basic Instruction* conditions, the AI automatically goes to the next step without detailed guidance and explanation.

We also limited the kinds of pictures and objects that users can draw to conduct an accurate and controlled experiment. In every drawing task, the participants select one of three types of drawings: landscape, still-life, or portrait. Although Sketch RNN provides recognition and completion function for over 100 objects, there are quality differences depending on each

Step	Description
1	The leader starts to draw a part of an object.
2	The assistant completes the rest of the object.
3	The assistant draws the same object in a different style.
4	The assistant draws another object that matches the objects.
5	The leader freely draws on the canvas.
6	The assistant finds an empty space to draw a new object.
7	The leader draws an appropriate object in the empty space.
8	The leader chooses colors and marks them on each object.
9	The assistant colorizes the sketch with the chosen colors.

Table 1. Scenario of drawing a picture with DuetDraw

object. Thus we have selected three best recognized objects that would be easy to work with and assigned these to each category of the drawing. Accordingly, when users are in the leader role, they were asked to start the task by drawing a palm tree when chosen landscape, a strawberry when chosen still-life, and a left eye when in portrait.

Survey

We conducted a survey to quantitatively evaluate the user experience of DuetDraw. At the end of each task, the participants filled out the questionnaires about the condition. The survey consisted of 15 items. We selected 12 items from the criteria commonly used for user interface usability and user experience evaluations [1, 35] in consideration of the characteristics of the tasks: *1) useful, 2) easy to use, 3) easy to learn, 4) effective, 5) efficient, 6) comfortable, 7) communicative, 8) friendly, 9) consistent, 10) fulfilling, 11) fun*, and *12) satisfying*. In addition, we included three extra criteria that have been pointed out in the AI interface issue [18, 23, 48]: *13) predictability, 14) comprehensibility*, and *15) controllability*. Users evaluated each task on the survey with a 7-point Likert scale ranging from highly disagree to highly agree.

Think-aloud and Interview

We also conducted a qualitative study using the think-aloud method and semi-structured interviews to gain a deeper and more detailed understanding of user experience in collaboration with AI. Since we asked the participants to use the think-aloud method while performing the tasks [34], they could freely express their thoughts about the tasks in real time. We video recorded all the experiments and audio recorded all the think-aloud sessions.

After all tasks were completed, we conducted semi-structured interviews. In the interviews, the participants were asked about their overall impressions of DuetDraw, their thoughts on the two different styles of initiative and communication, and each of the functions of the AI. In this process, we used the photo projective technique [8], showing users the pictures they had just drawn so that they could easily recall their memories of the tasks. All the interviews were audio recorded.

Analysis Methods

From the study, we were able to gather two kinds of data: quantitative data from the surveys and qualitative data from the think-aloud sessions and interviews. We conducted quantitative analysis for the former and qualitative analysis for the latter, which are described in detail below.
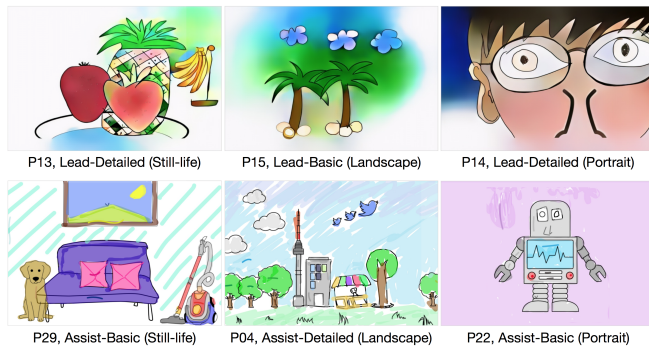
P13, Lead-Detailed (Still-life)	P15, Lead-Basic (Landscape)	P14, Lead-Detailed (Portrait)
P29, Assist-Basic (Still-life)	P04, Assist-Detailed (Landscape)	P22, Assist-Basic (Portrait)

Figure 3. Pictures drawn by participants in experiment.

Quantitative Analysis
In quantitative analysis, we aimed to examine if there was a significant difference between users' evaluation of each condition and the way in which these differences could be explained. As every participant performed all five tasks (within-subjects design), we analyzed the survey data using a one-way repeated-measures ANOVA, comparing the effect of each condition on the user experience of the interface. We also conducted Tukey's HSD test as a post-hoc test for pairwise comparisons.

Qualitative Analysis
The qualitative data from the think-aloud sessions and interviews were transcribed and analyzed using grounded theory techniques [14]. The analysis consisted of three stages. In the first stage, all research team members reviewed the transcriptions together and shared their ideas, discussing main issues observed in the experiments and interviews. We repeated this stage three times to develop our views on the data. In the second stage, we conducted keyword tagging and theme building using Reframer [57], a qualitative research software tool provided by Optimal Workshop. We segmented the transcripts into sentences and finally obtained 635 observations. While reviewing the data, we annotated multiple keyword tags in each sentence so that the keywords could summarize and represent the entire content. A total of 365 keyword tags were created, and we reviewed the tags and text a second time. Then, by combining the relevant tags, we conducted a theme-building process, yielding 30 themes from the data. In the third stage, we refined, linked, and integrated those themes into four main categories. (The quotes are translated into English.)

RESULTS
Through the user study, we obtained the questionnaire responses from the survey, transcriptions from the interviews and think-aloud sessions, and 150 drawings drawn by 30 participants (Figure 3). The results of the analysis are as follows.

