공학박사 학위논문

인공지능을 이용한 알고리즘 기반의 시스템과 사용자의 인터랙션에 대한 이해

Understanding How People Interact with Algorithm-based Systems Using Artificial Intelligence

2019년 2월

서울대학교 융합과학기술대학원

융합과학부 디지털정보융합전공

오 창 훈

인공지능을 이용한 알고리즘 기반의 시스템과 사용자의 인터랙션에 대한 이해

Understanding How People Interact with Algorithm-based Systems Using Artificial Intelligence

지도교수 서 봉 원

이 논문을 공학박사 학위논문으로 제출함

2019 년 1 월

서울대학교 융합과학기술대학원

융합과학부 디지털정보융합전공

오 창 훈

오창훈의 박사 학위논문을 인준함

2019 년 1 월

위 원 장	이 중 식	(인)
부 위 원 장	서 봉 원	(인)
위 원	이 준 환	(인)
위 원	권 가 진	(인)
위 원	김 주 호	(인)

Abstract

The recent development of artificial intelligence (AI) algorithms is affecting our daily lives in numerous areas. Moreover, AI is expected to evolve rapidly, bringing tremendous economic value. However, compared to the attention these technological improvements receive, there is relatively little discussion on human factors and user experience related to AI algorithms. Thus, this thesis aims to better understand how users interact with AI algorithms. Specifically, this work examined algorithm-based human–AI interaction in four stages, through various modes of human-computer interaction: The first study investigated how people perceive algorithm-based systems using AI, finding that people tend to anthropomorphize as well as alienate them, which is distinct from their perceptions of computers. The second study investigated how people interpret and evaluate the output from AI algorithms through a prototype, AI Mirror, which assigned aesthetic scores to images based on a neural network algorithm. The results revealed that people interpret AI algorithms differently based on their backgrounds, and that they want to understand and communicate with AI systems. The third study investigated how people build a sequence of actions with AI algorithms through a mixed method study using a research prototype called DuetDraw, a drawing tool in which users and AI can draw pictures together. The results showed that people want to lead collaborations while hoping to get appropriate instructions from the AI algorithm. Lastly, a case study on a practical application of AI was conducted with a research prototype called NewsRobot, which automatically generated news articles with different content and styles. Findings showed that users prefer selective news and multimedia news that have more functionality and modality, but at the same time they do not want AI to boast about its ability. With these distinct but intertwined studies, this thesis argues the importance of understanding human factors in the user interfaces of AI-based systems and suggests design principles to this end.

Keywords: Artificial Intelligence, Human-AI Interaction, Algorithm-based Systems Using Artificial Intelligence, Human-Computer Interaction, Algorithmic Experience

Student number: 2012-22466

Contents

List of Figures

List of Tables

1 INTRODUCTION

1.1 Background

In March 2016, a Go game held at a conference room in a hotel in Seoul was broadcasted live across the Korean peninsula. It was a Go match between the human Go champion Lee Sedol and AlphaGo, an artificial intelligence (AI) Go program developed by Google DeepMind. Since Go has long been regarded as the most challenging classic game for AI, the match garnered significant attention from AI and Go communities worldwide. By winning four of the five games, AlphaGo became the final winner of the match. The result of the event was a complete surprise to many people, as it showed that AI had evolved to a remarkable level, even surpassing humanity in an area requiring advanced intelligence. It was impossible to deny the scene in which the software, which had no form, mercilessly conquered the champion by moving the Go-stone through the human agent Aja Huang. After the first game, DeepMind founder Demis Hassabis tweeted, “#AlphaGo WINS!!!! We landed it on the moon [1].”

AI is no longer a distant future in Hollywood movies. At this point, less than three years after that event shocked us, media programs frequently discuss the rosy future of AI technology. Tech giants show off their technological prowess and market dominance by announcing new devices and services based on AI. Millions of users interact with intelligent devices or services that use AI technology, such as Apple's Siri and Amazon's Echo. It is expected that AI technology will become increasingly prevalent in various areas, including autonomous vehicles ([2], [3]), medical treatment ([4], [5]), game playing ([6], [7]) and customized advertisements [8].

However, despite the clear evidence that AI technology will have a profound impact on our society, relatively little work has been done to holistically understand how users interact with AI. Research on human factors and user experience considerations of AI technology is particularly insufficient. Most AI-related studies are focused on developing or improving AI algorithms themselves. Although machine learning and AI communities have recently started discussing Explainable AI to expand their research fields to converge with neighboring areas [9], there has still been relatively little consideration of users. Likewise, in the field of human-computer interaction (HCI), although research on some topics related to AI has appeared, the level of discussion is still low.

Against this background, this thesis uses various HCI methodologies to examine how users interact with AI algorithms. Through the results of four separate but intertwined studies, this thesis argues that users want to have a more human-like interaction with AI than the superficial relationship they have with existing computers; they want to understand and communicate with AI. Furthermore, this paper derives a set of guidelines for the design of interfaces based on AI algorithms.

Brief History of Artificial Intelligence

Artificial Intelligence refers to intelligence demonstrated by machines, in contrast to the natural intelligence displayed by human beings.[1] In computer science, the term applies to any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals [10]. More recently, in the business context, the term is defined as "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation [11]." AI often refers to machines that mimic cognitive functions that humans associate with other human minds, such as learning and problem-solving [12].

The field of AI research was founded at a workshop at Dartmouth College in 1956 attended by Herbert Simon and Allen Newell of Carnegie Mellon University, John McCarthy and Marvin Minsky of MIT, and Arthur Samuel of IBM Research [13]. The conference established not just the term AI, but also the goals and scope. Since that event, AI has received a great deal of attention. By the mid-1960s, the United States Department of Defense heavily funded AI-related research, and laboratories researching AI were established around the world [13]. This is called the "inference period" of AI history, and many studies focused on making computer systems capable of logical reasoning [14]. AI researchers were optimistic about the future of AI. Herbert Simon predicted that machines would be capable of doing any work a human could do within twenty years [15]. Marvin Minsky also predicted that within a generation, the problem of creating AI would substantially be solved [15]. In this period, early natural language processing programs that demonstrated the superficiality of communication between humans and machines, such as Eliza, were created [16].

However, despite their optimistic expectations, the first winter came to AI research ([12], [15]). Computing power was not fully developed, so complex

[1] From Wikipedia (https://en.wikipedia.org/wiki/Artificial_intelligence)

problems could not be addressed. Some argued that since AI could not solve real-world problems, researching AI was essentially making toys. The US government started to cut off exploratory research in AI, and most AI research laboratories had difficulty obtaining funding [17]. Surprisingly, HCI began to take its place as an academic discipline at this time.

In the early 1980s, after the first harsh winter of AI, spring began to arrive. Knowledge engineering became the new keyword in the field, beginning the "knowledge period [14]." Due to the commercial success of expert systems ([12], [13], [15], [18]), a form of AI program simulating the knowledge and analytical skills of human experts, AI was recognized as having practical value in specific areas. For example, John McDermott of CMU developed the R1 (also called XCON), a production-rule-based system, to assist in the ordering of VAX computer systems by automatically selecting components based on the customers' requirements [19]. Moreover, the introduction of multilayer perceptron and back-propagation solved some persistent problems in AI [20] and rejuvenated the field.

The field of AI research, however, reached winter again, later called the second winter of AI ([13], [15], [17]). At this time the cost of maintaining computing power was still high. AI research faced criticism that "expert systems" were too focused on particular topics and had no general value. At the same time, many universities in the US adopted HCI as a division of their computer science department [21]. Studies in computer engineering became interested in the practical use of computing systems and focused on user-related discussions, including the question of usability.

In the early 1990s, AI was revived with the introduction of computers with outstanding performance, called the "learning period [14]," the third period of AI research. There was a remarkable advance in computer power as described by

Moore's Law. Research topics in AI became more cautious, focusing on practical use. Additionally, mathematical and scientific advances were made ([12], [13]), including the introduction of machine learning algorithms with superior performance such as support vector machines (SVM) [22]. The spread of the Internet produced enormous amounts of data that became an essential basis for AI research.

Since the late 2000s, AI has reached its peak and expects a rosy future. This is largely due to the development of deep learning, a class of machine learning algorithms that use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation [23]. Deep learning algorithms can learn in a supervised or unsupervised way and can learn multiple levels of abstractions, forming a hierarchy of concepts. Deep learning architectures such as deep neural networks, deep belief networks [24], and recurrent neural networks (RNN) have been applied to various fields including speech recognition ([25], [26]), computer vision [27], natural language processing ([28], [29]), social network filtering [30], audio recognition [31], machine translation ([32], [33]), bioinformatics [34], medical image analysis ([35], [36]), and board game programs ([37], [38]). In these fields, they have produced results comparable to and superior to human experts. In recent years, generation-related algorithms such as variational autoencoders (VAEs) [39] and generative adversarial networks (GANs) [40] have gained attention and are being continually developed.

The recent advances in AI have also achieved remarkable results when utilized in applications. IBM's DeepQA developed Watson, a computer system capable of answering questions in natural language [41]. Watson competed on Jeopardy against legendary champions Ken Jennings and Brad Rutter. IBM then announced that Watson would be used for utilization management decisions in lung cancer treatment. Voice assistant services such as Apple's Siri and Samsung's Bixby have also been launched and are becoming commonplace. Smart speakers like Amazon Echo and Google Home also provide services based on AI technology. It is expected that AI technology will become increasingly prevalent in various areas ([2]–[8]). Accordingly, AI will affect the lives of more and more people.

Why is Human-AI Interaction important?

As artificial intelligence technology develops expertise to a level comparable to or beyond that of humans ([27], [42]), and various AI applications and services are emerging, its users are gradually becoming an essential factor in its success. As a result, discussions about human factors and user experiences of AI technologies and services are getting more attention. Despite the interest, however, comparatively little research has been done on this topic. Most studies on AI focus on technology-driven development, devoted to improving the performance of existing algorithms or developing new algorithms rather than exploring real-world problems that AI can solve. AI guru Ali Rahimi criticized AI technology as overfitting in technology at a conference keynote in 2017. The HCI field, which seeks to understand the various interactions between users and computing devices, is now resolutely focusing on the AI topic. However, many studies are still trying to fit AI into existing research topics, without a deep understanding of the developing AI technology.

Under these circumstances, exploring human-AI interaction as a separate topic can be beneficial. HCI provides a range of methods for studying the interactions between users and technology. Researchers in HCI design novel computer interfaces, optimize for desired properties, and evaluate factors such as usability. Moreover, they study the broader socio-cultural implications of technology use, and reflect upon the values that underlie computational design critically [43]. These viewpoints can complement the hitherto limited consideration of human factors in the AI community and help prevent technology-based overfitting.

Understanding human factors in AI can also improve the algorithms themselves. Humans influence various parts of the AI algorithm design process [44]: The raw data for learning that is the basis of an AI algorithm is, of course, information about humans, often collected by humans. The data is designed by humans, to solve problems framed by humans. Humans are deeply involved in the refinement of data, and humans choose its attributes. The math formulas are determined by humans. The model is evaluated and revised by humans. The predictive values of the model are given to humans, and humans decide how to interpret it appropriately, and where and how to use it. Predictions are often presented to humans through user interfaces designed by humans. In this way, it can be seen that many humans are involved in the algorithm design process, including both researchers and the general public. A deeper understanding of human factors would help researchers account for this human influence, while also enabling them to design algorithms that are more intuitive and useful to humans.

Lastly, by doing human-AI interaction research, HCI and AI could be synergistic [45]. There has been an unjustified perception in HCI that AI is unreliable, and in AI that interfaces are merely cosmetic, which is a counterproductive disagreement. However, AI's goal of intelligent interfaces would benefit enormously from the user-centered design principles of HCI, which would enable the design of more natural interactions with AI [45]. Likewise, applying AI technology to user interface design would provide users with more intelligent software [45].

Algorithm-based systems using AI

As discussed above, it is no exaggeration to say that AI has evolved in two directions: one is technology-focused (algorithm design and its performance evaluation), and the other is market-focused (services and products using AI technology). On the technology side, AI research has only considered creating the best performing model for a given set of data. Researchers in technology development areas such as machine learning, computer vision, and speech synthesis have relatively little concern of when and how their results will be used. On the market side, services and products related to AI give the impression that they are made to preempt the technology, rather than providing the user with a complete and optimized experience. Although both areas of development are important in human-AI interaction, the former has the shortcoming of ignoring users and practical value, while the latter is individual product-focused, lacking holistic understanding.

For this reason, this study narrows the scope of its exploration of Human-AI Interaction to algorithm-based systems using artificial intelligence (Figure 1-1). It includes algorithms that have advanced the development of AI in recent years and provides an opportunity to look more closely at how users interact with those algorithms. Through algorithm-based systems using AI can, the study is able to focus on users' interactions with actual algorithms, rather than merely evaluating the user experience of off-the-shelf products or services. In this way, this study strives to inform AI research of the importance of human factors and user experience, and at the same time find a contact point between the AI and HCI fields.

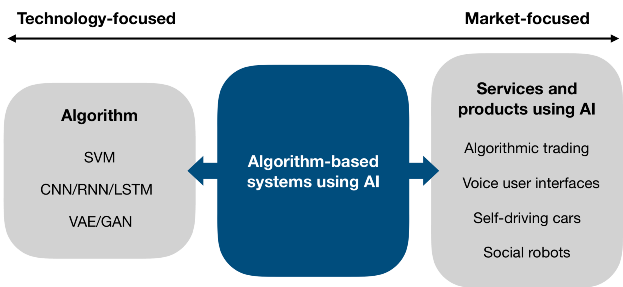


Figure 1-1. Algorithm-based system using AI

What is interaction?

Before examining human-AI interactions for algorithm-based systems using AI, this dissertation examines the concept of interaction and chooses an appropriate interaction model as a framework. To this end, this study refers to Kasper Hornbaek's work on the definition of interaction [46]. In a recent essay arguing that few attempts to directly define interaction have been made, he conducted a thorough literature review of the term and categorized the definitions into seven concepts: interaction as (1) dialogue, (2) transmission, (3) tool use, (4) optimal behavior, (5) embodiment, (6) experience, and (7) control [46].

Among these categories, this thesis focuses on the dialogue concept. This concept sees interaction as a cycle of communication acts between machine and human [47]. The communication cycle is a deep and ongoing, consisting of several stages. Thinking of interaction as a dialogue allows understanding from a longer-term perspective. Since this thesis seeks a deeper and more complete understanding of the interaction between humans and AI algorithms, the dialogue concept is an appropriate framework to apply.

One of the representative models of interaction as dialogue is Donald Norman's gulf of execution model ([48]–[50]). This model divides the steps in which the user interacts with the physical system into three areas: perception, interpretation and evaluation, and sequence of actions. Perception refers to users observing the state of the system, whereas the interpretation and evaluation phase covers users making sense of that state and assessing the outcome. Sequence of actions refers to the user forming an intention, specifying a sequence of actions, and executing the action sequence.

These stages are all significant topics in relation to AI algorithms. With regard to perception, there is the question of what a priori knowledge and ethics people have about AI and technology. With regard to interpretation and evaluation, debates are frequently raised about bias in AI algorithms, Explainable AI (XAI), and other topics related to user trust in AI. With regard to sequence of actions, there are discussions about how the user manipulates the interface and what factors should be considered in the user and AI partnership.

Applying the dialogue view of human-AI interaction to examine various aspects of AI algorithms, this dissertation aims to understand how to design an algorithm-based AI interface that better meets users' needs (Figure 1-2).

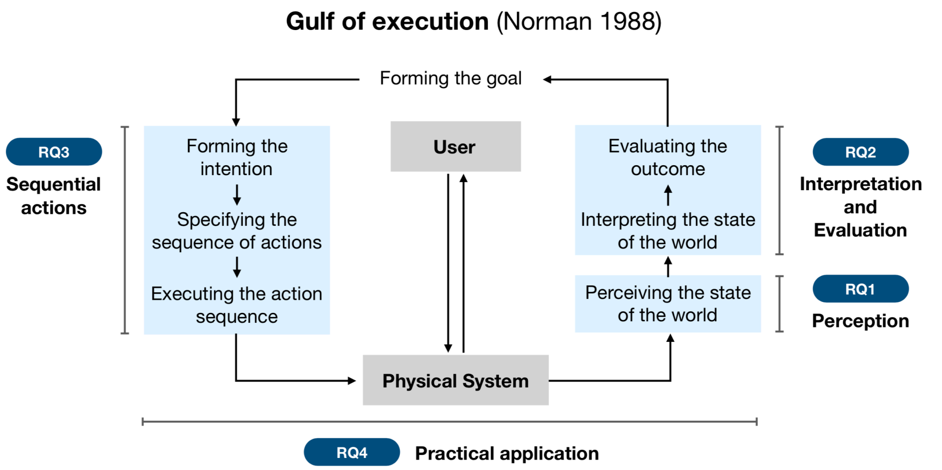


Figure 1-2. Norman's interaction model. Norman's gulf of execution model divides the steps in which the user interacts with the physical system into three areas: perception, interpretation and evaluation, and sequence of actions.

1.2 Research Goal

This thesis aims to examine the different stages of interaction between humans and algorithm-based AI through various HCI methods and derive design implications for AI-embedded user interfaces.

1.3 Research Questions

The main research question of this thesis is "How do people interact with algorithm-based systems using AI?" Under this question, it asks the following.

RQ1: How do people perceive algorithm-based systems using AI?

- What are the characteristics of human perception of AI algorithms?
- What factors affect people's perception of AI algorithms?

RQ2: How do people interpret and evaluate algorithm-based systems using AI?

- Can people interpret the principles of AI algorithms? If so, how do they try to interpret them? How does their interpretation affect user experience?
- How do people evaluate the output of AI algorithms? Do they think the results are reasonable? What factors affect these assessments?

RQ3: How do people build sequential actions with AI?

- Who do users prefer to take the initiative and which communication means should be provided to users in this process?
- What factors affect the user experience of the sequence of actions stage of interaction?

RQ4: How do people use a practical application of an algorithm-based system using AI?

- How does the user evaluate the addition of functionality and modality to algorithm-based systems using AI?
- What should be introduced into the AI interface design to provide functionality and modality that meets user expectations?

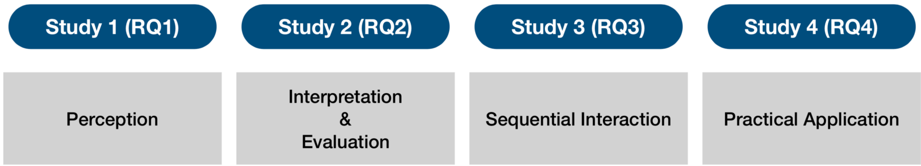


Figure 1-3. Thesis overview

The thesis consists of four sub-studies (Figure 1-3): (1) The first study investigates how people perceive AI algorithms; (2) the second study focuses on how people interpret AI algorithms; (3) the third study considers how people cooperate with AI algorithms; (4) the fourth study investigates the user experience of a news service generated by an AI algorithm as an application case study of how people interact with AI algorithms. Through these separate but intertwined studies, this thesis concludes that users want to have a more human-like interaction with AI than the superficial relationship they have with existing computers; they want to understand and communicate with AI. In addition, this thesis derives guidelines for interface design based on AI algorithms.

1.4 How People Perceive Algorithm-based Systems Using Artificial Intelligence

The first stage of investigation into human interaction with algorithm-based systems using AI sought to understand how people perceive these systems. This study uses the Google DeepMind Challenge Match as a case study. The result of the match shocked and amazed many people in Korea; it provoked public discussion and therefore provided a good opportunity to investigate the general public's opinions, responses, and concerns about AI. After recruiting 22 participants, semi-structured interviews about the match were carried out.

This study identified a dichotomous ("us vs. them") view of AI. Specifically: (1) People had preexisting stereotypes and prejudices about AI, mostly acquired from media sources such as Hollywood movies. Participants simultaneously believed that AI could cause harm to humans and that AI should assist and help humans. (2) People had ambivalent feelings about AI. They anthropomorphized but also alienated AlphaGo, evaluating it based on its perceived human characteristics while also focusing on features that differed from typical human traits. (3) People expressed concerns over a future society in which AI is widely used. They worried that AI would replace their jobs and they would not be able to control AI technology.

This work illustrates a confrontational relationship between users and AI and suggests the need to create a new kind of user experience in this nascent socio-technological space. It calls for a collaborative research effort from the HCI community to study and accommodate users in a future in which they interact with algorithms, not just interfaces.

1.5 How People Interpret and Evaluate Algorithm-based Systems Using Artificial Intelligence

For the second phase of understanding how people interact with algorithm-based systems using AI, this study investigated users' reasoning about and evaluation of an AI algorithm. A research probe called AI Mirror was designed, an application that algorithmically scored the aesthetic value of photographs users had taken or selected, based on a deep neural network model (Figure 1-4). A user study employing both quantitative and qualitative methods was conducted. A total of 18 participants were recruited, consisting of a balanced mix of AI/ML experts, photographers, and members of the general public. They performed a series of tasks in which they took photos using AI Mirror, reasoned about AI Mirror's algorithm through the think-aloud method, and answered a questionnaire to report their expected scores for the photos as well as rating the interpretability and reasonability of the AI's scores. Semi-structured interviews about how users experienced the system were also conducted. The results of the study can be summarized as follows.

(1) Users showed different characteristics in reasoning about the AI algorithm depending on their group (i.e. experts, photographers, or general public); they understood the AI using group-specific expertise. The photographers were able to best interpret the AI's aesthetic scores and considered them reasonable. On the other hand, the AI/ML experts had difficulty interpreting these scores and considered them relatively unreasonable. (2) Users adopted personal strategies to infer the AI's principles of evaluation over time. They used approaches that involved making subtle changes to various photo elements and testing their ideas through examples. If there was a difference between the users' thoughts and the AI's predictions, they then had difficulty interpreting the AI's predictions and considering them reasonable.

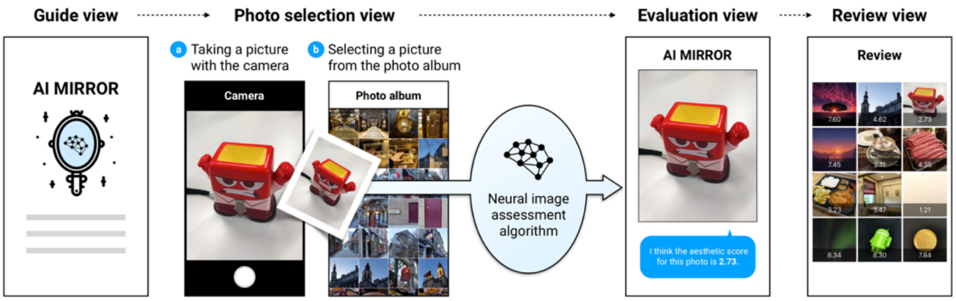


Figure 1-4. AI Mirror

Based on these findings, design recommendations for AI-powered user interfaces that convey subjective results to users were derived: (1) Integrate diverse expertise and user perspectives to ensure algorithm transparency and fairness. (2) Take advantage of people's curiosity about AI principles. (3) Help AI and users understand each other through mutual communication.

1.6 How People Build Sequential Actions with Algorithm-Based Systems Using Artificial Intelligence

The third phase of the research on algorithm-based human-AI interaction sought to understand how people cooperate with AI algorithms.

First, a prototype AI named DuetDraw was designed. The prototype allowed users to collaboratively draw pictures with an AI (Figure 1-5). Using state-of-the-art AI techniques, DuetDraw performed a variety of AI-based functions to help users draw. These drawing tasks include completing the rest of an object that a user started drawing, drawing the same object in a different style, suggesting an object that matched the picture the user had created, finding an empty space on the canvas, and automatically colorizing sketches.

To better understand users' experiences of user–AI collaboration, a user study of DuetDraw using a mixed method approach was designed and conducted. As the major factors, the effects of the type of communication (detailed/basic) and initiative (lead/assist) on the user experience were considered. By combining these two factors, four experimental conditions and one control condition (no AI) were designed. Thirty participants were recruited and asked to conduct a series of drawing tasks. Users' input was gathered with the think-aloud method during the tasks and a survey and semi-structured interview that followed. The results of the study indicated the following: (1) Users prefer detailed instructions over basic instructions when communicating with AI. Users always want to take the initiative and only want the AI to provide detailed explanations about its process when asked. (2) Although the AI can provide users with useful, effective, efficient, and fun experiences, it can lower the perceived predictability, comprehensibility, and controllability of drawing tasks.

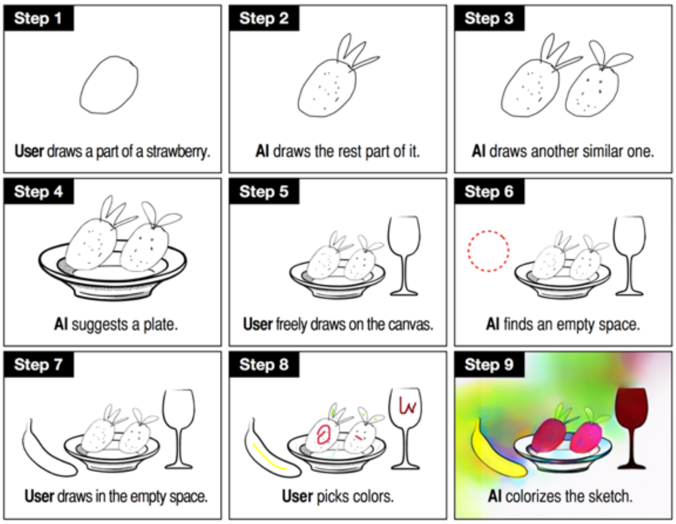


Figure 1-5. Drawing steps using DuetDraw

Based on these findings, design implications for user interfaces in which users and AI collaborate on creative work were derived: (1) Let users take the initiative. (2) Provide just enough instruction. (3) Embed interesting elements in the interaction. (4) Ensure a balance between AI elements.

1.7 How People Use a Practical Application of an Algorithm-based Systems Using Artificial Intelligence

As an applied case study of human-AI interaction, an automated journalism research project called NewsRobot was designed (Figure 1-6). NewsRobot automatically generated a series of summary news articles about the PyeongChang 2018 Winter Olympic Games in real time. It collected data on the results of major events and athlete details from official websites, processed the data, and input it into a predesigned article structure. The system was designed to generate news with two different types of content (general/selective) in three different styles (text/slide/video). By combining the two factors, it could produce six different types of news articles for every event.

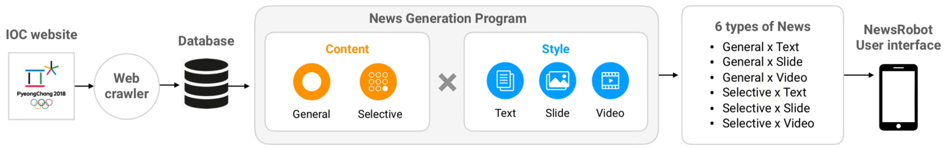


Figure 1-6. NewsRobot

A user study of NewsRobot with both quantitative and qualitative approaches was conducted. Thirty participants were recruited to watch Olympic Games events on TV and then shown six types of news article per event. They then answered questionnaires on each article and took part in semi-structured interviews. The result of the study can be summarized in the following three points: (1) Content: While users preferred selective news to general news, they considered selective news less credible than general news. (2) Style: As more news presentation elements were added, users' preference increased. People liked video news most, followed by slide news and then text news. In terms of quality, users rated slide news as clearer and more concise than video and text news. (3) Overall assessment: While people were satisfied with NewsRobot's accuracy, objectivity, personalization function, and various presentation elements, they found the articles were sometimes dull, repetitive, and out of context.

Based on these findings, the following design implications for user interfaces for algorithm-based automated news generation systems were derived: (1) Provide selective news with adaptable interfaces. (2) Present various multimedia elements without overwhelming the user. (3) Collect quality data to refine news generation algorithms.

1.8 Thesis Statement

Artificial intelligence algorithms and users have a subtle relationship. People tend to anthropomorphize as well as alienate AI. In particular, when given subjective information that was automatically computed by an algorithm, the user wants to know the rationale and communicate with the AI about it. The user does not want to lose control when continuing to interact with AI and wants a detailed description of the information they need. Users want to be provided with functionality that meets their expectations rather than AI showing off its capabilities.

1.9 Contributions

The core contribution of this thesis is to understanding AI algorithms in terms of human factors and user experience, investigating it with various topics and modes of HCI. Based on this, it can be divided into the following detailed contributions.

- **Empirical results on human–AI algorithm interaction:** Through both quantitative and qualitative approaches, this study closely observed the interaction between AI algorithms and users and discovered new aspects of this interaction. It investigated people's fear of AI from various perspectives and identified the confrontational "us vs. them" view between humans and AI, which is distinct from existing views on computers. This work also yielded experimental results showing how users' unique characteristics affected the process of interpreting the outcomes of AI algorithms in terms of strategy, and communication. Furthermore, the study provided insights on the user experience of an automated news generation system.

- **Research probes:** AI-powered user interfaces were designed for three of the studies, playing a crucial role in understanding the user's interactions: AI Mirror, a user interface that gives aesthetic scores to photographs based on a deep neural network model, DuetDraw, a collaborative drawing application based on neural network technology, and NewsRobot, an automated news generation system that produces multiple news articles considering content and style.

- **Design implications:** This thesis discussed design implications for intelligent user interfaces that are based on AI algorithms. These include implications for interfaces that deliver a variety of interpretable results, which could be utilized by both the AI/ML and HCI communities, interfaces with which users and AI can communicate and cooperate for creative work, and practical interfaces that provide users with information in various forms.

- **Theoretical contribution:** This thesis stresses the importance of AI algorithms and their human factors and user experience in the HCI field and suggests the concept of an expanded user interface and algorithmic experience.

1.10 Thesis Overview

Chapter 2 lays the groundwork for understanding how human interact with artificial intelligence algorithm by covering major research challenges in the context of related work.

The main part of the thesis introduces four studies of algorithm-based human-AI interactions. The first three of the four studies investigate the interaction between the algorithm and people with different stages. The final study is an application case study for practical applications. These chapters span motivations, methodology, findings and results of each user study.

- **Chapter 3** investigates how people perceive algorithm-based systems using AI with a case study of the Google DeepMind Challenge Match, a Go match between Lee Sedol and AlphaGo, in March 2016. This study explores the underlying and changing perspectives toward AI as users experienced this historic event.
- **Chapter 4** presents understanding of how people interpret and evaluate algorithm-based systems using AI. In this chapter, AI Mirror, an interface that tells users the algorithmically predicted aesthetic scores of photographs is introduced.
- **Chapter 5** presents understanding of how people build sequential actions with algorithm-based systems using AI. In this chapter, DuetDraw, an AI interface that allows users and the AI agent to draw pictures collaboratively is introduced.
- **Chapter 6** presents understanding how people use practical application of algorithm-based systems using AI. In this chapter, NewsRobot, a research prototype that automatically generates news on major events of the PyeongChang

2018 Winter Olympic Games is introduced.

Chapter 7 summarizes design lessons from the four studies and discuss major design dimensions and limitations to assist future researchers and practitioners in their human-AI design.

Finally, **Chapter 8** reviews the contributions of the thessis and proposes future research directions for human-AI interaction.

2 RELATED WORK

As preparation for this research, studies related to the following four themes were reviewed: human perception of AI algorithms, users' interpretation and evaluation of AI algorithms, how people build sequential actions with AI algorithms, and practical design of algorithm-based systems using AI.

2.1 Human Perception of AI Algorithms

This section addresses key topic areas related to human perception of AI algorithms: technophobia, and anthropomorphism.

2.1.1 Technophobia

For the most part, studies on technophobia have focused on investigating the relationship between computer anxiety and demographic variables, such as age [51], gender ([52]–[54]), personality [55], occupation [56], nationality [57], and cultural differences [58]. These studies have revealed the various factors affecting computer-using behaviors, such as the differences among users that can influence computer anxiety ([51], [56], [58]) or attitudes toward computers ([53], [54], [57]), and the factors that make customers hesitant to purchase computer devices [52].

This thesis differs from these previous studies regarding technophobia, in that it focuses on AI rather than simple computers. Although previous studies have shown the importance of understanding fear of technology, most have addressed the issue in terms of computer usage and made little attempt to account for AI technology. As AI includes *cognitive* functions that humans associate with other human minds, such as *learning* and *problem solving* [12], it can be distinguished from simple computers, which are regarded as tools to complete certain tasks. To elucidate and account for people's fear of AI technology, it is necessary to carry out a study approaching users' views on AI with novel perspectives, as well as embracing the previous studies.

2.1.2 Anthropomorphism

In order to further understand the human perspective on AI technology, this study reviewed the concept of anthropomorphism, the tendency to attribute human characteristics to inanimate objects. It involves attributing cognitive or emotional states to something based on observation in order to rationalize an entity's behaviors in a given social environment [59]. According to Dennett, people tend to interpret the behavior of an entity by treating it as if it were a rational agent governing its choice of action by considering its beliefs and desires [60]. Anthropomorphism is also linked to its inverse, dehumanization, the tendency to deny human-essential capacities to some agents and treat them like nonhuman objects [61].

Anthropomorphism has received attention in various disciplines beyond the field of psychology. It is well known that the concept has provided a useful mechanism for robot interaction design, especially in social robot research ([62], [63]). It has also long been discussed in the HCI domain. For example, Nass conducted various studies that investigated users' tendency to anthropomorphize computers when interacting with them ([64], [65]), and emphasized that people treat artificial objects like real humans [66] (CASA mode; computers as social actors).

2.2 User's Interpretation and Evaluation of AI Algorithms

This section addresses the related work of users' interpretation and evaluation of AI algorithms with three key topics: interpretability of algorithms and users' concerns; sense-making and gap between users and AI algorithms; and user control in intelligent systems.

2.2.1 Interpretability of Algorithms and User's Concerns

Despite the remarkable advances of artificial intelligence algorithms with the development of deep learning (DL), it has been pointed out that it is relatively difficult to understand how the internal principles and mechanisms of the algorithms work ([67]–[69]). To elucidate the principles of the algorithms, researchers of the AI/machine learning (ML) community have conducted various studies ([70]–[73]). Some research has been conducted on the topic of explanatory AI (XAI) ([9], [74], [75]). However, as algorithms extend to the domain of human creativity, where people can have various subjective interpretations, the issue of the interpretability of the results of algorithms and their principles will continue to be raised.

The HCI community has also regarded algorithms as an important research topic ([76]–[78]). In particular, many studies have focused on the fairness and transparency of algorithms ([79]–[81]). Some studies related to algorithmic fairness suggest that algorithms could be less objective than required, manifesting increased bias ([82], [83]). Furthermore, since algorithms affect diverse user communities, this can be extended not only to individual problems but also to social problems, such as racial injustice and economic in equality [84]. Users' concerns over the potential harm of algorithms could substantially affect their trust in user interfaces [84].

Ensuring the transparency of algorithms could be a useful way to overcome these problems. It can affect people's trust in a system and their interactions with the system ([79], [85]), and it can allow people to question and critique a system in order to develop appropriate reliance on it [86]. Therefore, it would be important in AI interface design to investigate whether users can understand the algorithm and consider it transparent and reasonable.

2.2.2 Sense-making and Gap between Users and AI algorithms

Sense-making is a set of processes initiated when people recognize the inadequacy (gap) of their current understanding of events ([87], [88]). In this situation, individuals build, verify, and modify their mental models to account for the unrecognized features. It is reported that knowledge and expertise in related fields could affect people's sense-making processes [89]. Since the concept has been considered as a framework to understand the interaction between people and information technology [90], numerous studies have used it as a research method ([91], [92]) or introduced interactive systems for supporting it ([93], [94]).

The concept and framework could also be applied to the 196 understanding of how people reason about the results of AI algorithms. As AI algorithms are producing and communicating results that go beyond what people can understand, there could be a difference between the results of AI and human perceptions. In addition, because people's expertise in AI/ML varies so widely, this can have a different impact on people making sense of AI algorithms. Under these circumstances, looking at the processes people use to reduce the difference between their thinking and the results of AI algorithms can provide important discoveries about how people interact with AI algorithms. It is also important to look at how this process will differ, especially when each person has different expertise, because AI algorithms are spreading to various fields of expertise and interacting with various users other than AI/ML experts.

2.2.3 User Control in Intelligent Systems

In the HCI community, there has been discussion of how users and automated systems communicate [95]. Some have conducted research based on the idea of developing an adaptive and intelligent agent that automatically responds to user behavior ([96], [97]). In contrast, other groups of scholars have argued that a system encouraging users' ability to manipulate interfaces directly should be considered [98]. In addition, a mixed-initiative viewpoint has been raised that combines the two to take advantage of each ([99]–[101]). Recent advances in AI algorithms have rekindled interest in these discussions, since AI algorithms can now respond to user behavior more intelligently than ever before and users are communicating in new ways rather than simply manipulating the interface. Accordingly, in this study, the interaction between the user and AI algorithms will be closely observed and analyzed to see what control and communication could provide value to users and extend this discussion.

2.3 How People Build Sequential Actions with AI Algorithms

This section addresses key topic areas related to how people build sequential actions with AI algorithms: AI, deep learning, and new user experience in creative works and communication and leadership between humans and computers.

2.3.1 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 ([23], [102], [103]), but also new interfaces ([104], [105]), 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 ([106], [107]), musical composition ([108], [109]), and drawing [110].