Result 1: Quantitative Analysis
The repeated measures one way ANOVA revealed that there are significant effects of conditions on users' ratings on user experience. Except for *fulfilling*, all the 14 items showed significant difference: *useful, easy to use, easy to learn, effective, efficient, comfortable, communicative, friendly, consistent, fun, satisfying, predictable, comprehensible, controllable* (F-values and p-values are shown in Figure 4).

Based on the result, we further conducted Tukey's HSD test as a post-hoc test to identify pairwise comparisons between each condition. As there were 150 comparisons and 58 significantly different pairs among them, we categorized the results focusing on the main issues below.

Detailed Instruction is Preferred over Basic Instruction
From the multiple pair comparisons, we observed that the participants tended to prefer *Detailed Instruction* to *Basic Instruction*. Specifically, we checked if communication mode significantly affected users' ratings when the drawing mode was the same.

First, when the initiative style was Lead, we identified that nine items among the 15 showed that *Detailed Instruction* was placed significantly higher than *Basic Instruction* (Comparison 1 in Table 2, t-values and p-values are shown in the table): *easy to use, easy to learn, effective, comfortable, communicative, friendly, consistent, comprehensible, controllable*. Even though the differences were not significant, these trends were the same in the remaining six items. Second, when the initiative style was Assist, we observed the same pattern and significant differences (Comparison 2 in Table 2): *easy to learn, effective, comfortable, communicative, friendly, consistent*. Even though the differences were not significant, these trends were the same in the remaining six items.

UX Could Be Worse with Lead-Basic than Assist-Detailed
One of the most interesting results of the survey analysis was that user experience could be lower when users were provided *Basic Instruction* with initiative than when provided *Detailed Instruction* without initiative. The pairwise comparison analysis result indicated that in 9 of the 15 items, (b) *Lead-Basic* produced significantly lower scores than (c) *Assist-Detailed* (Comparison 3 in Table 2): *easy to use, easy to learn, effective, comfortable, communicative, friendly, consistent, predictable, comprehensible*. Even though the differences were not significant, these trends were the same in the remaining items except for *fun*. This result suggests that the problem related to communication with AI could be more significant than that related to the initiative issue.

AI is Fun, Useful, Effective, and Efficient
We also identified that the treatment conditions received higher scores in all four tasks than the control condition (Comparisons 4–7 in Table 3, t-values and p-values are shown in the table): *useful, effective, efficient, fun*. In the case of *useful, effective,* and *efficient*, when the *Detailed Instruction* was provided, both *Lead* and *Assist* showed significantly higher scores than *Basic* (Comparisons 4, 6 in Table 3). These items are related to the basic usability of the interface, and we think that the interactions with AI could be helpful for users' task performance itself. Besides, in the case of *fun*, the treatment conditions showed significantly higher scores than the control condition in all four modes (Comparisons 4–7 in Table 3). This shows that the interaction with AI can bring fun and excitement to the user as well as enhance basic usability.

No-AI is more Predictable, Comprehensible, and Controllable
However, as pointed out in previous studies [18, 23, 48], the treatment conditions recorded lower scores for the *predictable,*
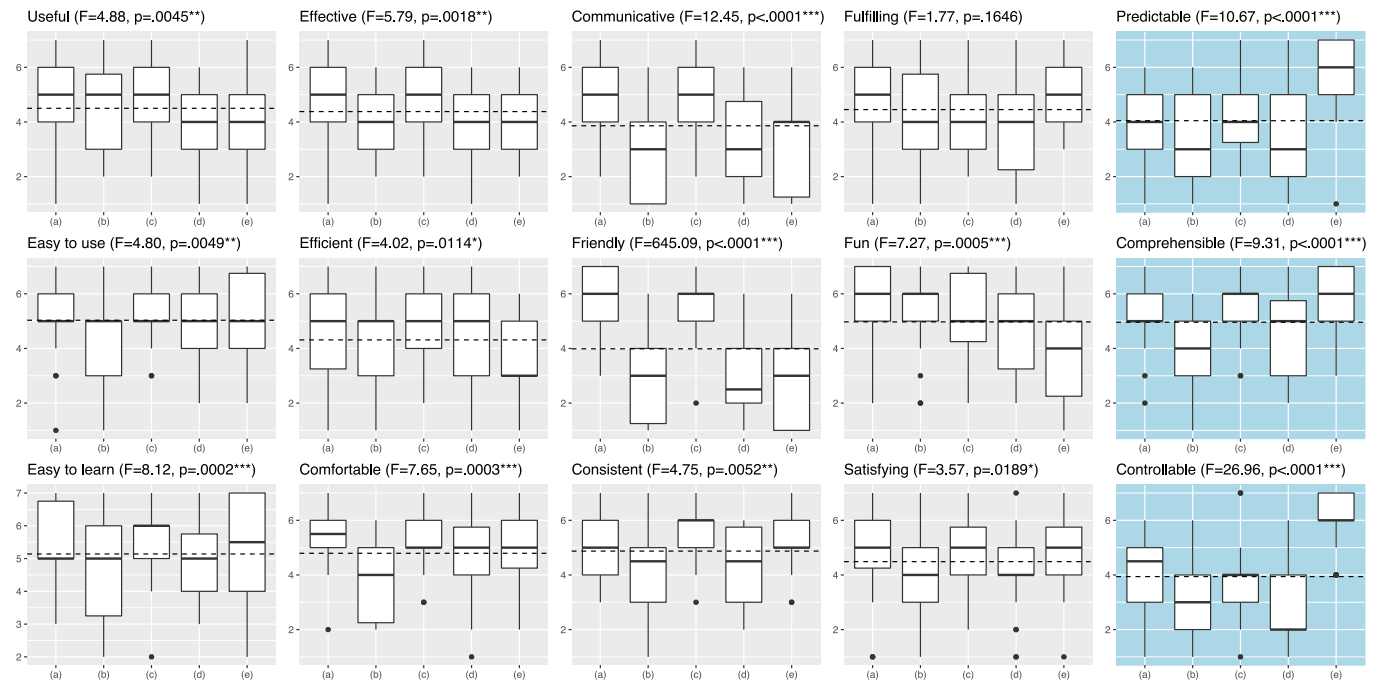
Figure 4. Box plots of user ratings of each item according to each condition and result of one-way repeated-measures ANOVA. Except for *fulfilling*, all items showed significant differences. ((a) *Lead-Detailed*, (b) *Lead-Basic*, (c) *Assist-Detailed*, (d) *Assist-Basic*, (e) *no-AI*, F(4, 26). The dotted lines represent the mean of each item. The items in the rightmost column with the light blue background are related to AI-specific issues. Statistically significant results are reported as *p < 0.001***, p < 0.01**, p < 0.05**)