AI algorithms are also being applied in creative areas considered unique to humans, such as writing ([106], [107]), musical composition ([108], [109]), and drawing [110]. In drawing, several new interfaces using deep learning have been introduced. *Quick, Draw!* [111] 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* [112] 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 ([113], [114]).

Various deep learning algorithms that can support these drawing interfaces have been developed. *Sketch-RNN* [115] 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 [116]. Gatys et al. also developed a neural network for blending the content of one image and the style of another image [117]. 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 ([11], [118]). In this respect, by combining AI algorithms and perspectives of prior studies, this study designed a prototype with which humans and AI can produce complex and creative output such as drawn pictures.

2.3.2 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 [43]. 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 ([43], [119]). 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 [120]; 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 ([95], [121]). 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 ([100], [122]).

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, this study focuses on the communication and initiative issues and explores how these can affect the user experience of interfaces in which users and AI work together.

2.4 Practical Design of Algorithm-based Systems Using AI

This section addresses the related work of practical design of algorithm-based systems using AI with three key topics: automated journalism; personalization of news content; and the effect of multimedia modality level on the user experience.

2.4.1 Automated Journalism

In line with the attempt to apply technology to journalism ([123]–[125]), automated journalism (also known as robot journalism), the generation of news articles using computer software instead of human reporters, has been introduced ([126]–[128]). Since the technology can create a large number of news articles quickly and accurately, it has attracted the attention of both the media industry ([129]–[131]) and scholars ([132]–[134]). Many researchers have searched for technical ways of generating news articles that are indistinguishable from those of human reporters ([135], [136]). In addition, these technical explorations have brought about extended discussion of the sociocultural impacts [137], such as authorship [128], labor substitution ([138]–[141]), and algorithm power ([142]–[144]).

In the field of HCI, several studies have suggested different kinds of authoring tools to support news content generation ([145]–[147]). Recently, Kim and Lee proposed a five-step robot journalism framework for automated news content generation [148], which consisted of (1) data crawling, (2) event extraction, (3) key event detection, (4) mood detection, and (5) news article generation. Based on this framework, they also created an algorithm-driven interactive news generation system, revealing its capability of generating news stories that are significantly more interesting and enjoyable than traditional news articles [149].

Meanwhile, recent discussions related to automated news emphasize dealing with the user experience of the news media more closely ([150], [151]). As news is being consumed through various online media, and various news curation algorithms are being created, many problems, such as filter bubbles and misinformation, are getting worse ([83], [152]). Furthermore, if the algorithms are actively involved in news generation as well as news curation, there is a possibility that these problems will worsen even more.

In this circumstance, this study aims to design and create an automated news system with various factors, identify the potential of automated journalism, and evaluate them from various perspectives, and create a basis for a discussion of a system design that promotes a desirable user experience.

2.4.2 Personalization of News Content

Personalization involves activities that tailor information or services to users employing knowledge about them to achieve targeted goals [153]. As an important social phenomenon generating significant economic value, it has drawn a lot of attention from various areas, such as economics [154], management ([155], [156]), marketing ([157], [158]), information systems ([159]–[161]), and computer science ([162], [163]).

Of course, personalization is one of the most critical issues in automated journalism ([164], [165]). In automated journalism, software algorithms can use the same data to tell stories from different angles, customizing languages, topics, tones, and styles according to an individual reader's preferences. Research has demonstrated that personalized messages can engage and persuade an audience more effectively than generic mass messages ([166], [167]).

On the other hand, another study revealed that the use of a selective news system can have a direct negative effect on knowledge [161]. Excessive personalization through automation also has the potential to allow users to collect and consume only biased information. The problem of transparency and fairness of algorithms in the personalization process is also constantly being raised.

As the personalization of news can lead to various positive and negative effects on the user experience, this work pursues further research on these issues and practical discussions to enhance users' experience of the system while preventing negative aspects.

2.4.3 Effect of Multimedia Modality on User Experience

As people now consume news media through smartphone web browsers, the resources used to produce news articles are no longer limited to text but also include images, audio/video clips, and graphic animations ([168], [169]). Adding multimedia resources to news articles can increase audiences' engagement in news articles [170]. However, most automated journalism research has focused on text news generation ([28], [135]). It is necessary to recognize the possibility of using such multimedia and to introduce these factors into news generation.

Generally, in interface design, the use of multimedia can increase the expressiveness, usability, and enjoyment of computer interfaces as well as content itself [171]. On the other hand, simply increasing the multimedia modality level of an information system can increase its complexity and adversely affect its ability to convey information [172]. Introducing multimedia modalities requires more attention to user needs and comprehensive field studies to investigate the most appropriate solutions for each kind of application and user category [172].

Based on these perspectives, this study aims to investigate how the user experience of watching news generated by an automated journalism system is affected as multimedia modality elements (from text to images and audio/video clips) increase and discuss the appropriate level of multimedia modality in automated news generation systems.

3 HOW PEOPLE PERCEIVE ALGORITHM-BASED SYSTEMS USING ARTIFICIAL INTELLIGENCE

Various forms of artificial intelligence, such as Apple's Siri and Google Now, have permeated our everyday lives. However, the advent of such "human-like" technology has stirred both awe and a great deal of fear. Many consider it a woe to have an unimaginable future where human intelligence is exceeded by AI. This chapter investigates how people perceive and understand AI with a case study of the Google DeepMind Challenge Match, a Go match between Lee Sedol and AlphaGo, in March 2016. This chapter explores the underlying and changing perspectives toward AI as users experienced this historic event. Interviews with 22 participants show that users tacitly refer to AlphaGo as an "other" as if it were comparable to a human, while dreading that it would come back to them as a potential existential threat. Our work illustrates a confrontational relationship between users and AI, and suggests the need to prepare for a new kind of user experience in this nascent socio-technological change.[2] It calls for a collaborative research effort from the HCI community to study and accommodate users for a future where they interact with algorithms, not just interfaces.

3.1 Motivation

The advancement of machine learning, the explosive increase of accumulated data, and the growth of computing power have yielded artificial intelligence technology comparable to human capabilities in various fields ([173]–[175]). It allows common users to interact with intelligent devices or services using AI technology, such as Apple's Siri and Amazon's Echo. It is expected that AI technology will become increasingly prevalent in various areas, such as autonomous vehicles ([2], [3]), medical treatment ([4], [5]), game playing ([6], [7]), and customized advertisements [8].

However, some people fear AI. Many scientists, including Stephen Hawking and Ray Kurzweil, have expressed concerns about the problems that could arise in the age of AI ([176], [177]). According to an online survey conducted by the British Science Association (BSA) [178], about 60% of respondents believe that the use of AI will lead to fewer job prospects within 10 years, and 36% believe that the development of AI poses a threat to the long-term survival of humanity. Such fears could lead to a downright rejection of technology [179], which could have negative effects on individuals and society [180]. Therefore, making a sincere attempt at understanding users' views on AI is important in the area of human-computer interaction.

[2] This chapter has adapted, updated, and rewritten content from a paper at CHI 2017 [77]. All uses of "we," "our," and "us" in this chapter refer to coauthors of the paper.

This chapter aims to investigate the various aspects of people's fear of AI and the potential implications for future interfaces. To this end, we took the Google DeepMind Challenge Match [181] as a case study. The match comprised a five-game Go match between Lee Sedol, a former world Go champion, and AlphaGo, a computer Go program developed by Google DeepMind. The match was held in Seoul, Korea, in March 2016. Before the match, Lee was expected to defeat AlphaGo easily. However, he lost the first three games in a row. Although Lee won the fourth game, AlphaGo won the overall match. The result shocked and amazed many people, provoking public discussion. The match provided a good opportunity to investigate the general public's opinions, responses, and concerns about AI.

To investigate and understand users' fear of AI, or more specifically of AlphaGo, we recruited 22 participants and carried out semi-structured interviews about the match. While conducting the study, we intentionally left the term AI undefined so that we could collect various conceptions on AI without prompting participants. We identified that people had a dichotomous ("us vs. them") view of AI. The findings from the study can be summarized as follows:

- People had preexisting stereotypes and prejudices about AI, mostly acquired from media such as Hollywood movies. They believed that AI could cause harm to humans, and that AI should assist and help humans.
- People's thoughts changed according to the result of each game of the match. At first, people were immensely shocked and apprehensive. As the match progressed, they began to cheer for Lee Sedol as if he were a representative of all of humanity.
- People not only anthropomorphized but also alienated AlphaGo. People evaluated AlphaGo based on its perceived human characteristics.

- People expressed concerns about a future society where AI is widely used. They worried that their jobs would be replaced by AI and humans would not be able to control the AI technology.

Based on these findings, we discuss the current awareness of AI from the public and its implications for HCI as well as suggestions for the future work. The rest of this chapter describes related works and the basic information on the match, then details the study design and findings, followed by a discussion.

3.2 Google DeepMind Challenge Match

The Google DeepMind Challenge Match [181] was a Go match between Lee Sedol, a former world Go champion, and AlphaGo, an AI Go program. It took place in Seoul, South Korea, between March 9 and 15, 2016. Since Go has long been regarded as the most challenging classic game for AI, the match brought a lot of attention from AI and Go communities worldwide. The match consisted of five games, and by winning four of the five games, AlphaGo became the final winner of the match. A detailed explanation of the players is as follows:

- ***AlphaGo*** is a computer Go program developed by Google DeepMind. Its algorithm uses a combination of a tree search and machine learning with extensive training from human expert games and computer self-play games [37]. Specifically, it uses a state-of-the-art Monte Carlo tree search (MCTS) guided by two deep neural networks: the "value network" to evaluate board positions and the "policy network" to select moves [37]. It is known as the most powerful Go-playing program ever, and the Korea Baduk Association awarded it an honorary 9-dan ranking (its highest).

- ***Lee Sedol*** is a South Korean professional Go player of 9dan rank. He was an

18-time world Go champion, and he won 32 games in a row in the 2000s. Although he is no longer the champion, he is still widely acknowledged as the best Go player in the world. Unlike the traditional Go playing style of slow, careful deliberation, he reads a vast number of moves and complicates the overall situation, finally confusing and annihilating his opponent. This style created a sensation in the Go community. Many Koreans consider him a genius Go player.

The result of the match was a big surprise to many people, as it showed that AI had evolved to a remarkable level, even outdoing humanity in an area requiring advanced intelligence. After the first game, Demis Hassabis, the DeepMind founder, posted the following tweet: *#AlphaGo WINS!!!! We landed it on the moon* [1]. This implied that it was a very important moment in the history of AI research and development.

However, in Korea where the game was held, the defeat of Lee caused a tremendous shock. This was partly due to the cultural implications of Go in Korea and its popularity. Go is considered one of the most intellectual games in East Asian culture, and it is extremely popular in Korea. Many people were very interested in the match before and after the event. In addition, people's expectation of Lee's victory was huge. Since most people expected that Lee would win the game, they could not accept the result that AlphaGo, the AI Go program, had defeated Lee, the human representative. Every match day, all national broadcasting stations in Korea reported the shocking news as top stories. Throughout the country, both online and offline, people talked about the event, expressing fear of AI as well as AlphaGo. We believed that the public discussion on this event could provide a unique opportunity to assess and understand people's fear of AI technology. Therefore, we tried to investigate the underlying implications of this event, and we designed and conducted a user study accordingly.

3.3 Methodology

To obtain insights about people's fear of AI, we designed and conducted semi-structured interviews with 22 participants from diverse backgrounds.

P#	Age	Sex	Occupation
P01	23	M	university student (computer science)
P02	24	F	university student (industrial design)
P03	25	F	blog writer
P04	26	M	university student (business)
P05	27	M	television producer
P06	28	F	vision mixer
P07	30	F	web designer
P08	30	M	lawyer
P09	31	M	environmentalist
P10	34	M	professional photographer
P11	35	M	Go player and teacher (amateur 7 dan)
P12	37	F	researcher (educational statistics)
P13	40	M	bookstore manager
P14	45	F	administrative worker
P15	45	M	consultant
P16	48	F	middle school teacher (English)
P17	49	F	food manager
P18	52	F	tax accountant
P19	55	M	taxi driver
P20	58	M	journalist
P21	58	M	accountant
P22	60	F	social worker

Table 3-1. Age, gender, and occupation of participants

3.3.1 Participant Recruitment

The interviews were designed to identify participants' fear of the match and obtain diverse opinions about AI. Inclusion criteria included basic knowledge of the match and experience watching the match through media at least once. In addition to this requirement, a demographically representative set of participants was sought. The target ages were divided into four categories: 20s and under, 30s, 40s, and 50s and over. We also considered the occupations of participants. We recruited participants living in the Seoul Metropolitan Area, disseminating the recruitment posters at local stores, schools, and community centers. At first, we recruited 15 participants who saw the posters and contacted us directly. Then we recruited seven additional participants through snowball sampling and contacts of the researchers so that we had an evenly spread group of participants in terms of age, gender, and occupation. A total of 22 diverse participants were recruited in the study, as shown in Table 3-1. The participants were each given a $10 gift card for their participation.

3.3.2 Interview Process

Each participant took part in one semi-structured interview after the entire match was over. As we aimed to collect various ideas on AI without prompting participants, we intentionally left the term AI undefined before and during the interview. Considering this, we designed the interview questions, and each interview was guided by the following four main issues: the participants' preexisting thoughts and impressions of AI, changes in their thoughts as the match progressed, impressions of AlphaGo, and concerns about a future society in which AI technology is widely used. As we sought to identify participants' thoughts over time, we carefully designed the questions separately according to the match schedule and provided detailed information of the match so that the participants could situate themselves in the match context. Then, to further induce the participants' diverse and profound thoughts, we provided them with 40 keyword cards covering various issues related to AI, which were extracted from AI and AlphaGo Wikipedia articles. We showed the cards in a set to the participants and let them pick one to three cards so that they could express their thoughts about certain issues they otherwise might have missed. We conducted the interviews at places of each participant's choosing, such as cafes near their offices. Each interview took about an hour. The participants responded more actively during the interviews than the authors had expected.

3.3.3 Interview Analysis

Interviews were transcribed and analyzed using grounded theory techniques [182]. The analysis consisted of three stages. In the first stage, all the research team members reviewed the transcriptions together and shared their ideas, discussing main issues observed in the 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 [183], a qualitative research software provided by Optimal Workshop. We segmented the transcripts by sentence and entered the data into the software. While reviewing the data, we annotated multiple keywords in each sentence so that the keywords could summarize and represent the entire content. A total of 1,016 keywords were created, and we reviewed the labels and text again. 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 five main categories, described below.

The research design protocol was reviewed and approved by the Institutional Review Board of Seoul National University (IRB number: 1607/003-011), and we strictly followed the protocol. All interviews were recorded and transcribed in Korean. The quotes were translated into English, and all participants' names were replaced by pseudonyms.

3.4 Findings

The interviews revealed that participants felt fear related to the match and AI and had a confrontational relationship as in "us vs. them." People had preconceived stereotypes and ideas about AI from mass media, and their thoughts changed as the match progressed. Furthermore, people not only anthropomorphized but also alienated AlphaGo, and they expressed concerns about a future society where AI will be widely used.

3.4.1 Preconceptions about Artificial Intelligence

We identified that people had preconceptions and fixed ideas about AI: AI is a potential source of danger, and AI should be used to help humans.

Artificial Intelligence as Potential Threat

Throughout the interviews, we identified that the participants had built an image of AI in their own way, although they had rarely experienced AI interaction firsthand. When asked about their thoughts and impressions of the term AI, most of the participants described experiences of watching science fiction movies. They mentioned the specific examples, such as Skynet from *Terminator* (1984), Ultron from *The Avengers* (2015), Hal from *2001: A Space Odyssey* (1968), sentient machines from *The Matrix* (1999), and the robotic boy from *A.I. Artificial Intelligence* (2001). In addition, P15 described a character from Japanimation from his youth. The characters were varied, from a man-like robot to a figureless control system.

Notably, the participants formed rather negative images from their media experiences, since most of the AI characters were described as dangerous. Many AI characters in the science fiction movies that the participants mentioned controlled and threatened human beings, which seemed to reinforce their stereotypes. Some of the participants agreed that this might have affected the formation of their belief that AI is a potential source of danger.

Meanwhile, the movie experiences made people believe that the AI technology is not a current issue but one to come in the distant future. Generally, the movies mentioned by participants were set in the remote future, and their stories were based on futuristic elements, such as robots, cyborgs, interstellar travel, or other technologies. Most of the technologies described in the movies are not available at present. For example, P14 said, "*The Artificial intelligence in the movie seemed to exist in the distant future, many years from now.*"

Artificial Intelligence as a Servant

It was found that many participants had established their own thoughts about the relationship between humans and AI. They believed that AI could not perform all the roles of people. However, they thought that AI could perform arduous and repetitive tasks and conduct those tasks quickly and easily. For example, P17 said, "*They can take on tasks that require heavy lifting, and they can tackle complex problems.*"

The expected roles of AI the participants mentioned were associated with their perceptions of its usefulness. Although some of the participants regarded AI as a potential threat, they partially acknowledged its potential convenience and abilities. For example, P15 said, "*They are convenient; they can help with the things humans have to do. They can do dangerous things and easy and repetitive tasks, things that humans do not want to do.*"

However, this way of thinking suggests a sense of human superiority in the relationship at the same time. More specifically, P13 commented, "*Man should govern artificial intelligence. That is exactly my point. I mean, AI cannot surpass humans, and it should not.*"

3.4.2 Confrontation: Us vs. Artificial Intelligence

Changes in participants' thoughts were observed according to the result of each game of the match. People indicated fear of AlphaGo at first, but as the match progressed, they began to cheer for Lee Sedol as a human representative. In this process, people tended to have a confrontational relationship with AlphaGo of an "us vs. them" type.

Prior to the Match: Lee Can't Lose

Before the match began, all the participants but two expected that Lee Sedol would undoubtedly win. For example, P12 said, "*I thought that Lee Sedol would win the game easily. I believed in the power of the humans.*" P04 said, "*Actually, I thought Lee would beat AlphaGo. I thought even Google wouldn't be able to defeat him.*" This showed his tremendous belief in Lee Sedol. The conviction about Lee's ability to win was almost like blind faith. During the interviews, we described AlphaGo's ability for the participants in detail, explaining its victory in the match with the European Go Champion and its capacity in terms of its overwhelming computing power and its learning ability. However, some of them said they already knew the information but still thought that AI could not win the game. Some participants provided us with several reasons for their conviction. P15 suggested the complexity of Go as a reason. He explained that the number of possible permutations in a Go game is larger than the number of atoms in the universe.

Game 1: In Disbelief (Lee 0:1 AlphaGo)

Even though many people expected that Lee would win the match, he lost the first game by resignation, which shocked many people in Korea. Although Lee appeared to be in control throughout much of the match, AlphaGo gained the advantage in the final 20 minutes and Lee resigned. He said that the computer's strategy in the early part of the game was "*excellent.*" Since the participants had been convinced of Lee's victory, their shock was far greater. P05 said, "*I thought, this can't happen. But it did. What a shock!*" P04 also showed his frustration, saying, "*AlphaGo won by a wide margin. It was shocking. Lee is the world champion. I couldn't understand how this could be.*"

In addition, some participants said that their attitudes toward the match changed after the first game. Before the game, they just intended to enjoy watching an interesting match; however, when they saw the result, they began to look at the game seriously. P22 commented, "*After the first game, I realized it was not a simple match anymore.*" Furthermore, some participants began to think that Lee might possibly lose the remaining games. On the other hand, some participants thought that Lee still had a 50-50 chance of winning. P13 said, "*I thought mistakes caused Lee's defeat. People can make mistakes. If he can reduce his mistakes, he can win the remaining games.*"

Game 2: We Can't Win (Lee 0:2 AlphaGo)

Lee suffered another defeat in the second match, and people began to realize AlphaGo's overwhelming power. During the post-game interview, Lee stated, "*AlphaGo played a nearly perfect game from the very beginning. I did not feel like there was a point at which I was leading.*" Now, people started to regard AlphaGo as undefeatable. P04 said, "*Lee was defeated in consecutive games. It was shocking. I began to acknowledge AlphaGo's perfection. Actually I didn't care that much when Lee was beaten in the first round. But after the second round, I had changed my mind. I realized AlphaGo could not be beaten, and it was terrifying.*" P01 commented, "*Before the match, I firmly believed Lee would win. But the second game totally changed my mind. It was really shocking.*" P02 also noted, "*After the second match, I was convinced that humans cannot defeat artificial intelligence.*"

Game 3: Not Much Surprise (Lee 0:3 AlphaGo)

When AlphaGo won the first three consecutive games, it became the final winner of the best-of-five series match. Interestingly, the impact of the third game was not as strong as that of the second or the first. As people had already witnessed AlphaGo's overwhelming power, most of them anticipated that Lee would lose again, and the result was no different from what they had expected. For example, P14 said, "*He lost again... yeah I was sure he would lose.*"

People became sympathetic and just wanted to see Lee win at least once. P07 said, "*I wanted to see just one win. I was on humanity's side. 'Yes, I know you (AlphaGo) are the best, but I want to see you lose.'* " P01 said, "*I didn't expect that Lee could win the remaining games. But I wished he would win at least once.*"

Game 4: An Unbelievable Victory (Lee 1:3 AlphaGo)

Surprisingly, in the fourth game, Lee defeated AlphaGo. Although Lee struggled early in the game, he took advantage of AlphaGo's mistakes. It was a surprising victory. AlphaGo declared its surrender with a pop-up window on its monitor saying "*AlphaGo Resigns.*" The participants expressed joy over the victory and cheered for Lee. Some of them saw the result as a "*triumph for humanity.*" P14 said, "*He is a hero. It was terrific. He finally beat artificial intelligence.*" P09 said, "*I was really touched. I thought it was impossible. AlphaGo is perfect, but Lee was victorious. It was great. He could have given up, but he didn't, and he finally made it.*"

Game 5: Well Done, Lee (Lee 1:4 AlphaGo)

AlphaGo won the fifth and final game. Before the beginning of the game, the participants mainly thought that AlphaGo would win. However, at the same time, since they had seen the victory of Lee the day before, they also slightly anticipated Lee's win.

P10 said, "*In a way, I thought Lee still had a chance to win the last game. I thought as AlphaGo learned, Lee might learn and build a strategy to fight.*" On the one hand, some participants were already satisfied by the one win. They were relieved that the game was over. P08 said, "*Anyhow, I was contented with the result the day before. The one victory made me relieved. I watched the last game free from all anxiety.*" P12 also stated, "*Lee played very well. I really respect his courage.*"

To sum up, throughout the five-game match, people were immensely shocked and apprehensive at first, but they gradually began to cheer for Lee Sedol as a human representative as the match progressed. The participants thought that Lee and AlphaGo had the fight of the century, humanity vs. AI. Through this match, people were able to recognize what AI is and how the technology has been developed. The participants also identified various characteristics of AI generally as well as AlphaGo specifically.

3.4.3 Anthropomorphizing AlphaGo

Throughout the interviews, we observed that the participants anthropomorphized AlphaGo. They referred to AlphaGo as if it were a human and differentiated AI technology from personal computers. They also thought AlphaGo was creative, which is usually considered a unique characteristic of human beings. This tendency has similar to Nass's concept of CASA (computers as social actors) [184].

AlphaGo is an "Other"

When describing their thoughts and impressions of AlphaGo, people always used the name "*AlphaGo*" as if it were a human being. Moreover, they often used verbs and adjectives commonly used for humans when mentioning AlphaGo's actions and behaviors, such as "*AlphaGo made a mistake,*" "*AlphaGo is smart,*" "*AlphaGo learned,*" and "*AlphaGo practiced,*" which indicates a tendency of anthropomorphization. One participant even called AlphaGo "*buddy*" from the beginning to the end of the interview. When asked if there was a special reason for this, she said, "*I know it is a computer program. But he has a name, 'AlphaGo.' This makes AlphaGo like a man, like 'Mr. AlphaGo.' I don't know exactly why, but I think because of his intelligence, I unconsciously acknowledged him as a buddy. 'You are smart. You deserve to be my buddy.'* " P06 went so far as to call AlphaGo someone we should acknowledge as our superior, saying, "*After the third game, I realized that we were bust. We had lost. AI is the king. We should bow to him.*"

AlphaGo Is Different from a Computer

Moreover, we identified that the participants anthropomorphized AI, as well as AlphaGo specifically, by drawing a sharp distinction between personal computers and AI technology. The participants uniformly described the two as different. While they regarded the computer as a sort of tool or implement for doing certain tasks, they described AI as not a tool but an agent capable of learning solutions to problems and building its own strategies. They thought that we could control a computer as a means to an end but that we could not control AI. They said AI knows more than we do and thus can undertake human work. P07 said, "*I think they (computers and AI) are different. I can't make my computer learn something. I just use it to learn something else. However, artificial intelligence can learn on its own.*" P08 also commented, "*When I first learned about computers, I thought they were a nice tool for doing things quickly and easily. But artificial intelligence exceeds me. AI can do everything I can do, and I cannot control it. The main difference between computers and AI is our ability to control it.*"

AlphaGo is Creative

Some of the participants said AlphaGo's Go playing style was somewhat creative, since AlphaGo made unorthodox, seemingly questionable, moves during the match. The moves initially befuddled spectators. However, surprisingly, the moves made sense in hindsight and determined the victory. In other words, to the viewers, AlphaGo calculated the moves in a different way from human Go players and thus finally won the game. Some participants thought that AlphaGo showed an entirely new approach to the Go community. P09 said, "*It was embarrassing. AlphaGo's moves were unpredictable. It seemed like he had made a mistake. But, by playing in a different way from men, he took the victory. I heard he builds his own strategies by calculating every winning rate. People learn their strategies from their teachers. But AlphaGo discovered new ways a human teacher cannot suggest. I think we should learn from AlphaGo's moves.*" P11, an amateur 7-dan

Go player and Go Academy teacher, also demonstrated this view based on his own experience. In the past, he had learned Go through the apprentice system. Although he learned the basic rules of Go, for him, learning was imitating the style of his teacher or the best Go players. His learning was focused not on how to optimize the moves with the highest winning rate but on how to find the weak spot of the champion of the day. He said, Go styles also have followed the main trend when a new champion appears. However, AlphaGo's moves were entirely different from this Go style and trend, which seemed creative and original to P11.

3.4.4 Alienating AlphaGo

People also alienated AlphaGo by evaluating it with the characteristics of a human. They sometimes showed hostility toward AlphaGo and reported feeling negative emotions toward it.

AlphaGo is Invincible

All participants agreed that AlphaGo, "*compared to a human being,*" has an overwhelming ability. AlphaGo was trained to mimic human play by attempting to match the moves of expert players from recorded historical games, using a database of around 30 million moves. Once it had reached a certain degree of proficiency, it was trained further by playing large numbers of games against other instances of itself, using reinforcement learning to improve its play. The participants concurred with the idea that a human's restricted calculation ability and limited intuition cannot match AlphaGo's powerful ability. P22 said, "*Lee cannot beat AlphaGo. AlphaGo learns ceaselessly every day. He plays himself many times a day, and he saves all his data. How can anyone beat him?*" P15 said, "*I heard that AlphaGo has data on more than a million moves, and he studies the data with* *various approaches for himself. He always calculates the odds and suggests the best move, and he can even look further ahead in the game than humans. He knows the end. He knows every possible case."* P15 even argued that the match was unfair. He contended that unlike Lee Sedol, who was trying to win alone, AlphaGo was linked to more than 1,000 computers, and this made its computation power far superior to that of human beings. For these reasons, he insisted that Lee was doomed from the beginning of the game and that the result should also be invalid. "*It was connected with other computers... like a cloud? Is it the right word? It is like a fight with 1,000 men. Also, computers are faster than humans. It is unfair. I think it was unfair."*

AlphaGo is Ruthless

Throughout the match, the participants referred to AlphaGo's ruthlessness and heartlessness, which are "*uncommon traits in humans."* Usually, when pro Go players play the game, a subtle tension arises between the players. Identifying the opponent's feelings and emotions could be significant, and emotional elements can affect the result of the match. However, AlphaGo could never express any emotion throughout the match. P20, who introduced himself as having a profound knowledge of Go, commented that there was not a touch of humanity about AlphaGo's Go style. He said, "*AlphaGo has no aesthetic sense, fun, pleasure, joy, or excitement. It was nothing like a human Go player. Most pro Go players would never make moves in that way. They leave a taste on the board. They play the game with their board opened. But AlphaGo tried to cut off the possibility of variation again and again."* At that time, the term "*AlphaGo-like"* became widely used as an adjective in Korea, meaning ruthless, inhibited, and emotionally barren. One of our participants, P07, used the term: "*Since the match, I often call my husband*

'Alpha Park' because there is no sincerity in his words."

AlphaGo is Amorphous

"Unlike human Go players," AlphaGo has no form and only exists as a computer program, which left a deep impression on the viewers. Since AlphaGo is only an algorithm, it showed its moves through the monitor beside Lee. Then, Aja Huang, a DeepMind team member, placed stones on the Go board for AlphaGo, which ran through Google's cloud computing, with its servers located in the United States. At first, people wondered who AlphaGo was. Some participants thought Aja Huang was AlphaGo, modeled on the human form. P03 said,

"My mom said she mistook Aja Huang for AlphaGo. I think people tend to believe that artificial intelligence has a human form, like a robot. If something has intelligence, then it must have a physical form. Also, artificial intelligence is an advanced, intelligent thing like a human, which makes people think its shape must also be like that of a human." One participant even believed Aja Huang was AlphaGo until the interview. She said, "*Wasn't he AlphaGo? I didn't know that. It's a little weird, don't you think?"*

AlphaGo Defeats Man

All participants said that AlphaGo induced negative feelings toward AI. As described above, usually, AI was still considered something that would only occur in the distant future. However, the AlphaGo event showed that the technology is already here. The event made people realize that AI was near. P08 said, "*Although Lee won once, he finally lost. This is a symbolic event of artificial intelligence overpowering humans. I'm sure this event made me feel the power of artificial* *intelligence in my bones."* P07 also said, "*Before I watched the match, I had no idea about who was developing artificial intelligence and how much the technology had developed. But now I know a few things about artificial intelligence. AlphaGo taught me how powerful the technology is."*

In this regard, we observed that AlphaGo affected the formation of people's negative emotional states. Some of the participants told us that they felt helplessness, disagreeability, depression, and a sense of human frailty and suffered from stress while watching the match. Furthermore, they said the result of the match knocked their confidence and increased their anxiety. If this was not true of themselves, they said they commonly saw the people around them suffering for the same reason. P03 noted, "*AlphaGo can easily achieve any goal. But I have to devote my entire life to reaching a goal. No matter how hard I try, I cannot beat AI. I feel bad and stressed. I've lost my confidence."* P08 also stated, "*The human champion was defeated by AlphaGo. He was completely defeated. Humans cannot catch up with artificial intelligence. I started to lose my confidence and feel hostility toward artificial intelligence. I became lethargic."* P20 said that, "*If I had a chance to compete with AlphaGo, I think I would give up because it would be a meaningless game."*

3.4.5 Concerns about the Future of AI

After witnessing the unexpected defeat of Lee Sedol, people also raised concerns over a future society where AI technology is prevalent. They especially worried that they would be replaced by AI and not be able to follow and control the advancement of AI.

Man is Replaced

People expressed their worry that AI will one day be able to perform their jobs, leaving them without work. They worried that as AI will be widely developed in many different fields, the technology will surpass the human endeavors in these areas. They thought that, as a result, because of the comparative advantages, AI will be preferred, and the demand for human labor will decrease. Moreover, they believed the problem of the replacement of humans is not confined to simple, and repetitive tasks. They thought it could happen in the specialized occupations, such as lawyers and doctors. For example, P08, a lawyer, recently had a talk about this issue with his colleagues. He said, "*Actually, lawyers have to perform extensive research into relevant facts, precedents, and laws in detail while writing legal papers. We have to memorize these materials as much as we can. But we can't remember everything. Suppose they created an AI lawyer. He could find many materials easily, quickly, and precisely. Lawyers could be replaced soon.*"

Fear of losing jobs raised the question of the meaning of human existence. Some participants said they felt the futility of life. P06 said, "*We will lose our jobs. We will lose the meaning of existence. The only thing that we can do is have a shit. I feel as if everything I have done so far has been in vain.*" P13 showed extreme hostility toward AI, saying, "*If they replace humans, they are the enemy,*" which is reminiscent of the Luddites, the movement against newly developed labor-economizing technologies in the early 19th century.

They worried that AI will also encroach on the so-called creative fields, the arts, which are regarded as unique to human beings. Some participants talked about a few news stories indicating that AI can perform comparably to human beings in painting, music composition, and fiction writing. They thought that there is nothing that human beings can do in such situations. P01 described his thoughts about this AI's encroachment on the art area based on his anecdote of seeing software that automatically transforms any image into Van Gogh's style. He said, "*Seeing AI invade the human domain broke my stereotype.*"

If human beings were replaced by AI in all areas, what would we do then? We also found this question raised with respect to education. P16, a middle school teacher, explained her difficulty in career counseling and education programs for students. She said, "*The world will change. Most jobs today's children will have in the future have not been created yet.*" Since she could not anticipate which jobs would disappear and which ones would be created in the future, she felt skeptical about teaching with the education content and system designed based on contemporary standards.

Singularity is Near

Some participants expressed their concerns about a situation in which humans cannot control the advancement of AI technology. This worry is related to the concept of the *technological singularity* [177], in which the invention of artificial superintelligence will abruptly trigger runaway technological growth, resulting in unfathomable changes to human civilization. According to the singularity hypothesis, an autonomously upgradable intelligent agent would enter a 'runaway reaction' of self-improvement cycles, with each new and more intelligent generation appearing more and more rapidly, causing an intelligence explosion and resulting in a powerful superintelligence that would far surpass all human intelligence. After seeing AlphaGo build his own strategies that went beyond human understanding and easily beat the human champion, the participants thought that the singularity could be realized soon in every field and that humans would not be able to control the technology. P06 said, "*It's terrible. But the day will come. I can only* *hope the day is not today."* The participants unanimously insisted that society needs a consensus about the technology and that laws and systems should be put in place to prevent potential problems.

AI could be Abused

The participants also expressed their concerns that AI technology might be misused. The AlphaGo match has demonstrated its ability to many people around the world. They worried that the overwhelming power of AI could lead some people to monopolize and exploit it for their private interests. They said that if an individual or an enterprise dominates the technology, the few who have the technology might control the many who do not. P04 said, "*I think that one wealthy person or a few rich people will dominate artificial intelligence."* P01 also noted, "*Of course, artificial intelligence itself is dangerous. But I am more afraid of humans, as they can abuse the technology for selfish purposes."* Some participants argued that if the technology were monopolized, the inequality between those who have it and those who do not would become more severe. For example, P13 said, "*I agree with the opinion that we need to control AI. But who will control it? If someone gets the power to control the technology, he will rule everything. Then we will need to control the man who controls AI."* People's worry about the misuse of AI eventually depends upon the decisions of man. This shows another "us vs. them" view: those who have AI vs. those who do not.

3.5 Limitations

There are several limitations of this chapter. While carrying out this study, we used the term AI in a broad sense, although it could be interpreted in many ways depending on its capabilities and functions. In addition, as our participants were all metropolitan Koreans who (mainly actively) volunteered to participate, the result of this study may not be generalizable. We also did not relate this research to previous related events, such as DeepBlue's chess match and IBM Watson's Jeopardy win.

3.6 Summary

This chapter has attempted to understand people's fear of AI with a case study of the Google DeepMind Challenge Match. Through a qualitative study, we identified that people showed apprehension toward AI and cheered for their fellow human champion during the match. In addition, people anthropomorphized and alienated AI as an "other" who could do harm to human beings, and they often formed a confrontational relationship with AI. They also expressed concerns about the prevalence of AI in the future.

This chapter makes three contributions to the HCI community. First, we have investigated people's fear of AI from various perspectives, which can be utilized in various areas. Second, we have identified the confrontational "us vs. them" view between humans and AI, which is distinct from the existing view on computers. Third, we have stressed the importance of AI in the HCI field and suggested the concept of an expanded user interface and algorithmic experience.

Based on the results of this study on people perception of AI algorithms, the next chapter will focus on how people interpret and evaluate algorithm-based systems using AI.

4 HOW PEOPLE INTERPRET AND EVALUATE ALGORITHM-BASED SYSTEMS USING ARTIFICIAL INTELLIGENCE

While artificial intelligence algorithms are making remarkable progress, it is often difficult for users to interpret their results. To understand how various users reason about AI algorithm results, we designed AI Mirror, an interface that tells users the algorithmically predicted aesthetic scores of photographs.[3] We conducted a user study of the system with 18 participants, including AI/machine learning (ML) experts, photographers, and general public members. They performed tasks consisting of taking photos and reasoning about AI Mirror's prediction algorithm with think-aloud sessions, surveys, and interviews. The results showed the following: (1) Users understood the AI using their own group-specific expertise; (2) Users employed various strategies to close the gap between their judgments and AI predictions over time; (3) The difference between users' thoughts and AI predictions

[3] All uses of "we," "our," and "us" in this chapter refer to contributors of the study.

was negatively related with users' perceptions of the AI's interpretability and reasonability. We also discuss design considerations for both algorithms and user interfaces.