	Comparison 1			Comparison 2			Comparison 3		
	(a) Lead-Detailed - (b) Lead-Basic			**(c) Assist-Detailed - (d) Assist-Basic**			**(b) Lead-Basic - (c) Assist-Detailed**		
item	*difference*	*t-value*	*p-value*	*difference*	*t-value*	*p-value*	*difference*	*t-value*	*p-value*
useful	0.57	1.653	0.4673	0.80	2.333	0.1420	-0.57	-1.653	0.4673
easy to use	1.00	3.411	0.0078**	0.54	1.819	0.3675	-1.04	-3.525	0.0054**
easy to learn	0.90	3.239	0.0133*	0.90	3.239	0.0133**	-0.94	-3.359	0.0092**
effective	1.03	3.198	0.0151*	1.03	3.198	0.0151**	-1.00	-3.095	0.0205*
efficient	0.66	1.817	0.3691	0.50	1.362	0.6528	-0.60	-1.635	0.4784
comfortable	1.34	4.828	<.0001***	0.80	2.897	0.0358*	-1.37	-4.949	<.0001***
communicative	2.53	6.898	<.0001***	2.00	5.446	<.0001***	-2.37	-6.444	<.0001***
friendly	2.97	8.830	<.0001***	2.53	7.540	<.0001***	-2.67	-7.937	<.0001***
consistent	0.96	3.400	0.0081**	1.14	3.986	0.0011**	-1.30	-4.573	0.0001***
fulfilling	0.47	1.233	0.7321	0.23	0.617	0.9722	0.07	0.176	0.9998
fun	0.20	0.605	0.9740	0.40	1.210	0.7455	0.33	1.008	0.8510
satisfying	0.87	2.574	0.0819	0.57	1.683	0.4484	-0.77	-2.277	0.1598
predictable	0.37	1.026	0.8427	0.84	2.333	0.1421	-1.07	-2.986	0.0280*
comprehensible	1.17	4.148	0.0006***	1.00	3.556	0.0049**	-1.47	-5.215	<.0001***
controllable	1.13	3.220	0.0141*	0.70	1.989	0.2780	-0.60	-1.705	0.4352

Table 2. Results of Tukey's HSD test. Results of Comparison 1 ((a) > (b)) and Comparison 2 ((c) > (d)) show that participants preferred *Detailed* to *Basic* Instruction. Results of Comparison 3 ((c) > (b)) show that *Assist-Detailed* provides a better experience than *Lead-Basic*.

	Comparison 4			Comparison 5			Comparison 6			Comparison 7		
	(a) Lead-Detailed - (e) no-AI			**(b) Lead-Basic - (e) no-AI**			**(c) Assist-Detailed - (e) no-AI**			**(d) Assist-Basic - (e) no-AI**		
item	*diff.*	*t*	*p*	*diff.*	*t*	*p*	*diff.*	*t*	*p*	*diff.*	*t*	*p*
useful	0.97	2.82	0.0441*	0.40	1.166	0.7704	0.97	2.82	0.0441*	0.17	0.49	0.9885
effective	1.13	3.51	0.0057**	0.10	0.309	0.9980	1.10	3.40	0.0080**	0.07	0.21	0.9996
efficient	1.30	3.54	0.0051**	0.64	1.726	0.4224	1.24	3.36	0.0091**	0.74	2.00	0.2734
fun	1.97	5.95	<.0001***	1.77	5.345	<.0001***	1.44	4.34	0.0003***	1.04	3.13	0.0187**
predictable	-2.20	-6.16	<.0001***	-2.57	-7.184	<.0001***	-1.50	-4.20	0.0005***	-2.34	-6.53	<.0001***
comprehensible	-0.53	-1.90	0.3252	-1.70	-6.044	<.0001***	-0.23	-0.83	0.9209	-1.23	-4.39	0.0002***
controllable	-1.94	-5.49	<.0001***	-3.07	-8.713	<.0001***	-2.47	-7.01	<.0001***	-3.17	-9.00	<.0001***

Table 3. Results of Tukey's HSD test. In *fun, useful, effective, efficient*, all treatment conditions produced higher scores than the control condition ((a), (b), (c), (d) > (e)). On the contrary, in *predictable, comprehensible*, and *controllable*, all treatment conditions produced lower scores than the control condition.

comprehensible, and *controllable* items than the control condition. In the case of *predictable*, all four treatment conditions recorded significantly lower scores than the control condition (Comparisons 4–7 in Table 3). For *controllable*, all four treatment conditions recorded significantly lower scores than the control condition (Comparisons 4–7 in Table 3). In the case of *comprehensible*, when the communication mode was *Basic*, the treatment conditions showed a significant difference (Comparisons 5, 7 in Table 3).