4.1 Motivation

The recent advances in artificial intelligence, specifically machine learning and deep learning (DL) [185], have been attracting more attention than ever from the academic and industrial fields. In various fields, such as computer vision ([186], [187]), speech recognition ([25], [26]), and natural language processing ([188], [189]), AI technology has already yielded results comparable to those of human experts ([173], [174], [189]). Beyond simply classifying objects or inferring values, the algorithms are evolving to generate new artifacts, such as pieces of writing [106] and artistic images ([190], [191]).

However, as AI algorithms have produced results in areas such as aesthetics in which people can take a subjective view, people can have difficulty in understanding their results, wondering how the algorithms work. Without expertise in AI/ML, it would be difficult for people to interpret and understand the results of AI algorithms. Moreover, AI algorithms sometimes do not fully explain their internal principles, which is sometimes referred to as the black box problem ([67]–[69]). In situations where people cannot accept the results of AI algorithms, if the transparency of an algorithm is not ensured, its users may lose confidence in the algorithm and not be immersed in it ([79], [85]). With this as a background, we aim to investigate how users reason about the results of an AI algorithm and discuss human-computer interaction/user experience considerations in the design of user interfaces with AI. First, we designed a research probe, AI Mirror, a user interface that tells users the algorithmically predicted aesthetic scores of the pictures they have taken or selected based on a deep neural network model (Figure 4-1). Then, we conducted a user study using both quantitative and qualitative methods. We recruited a total of 18 participants consisting of a well-balanced mix of AI/ML experts, photographers, and members of the general public. They performed a series of tasks consisting of taking photos using AI Mirror and reasoning about its algorithm with the think-aloud method and survey. In the questionnaire, we collected users' expected scores for their pictures and their interpretability and reasonability ratings for the AI's scores. We also conducted semi-structured interviews about how users experienced the system. The results from the study can be summarized as follows:

- According to their group (i.e., experts, photographers, general public), users showed different characteristics in reasoning about the AI algorithm. They understood the AI using their own group-specific expertise.
- The group of photographers was able to best interpret the AI's aesthetic scores and considered them reasonable. On the other hand, the AI/ML experts had difficulty interpreting them and considered them relatively unreasonable.
- Users employed various strategies to close the gap between their judgments and the AI's predictions over time.
- If there was a difference between users' thoughts and the AI's predictions, they had difficulty interpreting the AI's predictions and considering them reasonable.
- While interacting with the AI, users wanted to actively communicate with the AI.

Based on these findings, we discuss design considerations for AI-powered user interfaces that convey subjective results, such as aesthetic evaluations, to users. The main contributions of this work to the HCI community are as follows:

- This work yielded experimental results showing how the unique characteristics of users affect the process of inferring the outcomes of the AI in terms of group, strategy, and communication.
- These results have design implications for intelligent user interfaces that deliver a variety of interpretable results, which could be utilized by both the AI/ML and HCI communities.

4.2 AI Mirror

To address the research questions, we designed a research tool, AI Mirror.

4.2.1 Design Goal

In the design of the research tool, according to the research questions, we aimed to create an interface that allows users to interact with AI algorithms on a domain whose results can be interpreted in various ways. Therefore, we considered (1) using state-of-the-art neural network algorithms and (2) selecting a topic that allows users to produce their own artifacts and interpret the results of AI on them. Therefore, among the creative and open-ended domains, we selected aesthetics. We reviewed Augury [192], which evaluates a website's design by calculating the complexity and colorfulness of the website with a database of aesthetic preferences, and used the concept in the design process. Finally, we created an interface that can predict the aesthetic quality of photographs provided by users based on a state-of-the-art neural network algorithm and named it "AI Mirror" (Figure 4-1).

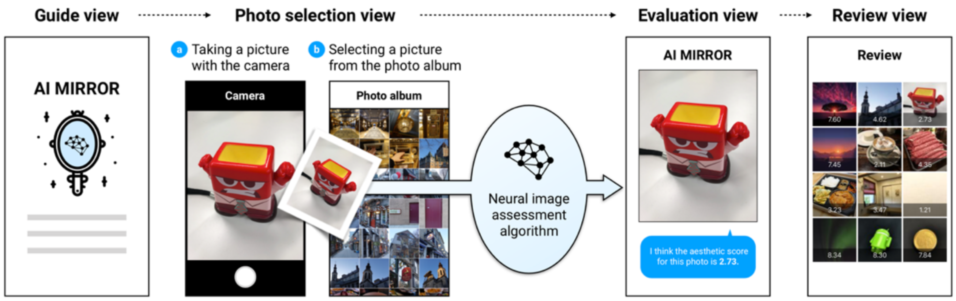


Figure 4-1. The overview of AI Mirror. We designed a research probe, "AI Mirror," a user interface that tells users the algorithmically predicted aesthetic scores of the pictures they have taken or selected based on a deep neural network model.

4.2.2 Image Assessment Algorithm

In the design of AI Mirror, we introduced Google's Neural Image Assessment (NIMA) [193], an AI algorithm to predict the aesthetic quality of images. This convolutional neural network (CNN) is trained to predict which images a typical user would rate as looking both technically good and aesthetically attractive. Instead of classifying images according to low/high scores or regressing to the mean score, the NIMA model produces a distribution of ratings for any given image on a scale of 1 to 10 [193]. Through the pilot study, we identified that as the mean scores of given images approximated the normal distribution, the scores concentrated on the average, and extreme values were rarely found. Since it was possible that users could not perceive the difference between the good and bad pictures, we performed a linear transformation of the normal distribution so that users fully utilized the algorithm in the experiment.

4.2.3 Design of User Interface

AI Mirror was developed as a web application that works on a mobile web browser and uses a camera and photo album. The user interface of AI Mirror is composed of four main views: guide view, photo selection view, evaluation view, and review view (Figure 4-1).

- **Guide view:** This screen is the first screen of the user interface and provides a basic explanation of the system with a simple concept image. A user can start using the system by entering his or her username.
- **Photo selection view:** On the next screen, the user can select a picture to be evaluated by the AI. In the process, the user can select one of two functions: (1) taking a picture with the camera app and (2) selecting a picture from the photo album. If the user chooses the former, the camera of the smartphone is activated so that the user can take a picture. If the user selects the latter, the photo album is activated so that the user can browse pictures and select a picture.
- **Evaluation view:** Right after taking or choosing a picture, AI Mirror shows the user the aesthetic score of the picture on a 10-point scale, stating in text, "I think the aesthetic score for this photo is 8.99."
- **Review view:** In this view, the user can view the pictures that have been evaluated by the AI so far. Photos are presented in a tile layout. If the user selects a photo, the photo is enlarged in the pop-up window. The aesthetic score is displayed at the bottom of the photo.

4.3 Study Design

To understand how users interact with the system, we designed a user study with a mixed-methods approach using both quantitative and qualitative methods. We strictly followed the user study design protocol reviewed and approved by the

Institutional Review Board of Seoul National University Institute.

Group	ID	Age	Sex	Characteristics
AI/ML expert	A1	37	M	CTO of video AI startup (author of CVPR paper)
	A2	35	M	CEO of video AI startup (author of CVPR paper)
	A3	32	M	CTO of sound AI startup (author of DCASE paper)
	A4	29	M	Researcher at IT research center (teaching ML/DL experience)
	A5	26	F	Researcher at IT research center (majoring in ML)
	A6	24	F	Researcher at IT company (AI field strategy)
Photo expert	P1	34	F	Photographer (10 years of field experience)
	P2	30	F	Amateur photographer (took camera education course)
	P3	34	M	Photographer (10 years of field experience)
	P4	31	F	Photographer (10 years of field experience)
	P5	30	M	Amateur photographer (11 years of field experience)
	P6	29	F	Photographer (majoring in fine arts)
No expertise	N1	36	F	Administrative worker
	N2	34	M	Researcher (urban planning)
	N3	25	F	Nursing teacher
	N4	32	F	Graduate school student (communication studies)
	N5	28	M	English teacher
	N6	28	M	Graduate school student (business)

Table 4-1. Participant information. (IDs: “A”=AI/ML expert, “P”=photography expert, “N”=no expertise.)

4.3.1 Participant Recruitment

In recruiting participants, we sought to balance the following three groups: AI/ML experts, photographers, and the general public. We set specific recruitment criteria for each group. First, the AI/ML experts group included only those who had majored in computer science-related areas, such as ML and DL, or had experience as specialists in related fields. The group of photographers included only professional photographers, people with training in photography, or non-professional photographers who had more than 10 years of photography experience. In particular, each group included only those without expertise in the area(s) of the other expert group(s). For the general public group, people who did not have these types of expertise were sought. We first posted a recruiting document on our institution's online community and then used the snowballing sampling method. We recruited a total of 18 participants, with the same number of participants for each group (Table 4-1). Some of the recruited AI experts ran AI-related startups, some had authored papers for top conferences in computer vision, while others were involved in related industries with relevant knowledge and expertise. The recruited photographers were people with more than 10 years of photography experience.

4.3.2 Experimental Settings

In the user study, we used a dedicated device, the iPhone X, as the main apparatus to control the experiment by providing the same conditions to all participants.

Since the experiment was done in the laboratory, it was necessary to prepare a variety of objects and additional material that participants could use to take pictures. Various objects of different colors and shapes (e.g., a yellow duck coin bank, a green android robot figure, a blue tissue box) were prepared so that users could combine various colors and attempt various compositions with them. In addition, we prepared backdrops so that users could keep the background clean and clear. We also prepared a small stand light. This setup allowed participants to freely take various photos. Meanwhile, to meet the users' various photo selection needs, we entered many pictures in the photo album of the experimental device beforehand. This photo album included a total of 80 photos, ranging from images that received 1 point from AI Mirror to images that received 8 points, with an equal number of high-scoring and low-scoring images.

4.3.3 Procedure

In the study, participants completed a series of tasks, including interacting with AI Mirror and reasoning about its algorithm with the think-aloud method and survey, and then took part in interviews. In a separate guidance document of AI Mirror before the experiment started, we provided the participants with a detailed explanation of the purpose and procedure of the experiment. Users were allowed to manipulate the system for a while to get used to it. On average, the experiments lasted about 60 minutes. All participant received a gift voucher worth $10 for their participation.

Task

The main task that participants were asked to perform in the experiment was to interact with AI Mirror and deduce the photo evaluation criteria of AI Mirror. Using AI Mirror, the participants took photos or selected photos from the photo album, and AI Mirror made aesthetic evaluations of the photos. There were no particular restrictions on the number of interaction trials or time.

Survey

In the questionnaire, we asked participants to answer three questions. The first asked participants about the expected score for the aesthetic evaluation of the photos they had taken or selected. Just prior to AI Mirror's aesthetic evaluation of the picture, they were asked to give their own score on a 10-point scale. The second question asked participants whether AI Mirror's aesthetic evaluation scores were interpretable for them. Participants rated this on a 5-point Likert scale (1-strongly disagree, 5-strongly agree). The third question asked participants whether AI

Mirror's aesthetic evaluation scores were reasonable. Likewise, participants rated this on a 5-point Likert scale (1-strongly disagree, 5-strongly agree). In addition to these three items–expected score, interpretability, and reasonability–we measured difference between the participant's aesthetic score (expected score) of the picture and that of AI Mirror. We also collected the trials over time along with these values to capture temporal changes in values for each participant.

Think-aloud Session and Interview

We conducted a qualitative study using the think-aloud method [194] and semi-structured interviews to gain a deeper and more detailed understanding of users' thoughts. While performing the tasks, the participants could freely express their thoughts about the tasks in real time. We 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 AI Mirror, especially focusing on its interpretability and reasonability. All the interviews were audio recorded.

4.3.4 Analysis Methods

From the study, we were able to gather two kinds of data: quantitative data from the surveys and the system logs and qualitative data from the think-aloud sessions and interviews.

Quantitative Analysis

In the quantitative analysis, we ran statistical analysis of the six types of data collected from each participant: group, trials over time, expected score, difference, interpretability, and reasonability. Based on this, we tried to determine the significant relation or difference between these variables. We used panel analysis, as it is specialized for analyzing "multidimensional" data collected "over time," and over the same individuals, which matched the data we gained from our experiment exactly. Besides, it can run the regression model of each variable of multidimensional data, so it can provide more concise and comprehensive results than an ANOVA, which produces results in an aggregated way without considering the time effect. Moreover, as we recruited users by group, we assumed that the unobserved variables were uncorrelated with all the observed variables and accordingly used a random effects model.

Qualitative Analysis

The qualitative data from the thinkaloud sessions and post-hoc interviews were transcribed, and analyzed using thematic analysis [195]. In the process, we used Reframer [183], a qualitative research software tool provided by Optimal Workshop. To organize and conceptualize the main themes, three researchers used line-by-line open coding. Through a collaborative, iterative process, we revised these categories to agreement and then used axial coding to extract the relationships between the themes.

4.4 Result 1: Quantitative Analysis

In the case of the statistical analysis results, since the number of responses was relatively small, the emphasis was on understanding the basic relationship or tendency between variables. We set difference, interpretability, and reasonability as dependent variables and sequentially investigated the effect of trials over time, group, and other independent variables on each dependent variable by conducting a panel data analysis. The average trials over time was 14.22, with a maximum of 27 and a minimum of 10 (SD=4.48). (Statistically significant results are reported as follows: p<0.001(***), 518 p<0.01(**), p<0.05(*).)

4.4.1 Difference

First, in the analysis on difference, based on the results shown in Table 4-2, we observed that trials over time had a significant 524 influence on difference (t-value=-2.66, p<0.01**). That is, as 525 trials over time increased, difference significantly decreased, 526 which means that as users continued to interact with the AI, 527 they reduced the difference between their expected scores 528 and the AI's scores.

Variable	*coefficient*	*Std. Error*	*t-value*	*p-value*	
(Intercept)	2.611	0.238	10.970	<0.001	***
Trials over time	-0.049	0.018	-2.656	0.008	**
General public	-0.203	0.244	-0.830	0.407	
Photographer	-0.373	0.236	-1.582	0.115	

$R^2 = 0.035,\ Adj.R^2 = 0.023,\ F(3, 252) = 3.02,\ p = 0.03^*$

Table 4-2. Results of panel data analysis of difference.

In addition, although we did not identify any significant effects of group, we found that there were slight differences in difference between user groups. Surprisingly, AI/ML experts showed the biggest difference from the AI (Mean: 2.14), followed by the general group (Mean: 2.07), and finally the photographers (Mean: 1.84).

4.4.2 Interpretability

In the analysis on interpretability, based on the results shown in Table 4-3, we observed that difference had a significant influence on interpretability (t-value=-7.63, p<0.001***). That is, as difference increased, interpretability significantly decreased, which means that users had difficulty interpreting AI scores when there was a big difference between their evaluations and those of the AI.

Variable	*coefficient*	*Std. Error*	*t-value*	*p-value*	
(Intercept)	3.034	0.251	12.090	<0.001	***
Trials over time	0.017	0.014	1.280	0.202	
Difference	-0.334	0.044	-7.615	<0.001	***
General public	0.488	0.274	1.782	0.076	
Photographer	1.323	0.272	4.860	<0.001	***

$R^2 = 0.280,\ Adj.R^2 = 0.268,\ F(4, 251) = 24.36,\ p < 0.001^{***}$

Table 4-3. Panel data analysis of interpretability.

In addition, we identified that group had a significant effect on interpretability, especially for photographers (tvalue=4.86, p<0.001***). The photographer group (Mean: 3.90 out of 5) showed a higher level of interpretation of the aesthetic scores evaluated by the AI compared to the AI/ML experts (Mean: 2.44). Although it was not a significant difference, the general public (Mean: 2.96) also showed a higher level of interpretation than the AI/ML experts. Meanwhile, trials over time also showed a slightly positive effect on interpretability, but it was not significant either.

4.4.3 Reasonability

In the analysis of reasonability, based on the results shown in Table 4-4, we observed that difference had a significant influence on reasonability (t-value=-12.02, p<0.001***). That is, as difference increased, reasonability significantly decreased, which means that users did not think the AI score was reasonable when there was a difference between their thoughts and those of the AI.

Variable	*Coefficient*	*Std. Error*	*t-value*	*p-value*	
(Intercept)	3.797	0.268	14.156	<0.001	***
Trials over time	-0.024	0.013	-1.884	0.061	
Difference	-0.485	0.040	-12.021	<0.001	***
General public	0.277	0.318	0.872	0.384	
Photographer	1.057	0.317	3.334	<0.001	***

$R^2 = 0.410,\ Adj.R^2 = 0.401,\ F(4, 251) = 43.63,\ p < 0.001^{***}$

Table 4-4. Results of panel data analysis of reasonability.

We also identified that group had a significant effect on reasonability in the case of photographers (t-value=3.33, p<0.001***). The photographers gave higher reasonability scores (Mean: 3.74 out of 5) than AI/ML experts did (Mean: 2.43). Although it was not a significant difference, the general public also gave higher reasonability scores (Mean: 2.92) than AI/ML experts did.

On the other hand, trials over time slightly lowered the reasonability, but it did not show any significant effect.

To summarize the results of the quantitative analysis, first, we partially identified that users in different groups showed differences in the process of interacting with the AI. The group of photographers showed the highest interpretability and reasonability among the three groups, with AI experts having the lowest. Second, users were able to narrow the gap between their evaluation scores and those of the AI as they continually interacted with AI. Third, the difference between users' thoughts and the AI's predictions lowered both the perceived interpretability and reasonability of the AI.

4.5 Result 2: Qualitative Analysis

In the qualitative analysis results, we focused on finding detailed features not revealed in the statistical analysis. Here, we report the characteristics of each group, the strategies users showed in the reasoning process, and the factors users considered important in their interpretability and reasonability evaluations.

4.5.1 People Understand AI Based on What They Know

Through the qualitative analysis, we identified that while interacting with AI Mirror, the participants showed distinctive characteristics according to their group. In particular, we observed that the vocabulary they used reflected their expertise. Each participant also attempted a distinct approach in the process of reasoning.