Meanwhile, *Detailed Instruction* could be a way to overcome these shortcomings of the AI interface. Although they received lower scores than the control condition, the *Detailed Instruction* conditions received higher ratings than the *Basic Instruction* conditions for all three items: *predictable, comprehensible*, and *controllable*. In the case of *comprehensible*, every *Detailed Instruction* condition recorded significantly higher scores than the *Basic Instruction* conditions (Comparisons 1, 2 in Table 2). In the case of *controllable*, in the *Lead* conditions, the *Detailed Instruction* conditions received significantly higher scores than the *Basic* conditions (Comparison 1 in Table 2). We could identify the same tendency in all other cases, even if this was not to a significant degree.

Even if Predictability is Low, Fun and Interest Can Increase

Through further analysis, we investigated the correlation between the *predictable* scores and the *fun* scores, which showed the opposite trend. The result revealed that there was a significant negative correlation between *predictable* and *fun* (*correlation coefficient: -0.847, p=.0010***). This means that although the AI interface has the disadvantage of low predictability, at the same time, it can provide users with a more fun and interesting experience [19].

Result 2: Qualitative Analysis

In the qualitative analysis, we aimed to investigate the users' thoughts in more depth and derive hidden characteristics behind the survey results. Specifically, we sought to identify users' perceptions of initiative and communication methods, the features they showed, and the factors they valued in interacting with the AI. We identified that users wanted the AI to provide detailed instructions but only when they wanted it to do so. In addition, they wanted to make every decision during the tasks. They sometimes anthropomorphized the AI and demonstrated a clear distinction between human and non-human characteristics. Finally, they reported that drawing with AI was a positive experience that they had never had before.

Just Enough Instruction

Overall, the participants wanted the AI to provide enough instruction during the tasks. However, at the same time, they did not want the AI to give too many instructions.

As seen in the survey results, we also identified that participants preferred *Detailed Instruction* to *Basic Instruction* in the qualitative analysis. Participants said *Detailed* provided a better understanding of the system and made them feel they were communicating and interacting with another intelligent agent. For example, P28 said, "I like the fact that it tells me what to do next." P27 also said, "It's a lot better. This guide makes me feel like I'm doing it right." Interacting with the

AI also increased the users' confidence. P24 said, "I liked the *Detailed* mode. I think it improved my confidence. I felt like I was communicating with someone." P02 said, "I like the way it talks to me. It confirms that I am doing a good job. It's like I'm being praised." In contrast, users expressed negative feelings about the *Basic Instruction*. They thought that in the *Basic* mode, it was hard to understand the system's intended meaning. Besides, they worried that they would miss the guidance, as it would pass quickly without their noticing. P14 said, "There is no explanation. It's not clear what I have to do. Does this mean that I have to draw something here? What should I draw?" P10 said, "It was my first task, so I didn't know what to do. I did not see the guide once, as it disappeared too quickly."

However, we also observed that some participants preferred *Basic Instruction*. They thought that in the long term, the *Basic* mode might have an advantage if users become more accustomed to the interface. They believed that straightforward and clear instructions would ultimately be more efficient. P08 said, "I think it [*Basic*] would be nice if I get used to the communication with the icon." P27 said, "If I become accustomed to it, I think I will pass on the *Detailed Instruction*. *Basic* could be more helpful."

Meanwhile, even the *Detailed* mode did not always guarantee a good experience. If the words of the AI seemed to be empty or automatic, users felt frustrated. When the system showed the message "It's a nice picture" as a reaction to a drawing, P27 said, "I think that it is an empty word; I mean, it just popped up automatically." P22 also talked about a similar experience; when he finished drawing an object, he was not satisfied with his drawing. However, immediately after he recognized that feeling, the AI praised his drawing, which made him feel disappointed. He said, "Do you really think it is nice? I want the AI to give me sincere feedback considering how I feel about my drawings. I felt like he was teasing me because I was not satisfied with my picture."

Participants wanted detailed communication rather than pre-set phrases. P15 commented, "When I drew this, I was thinking about a building like the UN headquarters in NYC. I wanted the system to be aware of my thoughts and give me more detailed feedback." They thought it would be better if the AI mentioned the details of the picture based on the drawn object rather than automatically showing a list of pre-set words. Besides, P05 said he did not want to get simple comments from the AI. Rather, he wanted to be able to actively share opinions with the AI about the drawn pictures. He said, "I want it to pick on my drawing, like 'Do you really think it is right here?' I want a more interactive chat like I have with my friends or my girlfriend."

Users Always Want to Lead

One of the most important characteristics that participants showed during the experiments was their strong desire to take the initiative, although *Lead* was not significantly preferred in the survey. Users' ability to make the decision at every moment seemed more important than being in the *Lead* mode itself. Most of the participants "always" wanted to take the initiative. Even in the *Assist* mode where the AI leads and the

user assists its drawing, they tried to take the initiative. P16 mentioned, "Of course, I know that I should help the AI in the *Assist* mode, but I couldn't be absorbed in that mode at all. Why should I support a computer? I cannot understand." P06 said, "Well, I think it's a very uncommon situation." P07 also said, "I did my best to do my role in the *Assist* mode, but it did not seem to be helpful. So I didn't know why I should help it."

Participants wanted to distinguish their roles from those of the AI. They thought that humans should be in charge of making decisions and that the AI should take on the follow-up work created by these decisions. In particular, they often expressed that AI should do the troublesome and tiresome tasks for humans. Some thought that repetitive tasks, such as colorizing, were arduous for them and did not want to perform them at all. P26 said, "It's very annoying. Why doesn't the AI just do this part?" P22 argued that people and AI should play different roles according to the nature of the work. He commented, "I feel a little annoyed with coloring the whole canvas. It is very hard. I wish you [AI] would do this colorizing. We humans don't have to do this. Humans have to make the big picture, and the AI has to do the chores." P21 also argued that people should have the right to make decisions in creative work. He said, "It's like I'm doing a chore [colorizing]. I like to make the decisions, especially when I do something artistic like this. It's fun to see what [the AI] is doing, but I don't want to do this myself."

Participants felt as if they were being forced when the AI made unilateral decisions. P07 said, "What are you doing? This is not co-creation. It seems like one person is letting the other person do it. I don't feel like we're drawing pictures together at all." P05 also felt as if he had become a passive tool for AI, saying, "I think he is using me as a tool." Some participants even said that this forced experience strengthened their negative feelings toward AI. Some of them stated that they felt frustrated and discouraged. P23 said, "Do I have to color myself? This is so embarrassing." P01 also said, "Anyway, I colored this vacuum cleaner and this sofa with the colors that the AI requested. Actually, it was not pleasant. I felt as if I was being commanded."

When asked to fulfill the AI's requests, some of the participants wanted to know why the AI had made those decisions. When asked to complete the colorizing with the colors that the AI had specified, P12 said, "So I wonder why he recommended these colors." Furthermore, participants wanted to negotiate with the AI so that their thoughts could be reflected in the drawing or to have more options from which they can choose. P19 said, "Usually, if I do not agree with someone's idea, I try negotiating. But it does not seem to be a negotiation. If I could negotiate, I would feel more like drawing artwork with the AI." P16 also said, "I think it would be better if I had more options or more room to get involved." P15 commented, "I don't like the position of these birds. I want to move them a little. I want to give him a lot more feedback."

Furthermore, some participants even wanted to deny the AI's requests. They tried to ignore the AI's requests and change the picture in a way that they thought was more appropriate. P09 said, "Why is the cleaner red? It is weird. I wanna

change it to a different color. I don't like the color of the sofa either." P06 also said, "I don't think I should follow its request." Meanwhile, P29 said, "These colors are a little bit dull. I'm gonna put on different colors."

AI is Similar to Humans But Unpredictable
During the task, we observed that participants tended to anthropomorphize the AI. People personified it as a human based on its detailed features [40]. They considered it an agent with a real personality. Furthermore, they did not regard the AI as being equal to human beings; rather, they regarded it as a subordinate to people. P13 mockingly said, "But I do not know what my robot master wants. Hey robot master, what do you want?" P14 also regarded the AI as someone with a personality. When the AI made a mistake, he said, "Oh poor thing, I forgive you for your mistake." P22 argued that the AI should be polite to humans. She said, "I don't like this request. He just showed me the message and told me to draw it. It's insulting. He should be polite, of course." P01 said, "I am trying to teach him something new, because he is not that fun yet. I heard that AI should learn from humans." This implies that she believed that AI is imperfect and must go through the process of learning through human beings.

Participants also found human-like features and non-human-like features of the AI. People felt the AI was like a human being when it drew objects imitating their drawing style, drew pictures in a natural way, or showed the process of its drawing. P18 said, "I felt as though it was a real human when it drew in a similar manner to how I draw." P23 said, "Well, this is not a well-drawn picture, but it makes me think it's drawn by a person. It seems to be drawn in a very natural way." On the other hand, participants felt the AI did not seem human when it drew objects too precisely and delicately, did not show its drawing process, and drew objects more quickly than expected. P30 said, "This is too sophisticated and too round. It's like a real coconut. It's too computer-like." P18 said, "I know it's not a human. It draws too quickly."

The problem was that the users felt uncomfortable when the AI went between being human-like and non-human-like. P11 told us that he felt it was awkward when it drew a clip art picture that was like a sophisticated and perfect object right after drawing a picture that was like a hand drawing. He said, "This nose is a bit different. It's like a clip art picture in a Google image search, and it makes this entire picture weird. Some of these pictures look hand-drawn, and some are elaborately drawn, as if made by a computer. It seems unbalanced." P22 also argued that pictures that had a mix of low- and high-quality parts seemed dissonant. He said, "It is a mixture of an excellent picture and a very poor picture. It's like someone wearing a cheap t-shirt but at the same time wearing luxury shoes."

Besides, users said they felt unhappy when the AI drew pictures that were much better than their pictures. They sometimes compared their drawings with those of the AI, which hurt their confidence. P20, comparing the part the AI had drawn to the part he had drawn, said, "If he had drawn it alone, it would have been better." He added that his role seemed to be meaningless. P18 also said, "It could be a perfect palm tree

if he took out the part I drew." P24 even told us that she felt like she was being ridiculed. She said, "Of course I like it. But AI seems to be teasing me."

Co-Creation with AI
Despite some of the inconvenience and the awkwardness of DuetDraw, most of the participants described drawing with AI as a pleasant and fun experience. This was also confirmed by the survey results, and we examined the elements in more detail in the qualitative analysis. P11 commented, "I think this program is fun and enjoyable. It is definitely different from conventional drawing." P01 said, "It was a bit of a new drawing experience. I was satisfied with it even though my drawing was not that good." Participants also stated that the AI allowed them to complete drawings quickly and efficiently. They said that the AI led them to the next step and helped with much of the picture. P29 said, "When I paused, the AI guided me to the next step quickly." P07 said, "It is fast. The AI does a lot of work for me."