First, while interpreting the AI's results, AI/ML experts commonly used words that reflected specialized knowledge of ML and DL, such as "algorithm," "dataset," "training," "model," "black box," "pixel," "classification," and "feature," which were never mentioned by the other groups. For example, A1 said, "*It's like evaluating a model. It's like putting unseen data into the test set and seeing if it works or not.*" A6 said, "*There may be some problems with the learning process and the database. It depends on if it was based on social media data, like Instagram. You know, colorful photos usually get a lot of likes.*" A5 said, "*And I think we should open the black box if possible and make it a white box.*" People in this group also used their AI/ML expertise in inferring the AI's criteria. For instance, A4 said, "*I think the boundaries of this object are not clear. It seems the algorithm is not detecting this object well. Normally vision technology needs to know the boundaries of objects.*" A4 then edited the photo of a white egg with a white background by drawing the outline of the egg. However, unexpectedly, AI/ML experts did not receive high scores overall and eventually said they were not confident in their understanding of the AI's standards. In browsing the pictures that he took on the review view of AI Mirror, A2 said, "*I do not know why this score is high .... and this is too low a score.*" A3 concluded the experiment by saying, "*The experiment itself is interesting .... but my pictures scored much lower than I expected.*"

Secondly, the photographers interpreted and inferred the results of the AI using their expertise in photography. They often mentioned important elements of photography, such as "light," "color," "moment," "composition," and "distance," and camera controls, such as "focus," "aperture," and "lens." For example, P02 said, "*This picture has a low depth of field, so I think it will get a higher score than the previous one.*" P04 said, "*The composition of this picture follows the rule of thirds well.*" P01 said, "*The light is concentrated toward the black background, so this doll is too bright. So I'm going to adjust the light by touching it on the camera app screen. I often do this. This makes the background darker and darker.*" When choosing images in photo albums, people in this group also picked the pictures that seemed likely to get high scores from the AI, taking advantage of their expertise. Taking the viewpoints of the photographer of the picture that he picked the album, P3 said, "*This is definitely a good picture. The photographer must be proud of such a beautiful picture. He must have waited for this moment.*" Emphasizing the importance of photoshopography, P6 also said, "*I think this photographer did photoshopography on this image to express the colors of various spectrums.*" P5 also assessed the quality of the photo selected from the album by reasoning about the weather at the time of the picture. Overall, the group of photographers took or picked high-scoring pictures, often showing expected scores similar to those of the AI. When he realized that his score was almost identical to the AI's score, A1 was surprised, and he said, "It was totally creepy. I think this AI is in my head."

Third, the general public group took pictures in the way that they typically take pictures without specific professional knowledge. They mainly took pictures of their favorite objects from what we had prepared for the experiment or chose pictures of famous landmarks or beautiful landscapes from the photo album, believing that the AI would appreciate these pictures. For example, N1 said, "*This [a yellow duck coin bank] is really cute. I'll take this." N3 said, "I'm just looking for a picture that looks pretty. This picture is pretty. Everything in the picture looks pretty. It looks like a house from a fairy tale."* Looking at the photos in the photo album, N6 said, "*And I think I've seen this quite a few times. It's the Louvre Museum,"* and picked the photo. However, they did not fully comprehend the scores of the AI. N2 said, "*I think there must be a certain standard ... But I cannot quite grasp it. I do not know if it's really aesthetic judgment."* N5 said, "*I think the AI has another evaluation criterion. The AI does not think this picture is pretty."* N4 even complained, saying, "*I think it'll give a very high score to this picture. Actually, I do not think this picture is pretty. However, the AI has always been so contrary to me, so this picture will have a high score."*

4.5.2 People Reduce Difference Using Various Strategies

Next, we identified that as they continued to interact with AI, users adopted their own personal strategies to infer the AI's principles of evaluation. They used approaches that involved making subtle changes to various picture elements, and they extended their ideas through various examples.

First, when users took pictures, they tried to experiment with the AI by making slight changes to the pictures. They changed the background color of an object or the composition of the same object. They sometimes added objects one at a time and looked at the AI's reactions as different colors were added. P1 said, "*The next thing I wanted to do was keep the background white and add another colored object. I wanted to see how the score changed when I did that.*" N6 said, "*This time, I'll take the same background and object from a distance. It makes the object look small in the picture. I have to change only one element .... Oh .... 4.75 points. Size does not matter. Now I understand more.*" N3 said, "*And this time, I'll take this same object on a yellow background. I think if the background is yellow, somehow it looks like the background will be more in focus than the object, so the score will be lower. (Score: 2.19) Now I know more. I think the AI and I have a similar idea.*" Through this process, most of the users found that the AI gave high scores (8 points on a 10-point scale) when one bright object in the photo stood out against a black background. Photographers tried these kinds of pictures relatively earlier in their trials than the other participants did.

Second, some users even used the editing features of the iPhone photo app to actively modify the photos they took or the photos they picked from the album and asked the AI to evaluate the modified photos. A4 described, "*I'll edit this photo of the macaroons. Let me get rid of the color. The reason for doing this is to know if the color is important or not. The color has gone and I think it will be lower than 7.22.*" P5 said, "*I'll crop the photo. Let's move the object to the center. I just changed the position of the object. I think this picture will be rated at ab out 8 points. (Score release) Uh-oh (...) The score is lowered (...) The composition is not a problem.*" In this way, participants developed a better understanding of the characteristics of AI by creating slightly different versions of the photographs.

They all stated that this process enabled them to better understand and experience AI principles.

Third, participants transferred their speculations about how AI works and applied them to different cases. They continued their testing of the aesthetic evaluation criteria of AI by using similar examples. They wanted to know whether the criteria they had grasped could be applied to other photos with similar characteristics (e.g., composition, contrast, color) but different objects from the photos they had taken or identified so far. N2 explained, "*I'll pick this crab signboard picture. I think this is going to have a score similar to the picture I took before. What was the score of the photo with the white background and the red toy?*" A5 described, "*I'll pick a photo with a variety of objects and a central object in it. That's the standard I've figured out so far.*" After getting a high score for a photo with a black background, P1 said, "*Then, this time, I'll pick a picture with a black background similar to the last one.*" Through this process, users were able to confirm whether their criteria were correct and narrow the gap between their thoughts and those of the AI.

Lastly, we identified that participants tried to find new standards that they had not seen so far by choosing completely different pictures from the photo album. After finding a certain way to get a high score, some participants additionally attempted to look at new types of photos. P4 described, "*I'll take a look at the kinds of pictures I have not seen before. I'll try this .... I have to review the various pictures to see what it likes and what it does not like.*" P3 said, "*I'll try it again. Um .... I'll take this. This is just a pattern that I have not picked up so far.*" N3 also remarked, "*I just want to try something I have not tried yet. I think it likes pictures of things that are distinct and colored. But from now on, I do not think I should choose things like that.*" Through this process, users were able to find new and unexpected criteria, such as "*a preference for photographs with repetitive* *patterns."*

Overall, based on these various strategies, while interacting with AI Mirror, participants were able to understand its scoring system and narrow the gap between its scores and their scores.

4.5.3 People Want to Actively Communicate with AI

Finally, regarding users' perceptions on the interpretability and reasonability of the AI algorithm's aesthetic evaluations, the participants wanted to actively communicate with AI Mirror in the experiment.

During the think-aloud sessions and interviews, regardless of their group, users recounted interacting with the AI as a positive experience. Most participants described the process as interesting, fun, and enjoyable. In particular, while reasoning about the criteria of AI Mirror's aesthetic evaluations, participants felt curious about the principles of AI and wanted to know about it. P4 described, "*It was fun and interesting. It got me thinking. It stimulated my curiosity."* N1 expressed, "*It was fun to find out the criteria it used to rate them. It was just an experiment, but I was really curious."* Participants were also delighted when the difference between the AI score and their expected score was not that large. They were even more delighted when the AI gave a higher score than they expected. They expressed that it was as if AI Mirror had read their thoughts and that they felt like they were being recognized and praised by the AI. N3 said, "*Later, I felt good about the AI, because it was well aware of the points I had intended and appreciated my effort."* N5 said, "*I feel good because I got a high score. I feel like I'm being praised by the AI."* Some participants even asked us to send the URL link to the AI Mirror webpage at the end of the experiment. They wanted to get ratings on their personal smartphone photos and to interact more with the AI.

Nonetheless, most participants stated that they also felt negative emotions during the interaction. When their expected scores differed significantly from those of the AI, especially when they were rated very poorly by the AI, participants felt embarrassed, unhappy, and frustrated. For example, N5 described, "*Oh .... I feel terrible. This score is lower than the previous one. I took more care with it. I feel worse as my score drops. It's pretty unpleasant.*" Participants told us that they could not understand why the AI's scores were lower than they thought and that they had difficulty interpreting the results. N6 said, "*I'm so frustrated because I do not know why my score is so low.*" A2 complained, saying, "*This is really low, but I do not know why .... This is too low .... I know this is an ugly picture. But is there a big difference from the photo I took earlier?*" Some even expressed that they could not understand the AI and regarded this interaction as meaningless. P6 said, "*Maybe it just thinks so. It is just being like that. I do not want to deduce anything. My overall level of interest is .... pretty low. I have no understanding of it.*" These unpleasant experiences also reduced participants' trust in the system as well as their confidence that they could take pictures well. P2 said, "*I think this picture will get 6 points. I have lost my confidence. I think my expectations for my picture have been lowered too.*"

In such a situation, the absence of communication between users and AI can be considered the main cause of the negative emotions of users. During the interviews, participants uniformly expressed a desire to communicate with the AI. They wanted the AI to explain not only the calculated scores but also the detailed reasons. N6 said, "*I wanted to know the elements of the scores. I think it would be better if it could tell me more specifically.*" P6 expressed, "*It would be much better if it could tell me why it came up with this score. Then I could take better pictures.*" Furthermore, users wanted to let the AI know their thoughts. P4 said, "*I want to let the AI know this is not as good a picture as it thinks.*" A6 described, "*I had a* *lot of disagreements with the AI. I think it would be nice if it could learn my thoughts on the points on which we disagreed. It is my app, and it has to appreciate what I think is beautiful."* Some users said that in this one-sided relationship, even though they could interpret the evaluations of the AI, they could not see them as reasonable. P1 said, "*The weather in the photos is not that sunny, but I like the cloudy weather. I'm sure that AI Mirror will rate this picture too low. It only likes those pictures that are high contrast. I can clearly see why the score is low, but I cannot say that it is reasonable."*

The various emotions that the participants experienced during the user study and their strong desire for communication for improved interpretability and reasonability suggest that in the design of user interfaces with AI (namely, algorithms), additional and careful discussion is needed.

4.6 Limitations

There are several limitations of this chapter. First, in the questionnaire analysis, the explanatory power of the model was relatively low, although several significant relationships and differences were found. The reason seems to be that the numbers of participants and trials were too small due to limitations of the experimental environment. Second, we limited the experimental context to a special situation (i.e., the aesthetic domain) and did not reflect the diverse capabilities of AI technology. Third, we assumed a one-sided relationship with AI and did not measure the effect of users' various communications with AI.

4.7 Conclusion

In this chapter, we investigated how users reason about the results of an AI algorithm, mainly focusing on their interpretability and reasonability issues. We designed AI Mirror, an interface that tells users the algorithmically predicted aesthetic scores of pictures that the users have taken or selected. We designed and conducted a user study employing both quantitative and qualitative methods with AI/ML experts, photographers, and the general public. Through the study, we identified that (1) users understood the AI using their own group-specific expertise, (2) users reduced the thought gap with the AI by interacting with it through various strategies, and (3) the difference between users and the AI had a negative effect on interpretability and reasonability. Finally, based on these findings, we suggested design implications for user interfaces where AI algorithms can provide users with subjective information. We hope that this work will serve as a step toward a more productive and inclusive understanding of users in relation to AI interfaces and algorithm design.

Based on this study of the process of inference by humans about artificial intelligence algorithms, the next chapter will discuss further on the process of artificial intelligence and user collaboration.

5 HOW PEOPLE BUILD SEQUENTIAL ACTIONS WITH ALGORITHM-BASED SYSTEMS USING ARTIFICIAL INTELLIGENCE

Recent advances in artificial intelligence 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.[4] 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.

[4] This chapter has adapted, updated, and rewritten content from a paper at CHI 2018 [232]. All uses of "we," "our," and "us" in this chapter refer to coauthors of the paper.

(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.

5.1 Motivation

It is the age of artificial intelligence, and recent advances in deep learning have yielded AI with capabilities comparable to those of humans in various fields ([173]–[175]). Many interactions have been introduced based on this technology, such as voice user interfaces and autopilots of self-driving cars.

AI is expected to become increasingly prevalent in numerous areas ([196]–[199]). It will not only assist humans in repetitive and arduous tasks but also take on complex and elaborative works ([7], [31], [37]). Moreover, while humans can guide AI, AI can also guide humans ([200], [201]). They can even work together to produce reasonable results in various creative tasks, including writing, drawing, and problem solving ([106], [202], [203]).

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], [118], [121], [204]). Many HCI researchers have conducted user studies on various AI interfaces ([205]–[207]), and the concept of *algorithmic experience* has been suggested as a new perspective from which to view the user experience of AI interfaces [77]. 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 AIbased 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 5-1).

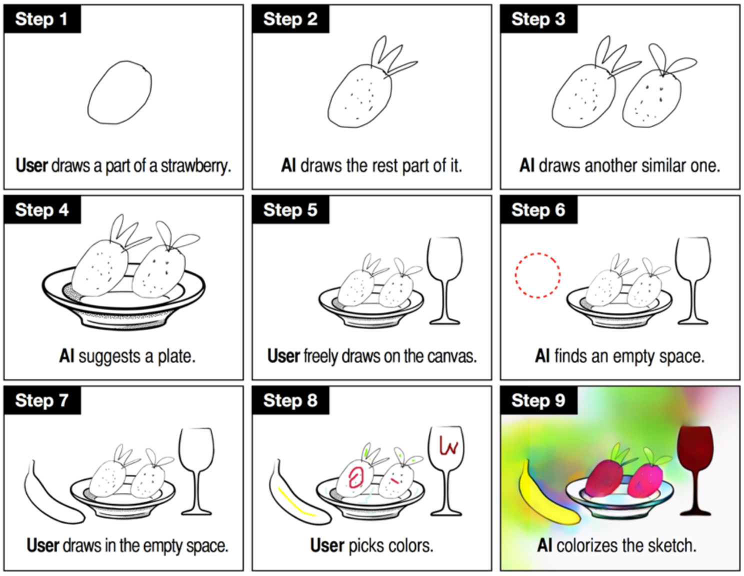


Figure 5-1. Drawing using DuetDraw in the Lead mode. With DuetDraw, users and AI can collaboratively draw pictures.

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.

5.2 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 5-1). The tool runs on a Chrome browser using P5.js [208], a JavaScript library for sketchbook software. For AI-based functions, DuetDraw uses the open source code of Google's Sketch-RNN [115] and PaintsChainer [116]. 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.

5.2.1 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 5-1). It is based on Google's Sketch-RNN [115].
- ***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 5-1). The object is drawn to the right of the existing object and at the same scale. It is also based on Sketch-RNN [115].
- ***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 5-1).

- ***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 5-1).

- ***Colorizing sketches with recommended colors:*** This function enables the AI to colorize sketches based on a user's 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 [116], a CNN-based line drawing colorizer (Step 9 in Figure 5-1).

5.2.2 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 5-2).

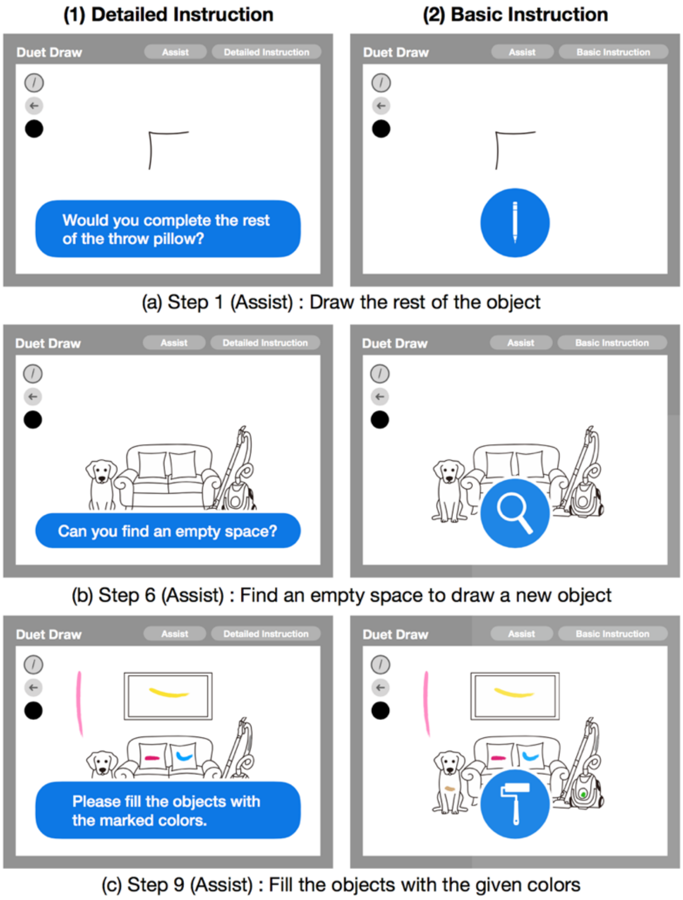


Figure 5-2. Examples of two different communication styles of DuetDraw.

5.3 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.

5.3.1 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.

5.3.2 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) *LeadBasic*, (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.

5.3.3 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 5-1) in which the AI and the user drew a picture together.

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 the emptiest space on the canvas.
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 5-1. Scenario of drawing a picture with DuetDraw

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 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.

5.3.4 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 ([209], [210]) 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 ([95], [211], [212]): *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.

5.3.5 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 thinkaloud method while performing the tasks [213], 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 [214], 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.

5.3.6 Analysis Methods

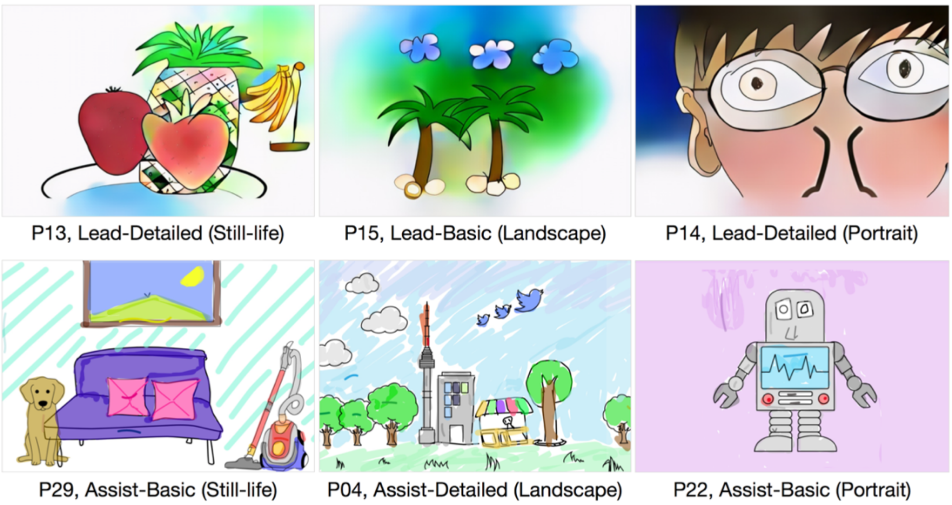
From the study, we were able to gather two kinds of data: quantitative data from the surveys and qualitative data from the think-aloud sessions and interviews. We conducted quantitative analysis for the former and qualitative analysis for the latter, which are described in detail below.

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 repeatedmeasures 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 [215]. 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 [183], 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.)