Users also positively assessed each function of AI. In particular, they were very satisfied with its ability to colorize sketches semi-automatically. Almost all the participants were impressed with the artistic work of the AI. P13 said, "Now he is going to colorize it like a *décalcomanie*. Please surprise me! (pause) Oh! Wow cool! It is terrific. This is a masterpiece!" P17 said, "Oh my god, I love this. It looks like an abstract painting. I am so satisfied." The participants also evaluated that the drawing function for the rest of the object was both wonderful and interesting. P25 commented that when the AI drew every element of the object that she was about to draw, she was delighted. She said, "That was incredible. Well...I am so surprised that he can recognize what I was drawing and what I was gonna draw. He completed my strawberry. He drew all the elements of a strawberry." In addition, after seeing the AI draw the rest of his object, P17 expressed his greater expectations regarding the AI's abilities. He said, "It's wonderful. This makes me look forward to seeing his next drawings. What will he do next?" Some participants were satisfied with its ability to recommend a matching object. As described above, when recommending the object, the AI presented a clip-art-style object. Although some of the participants disliked it, as it was more like a computer than a human, other participants enjoyed the feature. They said that the clip art helped to increase the overall quality of their picture. P08 said, "He painted the plate very well. It is beautiful. I like beautiful things. They're certainly better than ugly things. I think this pretty dish is much better than my strange strawberry." The participants were also pleased with its ability to find an empty space on the canvas. Although finding the blank space itself was not that impressive, they believed that this feature allowed them to think about what was needed in their paintings. P28 said, "It was terrific, as it let me think about what kind of object I could draw. I know it is not that useful. But it seemed to stimulate my imagination a little more." This shows that AI can help to foster human creativity in collaboration.

Meanwhile, participants were highly satisfied with the AI when confronted with unexpected results. Users were amazed and pleased when the AI suddenly painted objects they wanted but did not expect AI to draw. They were also delighted when the AI drew a picture that differed from what they had expected. P30 said, "When I let him know about this empty space, I vaguely thought that a plane or birds flying around the sky would fit here. Of course, I didn't expect that the AI would understand my thoughts. But the AI drew birds! I was thrilled." P21 also described his similar experience. He said, "I think art sometimes needs uncertainty. Some painters just scatter paint on the canvas without any purpose. I thought the AI was like this. I just picked the color, and the AI painted it. The result was totally different from what I had expected, and I was delighted." P17 said, "I think this is the best part of this experiment. The AI has drawn pictures in a way I have never thought of before."

Some users said that the experience of drawing with the AI made them feel as if they were with someone. P29 said, "When I was drawing this picture, I felt like I was drawing with someone." P11 said that drawing with the AI made it possible to create a picture that would never have been created independently. He said, "If I had drawn alone, I would not have drawn this. Before I started this, I never knew I was going to paint this picture." P02 mentioned that drawing together even made him feel more stable. He also said, "I think drawing is like putting the thoughts in your head on paper. Usually, we do this alone, but it's hard. But in this experiment, I felt like someone was involved in this process. I felt like I was talking with an agent and sharing my thoughts with him."

Lastly, DuetDraw made users curious about the principles of its algorithms. During the tasks, the participants wanted to see how the AI algorithms worked underneath the interface and tried to test their guesses. During the task, P15 said, "How do you know this is a tree? You are so amazing. What made you think it was a tree?" P17 was more curious about the AI algorithms and created and tested hypotheses. He said, "I was curious about the principle of this colorizing. So I deliberately picked a variety of colors inside this contour, not just one color. If he recognized the object as a whole then the coloring would not seem out of line." P14 also said, "Well, now I see. The AI seems to divide the area and color each sector differently." P08 also said, "This is so smart. He mixed the colors and made a gradient. Hmm...I'm still curious about the criteria he used to paint each area differently."

DISCUSSION
In this section, we discuss the findings of the study and its implications for user interfaces in which users and AI collaborate. We also report our plans for future work as well as the limitations of the study.

Let the User Take the Initiative
As we have seen in the qualitative research, users wanted to take the initiative in collaborating with the AI. To enhance user experience in this context, it would be better to let users make most of the decisions. Even if a user receives an order or request from the AI, it might be better to provide him or her with options or ask permission for the request. In addition, if a user and AI have to do their tasks separately, repetitive and arduous tasks should be assigned to the AI and creative and major tasks should be assigned to the user.

Meanwhile, it should be noted that the feeling of taking the initiative is not always guaranteed just because the user is in the leader role. This was also revealed in the survey results, in that there was no significant difference between the effect of the *Lead* and *Assist* modes on users' evaluations of each item. Regardless of whether a user takes the role of the leader or the assistant, he or she always wants to take the initiative in the collaboration process. Rather than simply naming the user the leader, it would be more appropriate to give him or her the initiative at every decisive moment through close communication.

Provide Just Enough Instruction
As we have seen in both the survey and the qualitative research, users prefer AI to provide detailed instructions in their collaboration with AI but only in the way they want. In this context, cordial and detailed communication should be considered in AI and user collaboration first. As the survey results revealed, offering users detailed explanations could be an effective way to enhance the overall user experience of user–AI collaboration. Furthermore, it can improve users' perceived predictability, comprehensibility, and controllability of the drawing tasks, all of which have been pointed out as shortcomings of AI interfaces in previous studies [18, 23]. *Detailed Instruction* can also make users understand the tasks more easily, feel as if they are with somebody, and feel confident.