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 5-3. Pictures drawn by participants in experiment.

5.4 Result 1: Quantitative Analysis

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 5-3). The results of the analysis are as follows.

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 5-4).

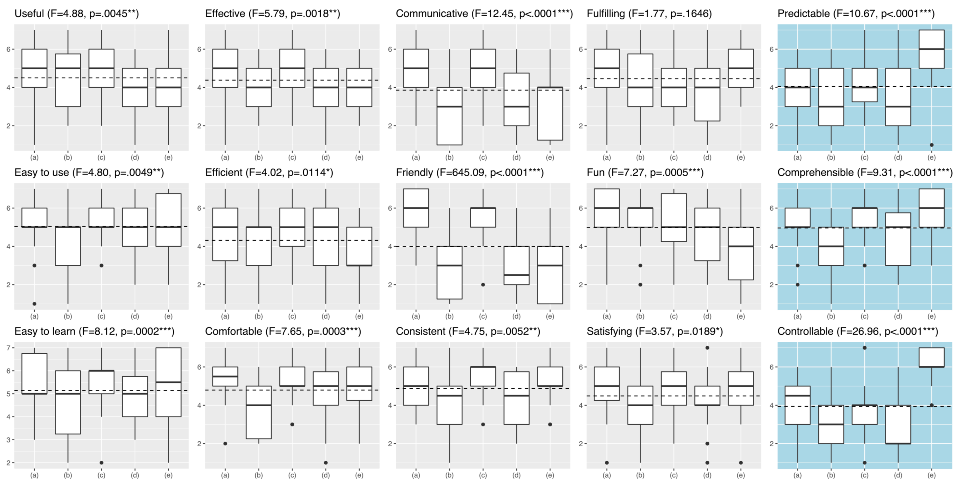


Figure 5-4. Box plots of user ratings of each item according to each condition and result of one-way repeated-measures ANOVA. Except for fulfilling, all items showed significant differences. ((a) Lead-Detailed, (b) Lead-Basic, (c) Assist-Detailed, (d) Assist-Basic, (e) no-AI, $F(4, 26)$. The dotted lines 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 $p < 0.001$***, $p < 0.01$**, $p < 0.05$*)

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.

5.4.1 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.

	Comparison 1			Comparison 2			Comparison 3		
	(a) *Lead-Detailed* - (b) *Lead-Basic*			(c) *Assist-Detailed* - (d) *Assist-Basic*			(c) *Assist-Detailed* - (d) *Assist-Basic*		
item	**diff.**	**t**	**p**	**diff.**	**t**	**p**	**diff.**	**t**	**p**
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	0.4	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 5-2. 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.

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 5-2, 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 5-2): *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.

5.4.2 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 5-2): *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.

5.4.3 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 5-3, 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 5-3). 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 5-3). This shows that the interaction with AI can bring fun and excitement to the user as well as enhance basic usability.

	Comparison 5			**Comparison 6**			**Comparison 7**			**Comparison 8**		
	(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	2.819	0.0441*	0.17	0.486	0.9885
effective	1.13	3.51	0.0057**	0.10	0.309	0.9980	1.10	3.404	0.0080**	0.07	0.206	0.9996
efficient	1.30	3.54	0.0051**	0.64	1.726	0.4224	1.24	3.361	0.0091**	0.74	1.998	0.2734
fun	1.97	5.95	<.0001***	1.77	5.345	<.0001***	1.44	4.336	0.0003***	1.04	3.126	0.0187**
predictable	-2.20	-6.16	<.0001***	-2.57	-7.184	<.0001***	-1.50	-4.199	0.0005***	-2.34	-6.531	<.0001***
comprehensible	-0.53	-1.90	0.3252	-1.70	-6.044	<.0001***	-0.23	-0.830	0.9209	-1.23	-4.385	0.0002***
controllable	-1.94	-5.49	<.0001***	-3.07	-8.713	<.0001***	-2.47	-7.008	<.0001***	-3.17	-8.997	<.0001***

Table 5-3. 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.

5.4.4 No-AI is more Predictable, Comprehensible, and Controllable

However, as pointed out in previous studies ([95], [211], [212]), the treatment conditions recorded lower scores for the *predictable, 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 5-3). For *controllable*, all four treatment conditions recorded significantly lower scores than the control condition (Comparisons 4–7 in Table 5-3). In the case of *comprehensible*, when the communication mode was *Basic*, the treatment conditions showed a significant difference (Comparisons 5, 7 in Table 5-3).

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 5-2). 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 5-2). We could identify the same tendency in all other cases, even if this was not to a significant degree.

5.4.5 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 [216].

5.5 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 nonhuman characteristics. Finally, they reported that drawing with AI was a positive experience that they had never had before.

5.5.1 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 preset 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.*"

5.5.2 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.*"

5.5.3 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 [64]. 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-humanlike 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."*

5.5.4 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.*"

5.6 Limitations

We have identified three limitations of this chapter. 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.

5.7 Conclusion

This chapter 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.

Over the past three chapters, this thesis has stepped through an understanding of how people interact with artificial intelligence algorithms. Based on the results, in the next chapter, a case study of a practical application working on algorithms and its user experience will be presented.

6 HOW PEOPLE USE A PRACTICAL APPLICATION OF AN ALGORITHM-BASED SYSTEM USIGN ARTIFICIAL INTELLIGENCE

Automated journalism refers to the generation of news articles using computer programs. Although it is widely used in practice, its user experience and interface design remain largely unexplored. To understand the user reception of an automated news system, we designed NewsRobot, a research prototype that automatically generates news on major events of the PyeongChang 2018 Winter Olympic Games.[5] It produced six types of news by combining two kinds of content and three styles. A total of 30 users participated in using NewsRobot, completing surveys and interviews on their experience. Our findings were as follows: (1) Users preferred selective news yet considered it less credible, (2) more presentation elements were appreciated but only if their quality was assured, and (3) NewsRobot

[5] All uses of "we," "our," and "us" in this chapter refer to contributors of the study.

was considered factual and accurate yet shallow in depth. Based on our findings, we discuss implications for designing automated journalism user interfaces.

6.1 Motivation

Automated journalism (also known as robot journalism) refers to the generation of news articles using algorithmically designed computer programs ([126]–[128]). They usually collect and process data and integrate it in a predesigned article structure, producing news articles [128]. Unlike human journalists, the programs can create articles on a large scale quickly, cost-effectively, and even accurately.

Many companies, such as Narrative Science and Automated Insights, are already producing and providing news articles based on this technology ([129], [130]), and traditional media companies, like The Los Angeles Times and Thomson Reuters, are also distributing automatically generated news articles [131]. The technology is also used in a variety of areas, such as sports event highlights, weather forecasts, and disaster and election reports ([128], [140], [217]).

Surprisingly, while the importance of automated journalism is often noted ([148], [149]), the topic is rarely studied in the field of human–computer interaction. Although not directly addressing automated journalism, recent discussions on the transparency and fairness of algorithms in automated systems show that it is important to not only focus on the technical design of the algorithm but also closely observe the user experience of these systems ([81], [84]). This calls for the active involvement of designers and HCI researchers in the design of automated news generation systems.

In this chapter, we aim to explore the user experience of automated journalism and discuss design implications for automated news generation systems based on the findings. In particular, we focus on the issues of content and style. Both are crucial factors in automated journalism as well as HCI, in that the former is effective when delivering a large amount of content to many users, and the latter could be a means of delivering content to users more effectively.

To assess the user experience of an automated news generation system, we designed a research prototype, NewsRobot (Figure 6-1). It automatically generates a series of summary news articles of the PyeongChang 2018 Winter Olympic Games in real time. It collects data on the results of major events and players' information from the official website [218], processes the data, and inputs it into a predesigned article structure. We designed the system to generate news with two different types of content (general/personalized) in three different styles (text/slide/video). By combining the two factors, it can produce six different types of news articles for every event.

We conducted a user study of NewsRobot with both quantitative and qualitative approaches. We asked 30 recruited participants to watch Olympic Games races on TV and then showed them six types of news articles per game. They then filled out questionnaires on each news article and took part in semi-structured interviews. The results of the study can be summarized in the following three points:

- **Content:** While users preferred selective news to general news, they considered selective news less credible than general news.
- **Style:** As more news presentation elements were added, users' preference increased. People liked video news the most, followed by slide news and then text news. On the other hand, in terms of quality, users rated slide news as more clear and concise than video and text news.
- **Overall assessment:** While people were satisfied with NewsRobot's accuracy, objectivity, personalization function, and various presentation elements, they

found the articles were sometimes dull, repetitive, and out of context.

Based on these findings, we discuss the design implications for user interfaces for algorithm-based automated news generation systems and the responsible role of designers in automatic news algorithm and interface design.

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

- NewsRobot, an interface of a real-time automated news generation system that produces multiple news articles considering content and style
- The results of the user study with quantitative and qualitative approaches, investigating the user experience of an automated news generation system, and new discoveries on various aspects of this experience
- Design implications for automated news generation systems, stressing the important role of system designers

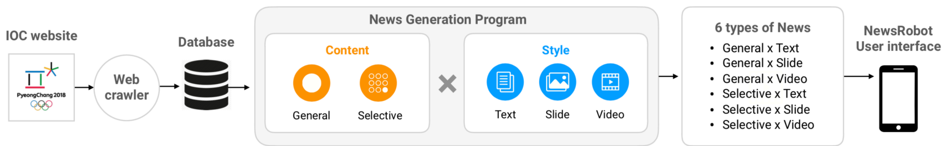


Figure 6-1. The overview of NewsRobot. To assess the user experience of an automated news generation system, we designed a research prototype, NewsRobot. It automatically generates a series of summary news of the PyeongChang 2018 Winter Olympic Games. It collects data on the results of major events and players' information from the official website, processes the data, and inputs it into the pre-programmed article structure. We designed the system to generate news with two different types of content.

6.2 News Robot

In this section, we describe NewsRobot, the research prototype of this study. In the design of NewsRobot, we aimed to create a system that generates real-time news about an actual sports event so that users could immerse themselves in the experiment.

6.2.1 Selecting Main Event and Data Source

As the first step, we selected the main event that NewsRobot would report about and its relevant data source. Since we aimed to create a variety of news articles in real time, considering major sports events, we chose the PyeongChang 2018 Winter Olympic Games.[6]

During the PyeongChang Olympic Games, the International Olympic Committee updated the results of all matches on the official website in real time [218]. The committee prepared separate pages for every match, providing not just the basic information of the match, such as location and date, but also all participants' intermediate records and ranks, differences from other participants, speeds, and finish records in real time. The committee also created separate pages for all athletes participating in the Olympics and posted their personal information, such as nationality, birthdate, age, gender, event and rank, and even biographical information. Information on players' previous Olympic and world championships, injuries, family relationships, idols, mottos, and even nicknames was included in the biographical information. By building a crawler of both pages with Python programming, we were able to collect both the real-time data of the results of each match and athlete information.

It was necessary to create articles in the same format for the controlled

[6] It was an international winter multi-sport event held between February 9 and 25, 2018 in PyeongChang, South Korea. It featured 102 events over 15 disciplines in seven sports, and 2,914 athletes from 92 countries competed in the games.

experiment. We limited main matches to races where multiple players competed for the record so that NewsRobot could create articles for every player and compare records between them according to their intermediate records. Finally, we chose three games: (1) Speed Skating Men's 1,000 Meters (February 14, 2018), (2) Alpine Skiing Women's Downhill (February 21), and (3) Short Track Speed Skating Women's 1,000-Meter Final (February 22).

Component		General news	Selective news
Head	Headline	Olympic {Alpine Skiing Men's Downhill} results 2018: {{Norway} wins gold and silver}; {Switzerland} wins bronze	{Beat Feuz} of {Switzerland} wins {Alpine Skiing Men's Downhill} {bronze}
Lead	Most important information (5W1H)	{Aksel Lund Svindal} of {Norway} won the gold medal with his {1: 40.25} run in the {Alpine Skiing Men's Downhill} event at the {Jeongseon Alpine Center} on {February 15}.	{Beat Feuz} of {Switzerland} {won the bronze medal} with his {1: 40.43} run in the {Alpine Skiing Men's Downhill} event at the {Jeongseon Alpine Center} on {February 15}.
Body	Supporting details	He passed the main intermediate measurement points {9th-9th-8th-2nd-2nd}, passing the finish line first with {1: 40.25}. This record is {0.12} seconds faster than that of the silver medal player.	He passed the main intermediate measurement points {18th-3rd-3rd-3th-3rd}, passing the finish line {third} with {1: 40.43}. This record is {0.18} seconds slower than that of the gold medal player.
	Background details	{His most recent record was fourth place in the same event at the Sochi Olympic Games 2014.}	{His most recent record was first place in the same event at the Alpine World Ski Championships St. Moritz 2017.}
	General details	Meanwhile, {Kjetil Jansrud} of {Norway} won the silver medal with his {1: 40.37} run. {Beat Feuz} of {Switzerland} won the bronze medal with his {1: 40.43 run}.	Meanwhile, {Aksel Lund Svindal} of {Norway} won the gold medal, and {Kjetil Jansrud} of {Norway} won the silver medal.

Table 6-1. General news vs. selective news. The braces are interlocked with the database, so they are tailored to each news article. The original version is written in Korean, and this table contains the English-translated version.

6.2.2 Designing News Article Structure

After the main events and data were defined, we designed a news article template that could cover all three races. We made the template following the inverted pyramid structure ([219]–[221]), the most common method for writing news stories. In the inverted pyramid, the widest part at the top represents the most important information, while the tapering lower portion illustrates that other material should follow in order of diminishing importance. We placed the headline in the first sentence of the template, followed by the lead sentence and then the body sentences. The lead sentence summarizes the most important information according to 5W1H (who, what, where, when, why, and how) [222]. The body sentences are then composed of supporting details, background details, and general details (Table 6-1).

6.2.3 Content and Style

In addition to the template article structure, we considered two main factors, content and style.

Content

We designed NewsRobot to create two different types of news: general news and selective news. The former is provided to all users equally, while the latter is customized to the users' interests and is provided differently for each user (Table 6-1).

- **General news:** general news, the most common type, summarizes the overall game in a comprehensive way. It focuses on the players who reached the podium, especially the gold medalist. It details the record of the gold medalist (gap between silver records, intermediate ranks variation) with additional information (from biographical information) and then briefly describes the records of the silver and the bronze medalists. Only one general news article is made per game.

- **Selective news:** selective news is a recap of a particular athlete who participated in a match. Unlike general news, this details the record of the particular player (gap between gold records, intermediate ranks variation) with additional information (from biographical information) and then adds a brief summary sentence about the podium. In principle, the number of selective news articles made is the same as the number of players who participated in the match.[7]

Style

We designed NewsRobot to create news articles in three types of styles, text, slide, and video (see Figure 6-2), according to the level of multimedia modality.

- **Text news:** Text news refers to basic news consisting of only text. This includes numerical data, such as player record and rank.
- **Slide news:** Slide news is news that combines text with graphical features by sentence, turning it into multiple slides. Graphical features include the background color of the medal according to the player's rank, a pictogram of the sport, the player's image (from a real-time Twitter search using the player's name as the hashtag), the player's intermediate ranks variation graph, and a picture of the podium with the national flags. All of the features are

[7] However, considering the selection view of NewsRobot's user interface, we limited the number of players to nine per game. In consideration of the high interest of users, the nine players included mainly high-ranked players and national players. In the case of the Short Track Speed Skating Women's 1,000-Meter Final, as the total number of competitors was six, the number of articles was limited to six.

automatically integrated with each sentence of the article.

- **Video news:** Video news refers to news that plays slide news with sound. The voice API engine [223] automatically reads out the text of the article as an announcer, and the graphical features are displayed when the corresponding text is being read. Background music is included to boost the intensity.

By combining these two types of content and three types of styles, a total of six types of articles (general-text, general-slide, general-video, selective-text, selective-slide, selective-video) are created for each event.

Figure 6-2. An example of Slide news of NewsRobot. Slide news is news that combines text with graphical features by sentence, turning it into multiple slides. Graphical features include the background color of the medal according to the player's rank, a pictogram of the sport, the player's image (from a real-time Twitter search using the player's name as the hashtag), the player's intermediate ranks variation graph, and a picture of the podium with the national flags. All of the features are automatically integrated with each sentence of the article.

6.2.4 Generating News Articles

Based on the data, structure, and six types of news, we created the automatic news generation program with Python programming. The collected data were calculated, refined, and entered into the necessary parts of the news text. The news articles were also created with various multimedia elements, such as images, graphs, voice, and sound. Through the pilot test before the three matches that we selected, we were able to elaborate on the algorithm and source codes of news generation. On each game day, we ran the news generation program immediately after the game finished, and finally, we were able to successfully generate all the news articles we had planned, for a total of 111 news articles.[8] All news articles were generated within five minutes of running the program.

6.2.5 Designing NewsRobot User Interface

After successfully generating the news articles, we created the NewsRobot user interface. The interface was designed to operate on web browsers of users' smartphones, and it was made with JavaScript and HTML5 programming. Participants were able to access NewsRobot's webpage and watch various news articles on the Olympics.[9]

Specifically, users could select a game on the first screen of the user interface (Figure 6-3). On the next screen, the names of the athletes participating in each game were presented in the form of a tile with the national flags. NewsRobot then displayed the six types of news articles for that match in random order.

---

[8] 30 from Speed Skating Men's 1,000 Meters ((1 general + 9 personalized) × 3 styles), 30 from Alpine Skiing Women's Downhill ((1 general + 9 personalized) × 3 styles), and 21 from Short Track Speed Skating Women's 1,000-Meter Final ((1 general + 6 personalized) × 3 styles)

[9] Although the user study was conducted three months after the Olympics, no additional modifications were made to the news articles for the experiment. We only included the news that was generated in real time during the Olympics.

Figure 6-3. NewsRobot User Interface

6.3 Study Design

To assess the user experience of the system, we designed a user study. The design protocol was reviewed and approved by the Institutional Review Board of Seoul National University Institute.

6.3.1 Participants

In the recruitment of participants, we aimed to balance the age and gender of general users who have no trouble accessing news articles using smartphones. We posted a recruiting document on our institution's online community website and recruited a total of 30 participants (15 males and 15 females). Their average age was 30.4, and the SD value was 7.6 (M: Mean = 29.3, SD = 6.5; F: Mean = 31.5, SD = 8.6). Before the experiment started, we provided the participants with a detailed explanation of the purpose and procedure of the experiment as well as NewsRobot. They were told that the news was produced in real time by a computer program during the Olympics. We also ensured that NewsRobot worked properly on the users' smartphones. Users were allowed to manipulate the system for a while so that they could get used to using it. On average, each experiment lasted about 70 minutes. Each participant received a gift voucher worth $10.