However, it should be pointed out that the AI should only provide a description when the user wants it. Excessive or inappropriate descriptions can have an adverse impact on the user experience. These may make the user feel disturbed or disconnected from the tasks and even disappointed and frustrated. Rather than giving users automated utterances like template sentences or preset words, the AI should kindly and specifically comment on the actual behavior of the user or the result of the task.

Embed Interesting Elements in the Interaction
This is an important and challenging point. As we saw in the user study, people were pleased with the interaction with the AI, and they felt various positive emotions. Users were especially amused when the AI drew unexpected objects. In this respect, placing serendipitous elements in the middle of the interactions could be considered as a means of enhancing the user experience and the interface's usability. This could be a way of providing an interesting and pleasant user experience [33] during the task.

At the same time, each function of AI should be designed to foster user's curiosity and imagination for creative works. Traditionally, creativity support tool studies have revealed many principles for motivating users' creative actions, such as presenting space, presenting various paths, lowering thresholds, and so forth [47]. We believe that these principles could be still more significant elements in providing a good experience when users collaborate with AI, thus enhancing users' potential and unleashing their creative aspirations.

Ensure Balance
The last point centers on the imbalance that users felt in collaborating with the AI. From the qualitative study, we observed that the participants felt confused when the ability of the AI differed across functions. They found it strange when there was a mixture of high- and low-quality objects on the canvas. They felt frustrated when the AI showed human-like characteristics and machine-like characteristics in the same task and when it showed superior ability compared to them. Since the users tended to regard AI as an agent and sometimes personified it, their expectations of the interface might have been higher and more complex than those of other simple interfaces. For this reason, when it showed unbalanced and awkward qualities, they felt disappointed, leading to anthropomorphic dissonance [52, 54]. As the AI platform will likely introduce various technologies or open sources and face a broad variety of users, balancing the multiple elements and providing a harmonious experience for users could be a key point in AI platform design.

Limitations and Future Work
We have identified three limitations of this study. First, although DuetDraw was designed for user–AI collaboration based on neural network algorithms, it cannot represent all AI interfaces. Second, in the experiments, we had to control the participants' behaviors with a task-oriented scenario, and users were not able to use the interface freely. Third, we could not address the long-term experience of user–AI interaction, and the study results may have been influenced by users' initial impressions of the interface.

In future work, we will investigate user experience in a wider variety of extended interfaces beyond the framework of drawing tools. We also plan to improve DuetDraw so that users can use it more flexibly and explore the long-term experience of cooperation between AI and users.

CONCLUSION
This study examined the user experience of a user–AI collaboration interface for creative work, especially focusing on its communication and initiative issues. We designed a prototype, DuetDraw, in which AI and users can draw pictures cooperatively, and conducted a user study using both quantitative and qualitative approaches. The results of the study revealed that during collaboration, users (1) are more content when AI provides detailed explanations but only when they want it to do so, (2) want to take the initiative at every moment of the process, and (3) have a fun and new user experience through interaction with AI. Finally, based on these findings, we suggested design implications for user–AI collaboration interfaces for creative work. We hope that this work will serve as a step toward a richer and more inclusive understanding of interfaces in which users and AI collaborate in creative works.

ACKNOWLEDGMENTS

REFERENCES

1. William Albert and Thomas Tullis. 2013. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes.

2. Nick Babich. 2016. 5 Essential UX Rules for Dialog Design. (2016). Retrieved September 18, 2017 from `http://babich.biz/5-essential-ux-rules-for-dialog-design/`.

3. Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, and others. 2016. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316* (2016).

4. Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew M Dai, Rafal Jozefowicz, and Samy Bengio. 2015. Generating sentences from a continuous space. *arXiv preprint arXiv:1511.06349* (2015).

5. Alex J Champandard. 2016. Semantic style transfer and turning two-bit doodles into fine artworks. *arXiv preprint arXiv:1603.01768* (2016).

6. Chenyi Chen, Ari Seff, Alain Kornhauser, and Jianxiong Xiao. 2015. Deepdriving: Learning affordance for direct perception in autonomous driving. In *Proceedings of the IEEE International Conference on Computer Vision*. 2722–2730.

7. Keunwoo Choi, George Fazekas, and Mark Sandler. 2016. Text-based LSTM networks for automatic music composition. *arXiv preprint arXiv:1604.05358* (2016).

8. John Collier. 1957. Photography in anthropology: a report on two experiments. *American anthropologist* 59, 5 (1957), 843–859.

9. Nicholas Davis, Chih-Pin Hsiao, Kunwar Yashraj Singh, and Brian Magerko. 2016a. Co-creative drawing agent with object recognition. In *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*.

10. Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016b. Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 196–207.

11. Umer Farooq, Jonathan Grudin, Ben Shneiderman, Pattie Maes, and Xiangshi Ren. 2017. Human Computer Integration versus Powerful Tools. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1277–1282.

12. Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015).

13. Leon A Gatys, Alexander S Ecker, Matthias Bethge, Aaron Hertzmann, and Eli Shechtman. 2016. Controlling perceptual factors in neural style transfer. *arXiv preprint arXiv:1611.07865* (2016).

14. Barney Glaser. 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

15. Alex Graves. 2013. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850* (2013).

16. Hayit Greenspan, Bram van Ginneken, and Ronald M Summers. 2016. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging* 35, 5 (2016), 1153–1159.

17. David Ha and Douglas Eck. 2017. A Neural Representation of Sketch Drawings. *arXiv preprint arXiv:1704.03477* (2017).

18. Melanie Hartmann. 2009. Challenges in Developing User-Adaptive Intelligent User Interfaces.. In *LWA*. Citeseer, ABIS–6.

19. Rex Hartson and Pardha S Pyla. 2012. *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.

20. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.

21. Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 159–166.

22. Allen Huang and Raymond Wu. 2016. Deep learning for music. *arXiv preprint arXiv:1606.04930* (2016).

23. Anthony David Jameson. 2009. Understanding and dealing with usability side effects of intelligent processing. *AI Magazine* 30, 4 (2009), 23.

24. Jonas Jongejan, Henry Rowley, Takashi Kawashima, Jongmin Kim, and Nick Fox-Gieg. 2017. Quick, Draw! (2017). Retrieved September 18, 2017 from `https://quickdraw.withgoogle.com`.

25. Jerry Kaplan. 2016. Artificial intelligence: think again. *Commun. ACM* 60, 1 (2016), 36–38.

26. Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1301–1318.

27. Will Knight. 2016. The Dark Secret at the Heart of AI. MIT Technology Review. (2016). Retrieved September 18, 2017 from `https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/`.

28. Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504.

29. Google Creative Lab. 2017. AutoDraw. (2017). Retrieved September 18, 2017 from `https://experiments.withgoogle.com/chrome/autodraw`.

30. Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. 2015. Human-level concept learning through probabilistic program induction. *Science* 350, 6266 (2015), 1332–1338.

31. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

32. Honglak Lee, Peter Pham, Yan Largman, and Andrew Y Ng. 2009. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in neural information processing systems*. 1096–1104.

33. Tuck Leong, Steve Howard, and Frank Vetere. 2008. Choice: abidcating or exercising?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 715–724.

34. Clayton Lewis and John Rieman. 1993. Task-centered user interface design. *A Practical Introduction* (1993).

35. Arnold M Lund. 2001. Measuring Usability with the USE Questionnaire12. *Usability interface* 8, 2 (2001), 3–6.

36. Yotam Mann. 2017. AI Duet. (2017). Retrieved September 18, 2017 from `https://experiments.withgoogle.com/ai/ai-duet`.

37. Lauren McCarthy. 2017. p5.js. (2017). Retrieved September 18, 2017 from `https://github.com/processing/p5.js?files=1`.

38. Ian Millington and John Funge. 2016. *Artificial intelligence for games*. CRC Press.

39. Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, and others. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529–533.

40. Clifford Nass, Jonathan Steuer, Ellen Tauber, and Heidi Reeder. 1993. Anthropomorphism, agency, and ethopoeia: computers as social actors. In *INTERACT'93 and CHI'93 conference companion on Human factors in computing systems*. ACM, 111–112.

41. Changhoon Oh, Taeyoung Lee, Yoojung Kim, SoHyun Park, Bongwon Suh, and others. 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*. ACM, 2523–2534.

42. Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. Alexa is my new BFF: Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2853–2859.

43. Mengye Ren, Ryan Kiros, and Richard Zemel. 2015. Exploring models and data for image question answering. In *Advances in neural information processing systems*. 2953–2961.

44. Xiangshi Ren. 2016. Rethinking the Relationship between Humans and Computers. *IEEE Computer* 49, 8 (2016), 104–108.

45. Yvonne Rogers, Helen Sharp, and Jenny Preece. 2011. *Interaction design: beyond human-computer interaction*. John Wiley & Sons.

46. Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural networks* 61 (2015), 85–117.

47. Ben Shneiderman, Gerhard Fischer, Mary Czerwinski, Mitch Resnick, Brad Myers, Linda Candy, Ernest Edmonds, Mike Eisenberg, Elisa Giaccardi, Tom Hewett, and others. 2006. Creativity support tools: Report from a US National Science Foundation sponsored workshop. *International Journal of Human-Computer Interaction* 20, 2 (2006), 61–77.

48. Ben Shneiderman and Pattie Maes. 1997. Direct manipulation vs. interface agents. *interactions* 4, 6 (1997), 42–61.

49. David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, and others. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.

50. Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. 2013. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*. 1139–1147.

51. Ilya Sutskever, James Martens, and Geoffrey E Hinton. 2011. Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 1017–1024.

52. Luke Swartz. 2003. *Why people hate the paperclip: Labels, appearance, behavior, and social responses to user interface agents*. Ph.D. Dissertation. Stanford University Palo Alto, CA.

53. Gheorghe Tecuci, Mihai Boicu, and Michael T Cox. 2007. Seven aspects of mixed-initiative reasoning: An introduction to this special issue on mixed-initiative assistants. *AI Magazine* 28, 2 (2007), 11.

54. Stuart NK Watt. 1997. Artificial societies and psychological agents. In *Software Agents and Soft Computing Towards Enhancing Machine Intelligence*. Springer, 27–41.

55. Etienne Wenger. 2014. *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann.

56. Terry Winograd. 2006. Shifting viewpoints: Artificial intelligence and human–computer interaction. *Artificial Intelligence* 170, 18 (2006), 1256–1258.

57. Optimal Workshop. 2016. Reframer. (2016). Retrieved September 21, 2016 from `https://www.optimalworkshop.com/reframer`.

58. Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A New Chatbot for Customer Service on Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3506–3510.

59. Yan Xu, Tao Mo, Qiwei Feng, Peilin Zhong, Maode Lai, I Eric, and Chao Chang. 2014. Deep learning of feature representation with multiple instance learning for medical image analysis. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 1626–1630.

60. Taizan Yonetsuji. 2017. Paint Chaniner. (2017). Retrieved September 18, 2017 from `https://github.com/pfnet/PaintsChainer`.