6.3.2 Procedures

We designed a user study that consisted of watching racing events and reading/watching news articles on NewsRobot, followed by completing surveys and interviews.

Task

To enhance the users' immersion in the game situation, we prepared the pre-recorded and edited race events videos and played them on a TV screen in the experiment room. The videos of the three races were shown in random order. Before each race was played, the participants were asked to select one of the participating athletes on the NewsRobot interface on their smartphones according to their interests. At the end of each video, the participants were given six types of articles in random order and were asked to respond to the questionnaire for each article type.

Survey

The survey consisted of 18 items from Sundar's index, a representative evaluation measure for online news [224]. Sundar identified multiple criteria used by the public in evaluating news articles, combining them into four major factors: credibility, liking, quality, and representativeness. Credibility is composed of biased, fair, and objective, and liking consists of boring, enjoyable, lively, interesting, and pleasing. Quality has five sub-items: clear, coherent, comprehensive, concise, and well written. Representativeness is composed of accurate, believable, disturbing, informative, and sensationalistic. Users evaluated each news article on the survey with a 7-point Likert scale ranging from highly disagree to highly agree.

Interview

At the end of the experiment, participants were asked to take part in semi-structured interviews. In the interviews, they were asked about their overall impressions of NewsRobot and their thoughts on content and style.

6.3.3 Analysis Methods

After the experiment, we conducted a quantitative analysis on the survey data and a qualitative analysis on the interview data.

Quantitative Analysis

In the quantitative analysis, we aimed to determine whether the content and style of NewsRobot news articles had any significant effects on users' evaluations. As every participant consumed all six different news articles for the three different events (3×6 within-subjects design), we analyzed the survey data using a two-way repeated measures ANOVA. We also conducted Tukey's HSD test as a post-hoc test for pairwise comparisons.

Qualitative Analysis

The qualitative data from interviews were transcribed and analyzed using thematic analysis [195]. In the process, we used Reframer [183], a qualitative research software tool. We segmented the transcripts into sentences and finally obtained 471 observations. While reviewing the data, a total of 250 keyword tags were created. By combining the relevant tags iteratively, we conducted a theme-building process, yielding three main categories.

6.4 Results 1: Quantitative Analysis

In this section, we report the results of the quantitative analysis. As Sundar suggested [224], we explain the 18 items by grouping them into the four main factors: credibility, liking, quality, and representativeness.

6.4.1 Selective News Is Less Credible

We found that content had a significant main effect on three of the items that make up credibility: biased ($F1, 505 = 194.79$; $p<0.001$***), fair ($F1, 505 = 52.55$; $p<0.001$***), and objective ($F1, 505 = 34.07$; $p<0.001$***). In the case of biased, selective news received higher scores than general news (Figure 6-4). In contrast, in the case of fair and objective, general news scores were higher than selective news scores. In other words, it can be seen that users felt that general news was more credible than selective news. Meanwhile, there was no significant main effect of style on the credibility of news articles.

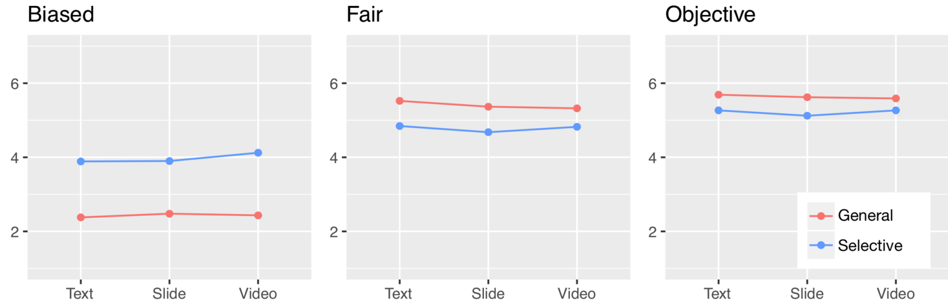


Figure 6-4. Users' evaluation of credibility of NewsRobot news articles

6.4.2 Users Like Both Multimedia and Personalization

We identified that style had a significant effect on all five items of liking: boring ($F2, 505 = 41.78$; $p<0.001$***), enjoyable ($F2, 505 = 60.18$; $p<0.001$***), lively ($F2, 505 = 53.21$; $p<0.001$***), interesting ($F2, 505 = 55.82$; $p<0.001$***), and pleasing ($F2, 505 = 25.91$; $p<0.001$***). The pairwise comparison from the post-hoc analysis revealed that boring scored the highest in text, followed by slide and video (Figure 6-5), showing significant differences between text and slide ($t=7.29$, $p<0.001$***) and text and video ($t=8.42$, $p<0.001$***). In the case of enjoyable, the mean values were in the order of video-slide-text, showing significant differences between video and slide ($t=7.64$, $p<0.001$***) and slide and text ($t=10.64$, $p<0.001$***). Likewise, regarding lively, the mean scores were in the order of video-slide-text (video-slide: $t=4.47$, $p<0.001$***, slide-text: $t=5.81$, $p<0.001$***). In the case of interesting, although the difference between video and slide was not significant ($p=0.43$), the scores showed the same pattern: video-slide-text (slide-text: $t=8.33$, $p<0.001$***). Lastly, pleasing showed the same pattern (video-slide: $t=1.25$, $p=0.64$, slide-text: $t=5.52$, $p<0.001$***).

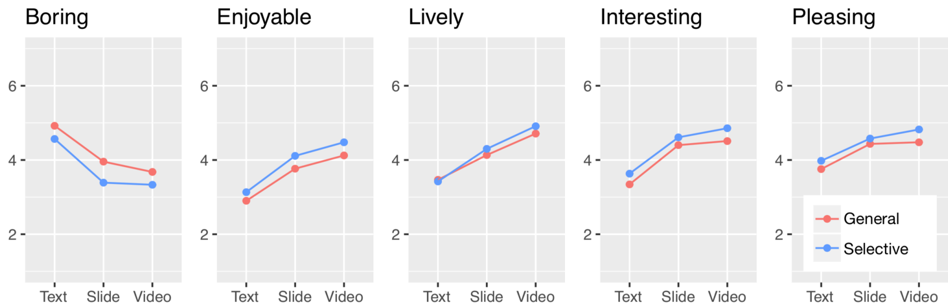


Figure 6-5. Users' evaluation of liking of NewsRobot news articles

We also identified that users liked general news more than selective news. In all the items of liking except for lively ($p=0.32$), content showed a significant main effect: boring ($F1, 505 = 12.36$; $p<0.001$***), enjoyable ($F1, 505 = 60.18$; $p=0.002$**), interesting ($F1, 505 = 7.99$; $p=0.005$**), and pleasing ($F1, 505 = 6.28$; $p=0.013$*). Meanwhile, we found no significant interaction effect.

6.4.3 Quality of Video Is Not Rated Highest

In terms of quality, first, we found that style had a significant main effect on clear ($F1, 505 = 13.15$; $p<0.001$***) and concise ($F1, 505 = 17.55$; $p<0.001$***). However, unlike the order of video-slide-text in credibility and liking, in both items, slide news showed the highest value, followed by video and text (Figure 6-6). The pairwise comparison from the post-hoc analysis revealed that, in the case of clear, there were significant differences in the mean scores between slide and video ($t=2.65$; $p=0.03$*) as well as video and text ($t=2.47$, $p=0.04$*). In the case of concise, the mean score difference between video and slide was significant ($t=3.77$; $p<0.001$***). Although it was not significant, video received a higher score than text ($p=0.12$). From this result, we identified that simply increasing the multimedia modality level does not guarantee the articles' quality, especially in relation to clearness and conciseness.

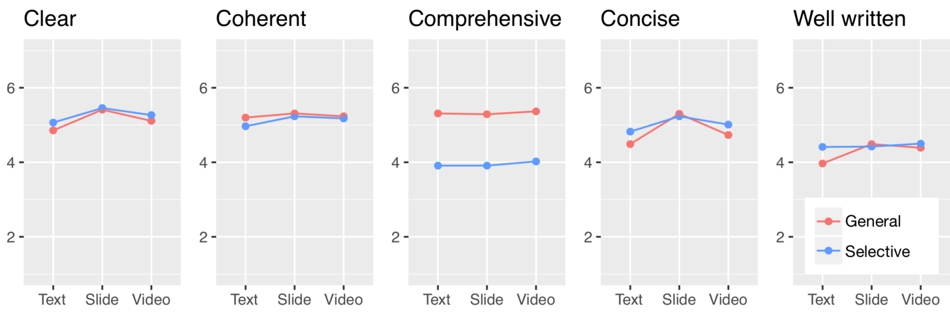


Figure 6-6. Users' evaluation of quality of NewsRobot news articles

Second, we examined whether content had a significant main effect on concise ($F1, 505 = 4.52$; $p=0.03$*) and comprehensive ($F1, 505 = 193.75$; $p<0.001$***). In the case of concise, selective news received higher scores than general news. On the other hand, for comprehensive, general news received higher scores than selective news. It seems that the selective news was considered clearer because it focused on one player without reporting any other information. In addition, it is presumed that general news was evaluated as more comprehensive because it dealt with various types of information in a comprehensive way. Meanwhile, we found no significant main effect on coherent or well written.

6.4.4 NewsRobot Is Accurate but Not Sensational

In the analysis of NewsRobot's representativeness, we focused on some unusual patterns rather than significant relationships between the variables. First, overall, users evaluated NewsRobot's news as relatively accurate and believable. Unlike any other items, all kinds of news articles received scores higher than 5 points in both items (Figure 6-7). Conversely, participants evaluated NewsRobot news as relatively less disturbing and sensationalistic (1–2 points).

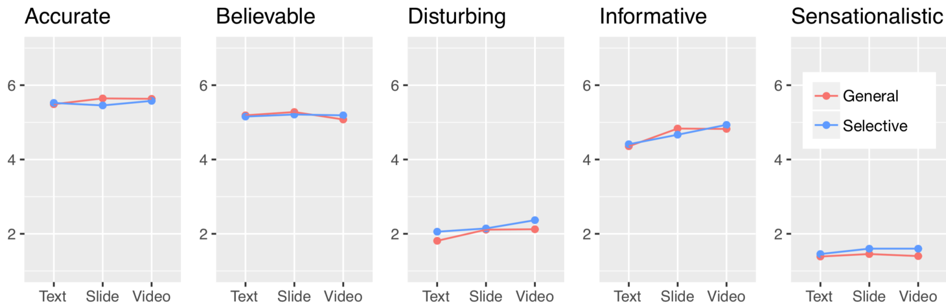


Figure 6-7. Users' evaluation of representativeness of NewsRobot news articles

Finally, style showed a significant main effect on informative ($F_{2, 505} = 12.60$; $p<0.001$***). The pairwise comparison analysis from the post-hoc test revealed that participants rated the NewsRobot news as more informative in the order of video, slide, and text (video-slide: $t=1.25$; $p=0.64$, slide-text: $t=3.59$; $p=0.0011$**, video-text: $t=4.84$; $p<0.001$***). This shows that increasing the multimedia modality level can make news articles seem more informative.

6.5 Results 2: Qualitative Analysis

In this section, we present users' in-depth thoughts about the news articles and NewsRobot.

6.5.1 Users Evaluate NewsRobot Features Highly

We were able to identify that several NewsRobot features provided a good experience for users.

Content

Participants highly appreciated that NewsRobot could generate selective news as well as general news. They told us that selective news was more satisfying, since it contained customized information tailored to their personal interests and delivered it first in the articles. For example, P30 said, "*I really liked the fact that the article started with the result of the player I selected.*" Some participants even told us that selective news made them feel their interests were recognized and respected by the system. P22 said, "*I felt that it [NewsRobot] was trying to learn more about me.*" In addition, some of the participants told us that simply being able to select a player according to their interests made them feel they were provided with a better experience. P07 said, "*I thought it was a simple selection procedure, but it was good. That's why I was so attached to the article that I created.*"

Various Presentation Elements

In addition, participants particularly appreciated NewsRobot's various presentation elements. We identified users' high preference for slide and video news, which contained more elements, such as background colors, pictograms, player images, graphs, and even voice and music. Regarding the colors that made up the slides, participants responded that they helped them intuitively understand how a particular player ranked. P27 said, "*When I first saw the color, I thought it was useful. If the background was gold, I knew immediately that this player had won the gold medal.*" Players' real-time images from Twitter also helped users understand the players. P13 said, "*When the athlete's image emerged in the background, I knew exactly who this athlete was.*" Moreover, users evaluated the use of the voice API in video news as useful. Some users thought the voice was better than expected. P19 said, "*I was curious about how it could read it so naturally.*" P03 said, “*I was able to concentrate more on the news, as the voice automatically read the news.*” Some argued that it would be more useful in particular circumstances, such as driving situations. P06 said, “*It would be great if I could use this while driving, when I cannot read news articles.*”

Among the various elements, participants were most satisfied with the intermediate ranks variation and comparison graphs including each player. Usually, it was difficult for users to know the relative performances of each player, because the players completed their races one after another in turn. However, these graphs showed how quickly each player passed through each point relative to other players, which was like providing users with an appropriate visualization of information that was not provided on the TV screens. P17 said, “*The graph was good, in that I could see the game at a glance.*” Some users told us that they were able to understand the players’ game flow through the graphs. P13 said, “*This player started slowly at the beginning of the game, but he gradually increased his speed in the second half and eventually placed first.*”

Fast Speed of News Generation

Participants were surprised that NewsRobot generated news articles so quickly all at once. Although they read the articles in the experiment after the Olympics, they were amazed and impressed when they heard that all the news articles were generated in real time during the event. For example, P09 said, “*Producing news articles quickly is very important. Of course, reporters can write quickly, but they cannot make so many articles at the same time.*” P30 said, “*It’s very fast. I am surprised that it could create so many articles as soon as the game was over.*” P20 said, “*A reporter could never make slides or video news at this rate.*” This showed that automated journalism could be a very effective and efficient way of producing news, in that it can generate a large variety of articles quickly.

6.5.2 NewsRobot Is Unbiased but Predictable

In line with the survey results, the interviews also revealed that participants described the NewsRobot news articles as accurate and matter-of-fact. They were convinced that since the system computed and included specific figures based on the input data, the information should be accurate. They noted that, for this reason, algorithm-based, automatically generated news articles provide objective and reliable information. P11 said, "*The numerical information gave me the impression that the news was objective.*" P25 also said, "*The figures were reliable, not biased. It seemed to be based on facts. I think the biggest advantage of algorithmic news is that it calculates and displays figures based on data.*" P16 also added, "*NewsRobot cannot lie; it only tells the truth,*" showing strong confidence in the accuracy of the news articles.

Participants also expressed that the lack of emotional elements and the exclusion of subjective judgments in the sentences increased the credibility of the news. For example, P01 said, "*Regardless of whether the player is loved or famous, it [NewsRobot] will describe him in the same way.*" People described algorithm-based news as objective, because it only provides predefined expressions, unlike the news, which often reflects the subjective opinions of reporters and anchors. P05 said, "*I could not find anything subjective in this news. You know, journalists often write using certain expressions to make their arguments more convincing.*" P09 said, "*Reporters' subjective thoughts often influence the atmosphere of the news, but the program reports it as it is.*"

However, despite these advantages, users also pointed out the inherent shortcomings of NewsRobot: It seems tedious, is unable to convey anything but data, and does not understand the context. Participants commonly expressed that the NewsRobot news articles were dull and uninteresting since all the articles followed the predesigned structure without any variation. P27 said, "*AI still seems to require a lot of refining. The biggest problem is that it is dull. I know it is accurate, but so far, it is only based on numbers.*" P07 commented, "*All the articles are almost identical in format. If I were to read more, I would feel bored.*"

Participants also complained that the algorithmic news articles could not deliver anything other than data and could not deliver in-depth news stories. As the system relied only on data, other information that was not included in the data was omitted. For example, in the Short Track Speed Skating Women's 1,000-Meter Final, two players collided with each other and one of them was disqualified. When NewsRobot did not explain the collision in detail and only reported the disqualification, P21 said, "*It just focused on the record; I could not get any more information. Shim Suk-hee was disqualified, and why she was disqualified was the most important part, but it did not discuss that.*" P23 said, "*If someone asked me to explain a race, I would first talk about the collisions between players. Reporters usually discuss the collisions in detail and mention the shocking reactions people have received. But it [NewsRobot] did not.*"

6.5.3 Benefits and Drawbacks of Using Multimedia

The final point of the qualitative study results relates to the benefits and drawbacks of using multimedia. As outlined in the survey results and previous sections, people were able to have a better user experience when NewsRobot produced news articles with multimedia, such as images and voice. They thought of video news as more vivid, pleasant, and informative than either slide or text. However, in terms of clearness and conciseness, video news provided a worse experience than slide news, although the latter had a lower multimedia modality level than the former. We were able to confirm this in the in-depth interviews.

Overall, people described that they were satisfied with the video news but found it strange for two reasons. First, participants discussed the dissonance between their expectations and the actual quality. They claimed that as the multimedia modality levels of news articles increased in order of text, slide, and video, their expectations also grew. However, the gap between their expectations and the actual quality of each style seemed to become larger. When reading text news, because of their low expectations for the text, they felt that the quality was somewhat good. In contrast, when watching video news, their high expectations for the news made them evaluate its quality as relatively low. P08 said, "*In the order of text, slide, and video, I felt they were very different from what I expected. I believe that in video news, a reporter should appear on the screen and inform viewers of the result of the match with a realistic and vivid voice.*"

On the other hand, some participants added that presenting news with multiple multimedia elements made the awkward parts of the news more noticeable. For example, when they read a sentence with an awkward expression, they just passed that part without noticing, whereas when they heard it, they suddenly felt strange and became less focused on the news. P09 said, "*In video news, when NewsRobot read the name of the foreign player aloud, it sounded awkward.*" P24 said, "*When the voice read the same word in succession, like 'third place, third place, third place', I suddenly found it strange. I did not notice it when I read the same thing in text news. If a reporter had read that part aloud, he would not have done that.*"

6.6 Limitations

We have identified several limitations of this chapter. First, we did not conduct a comparative study comparing NewsRobot's articles and actual articles written by human reporters in the evaluation. Second, although we generated the news articles during the Olympic Games in real time, the user evaluation itself was not conducted in real time, since we had to consider unexpected situations that could occur in a real environment and control the experiment. Third, the user survey was limited to one sports game. Lastly, we designed the structure of the articles using a rule-based template and did not make various changes.

6.7 Conclusion

This chapter examined the user experience of automated journalism, where news articles are generated by algorithms, mainly focusing on its personalization and multimedia modality issues. We designed a research prototype, NewsRobot, which generated news articles about the PyeongChang 2018 Winter Olympic Games, and conducted a user study of the system using both quantitative and qualitative approaches. The results of the study revealed the following. (1) Although people regarded general news as more reliable than selective news, they preferred the latter to the former. (2) People also liked news with a high multimedia modality level, but they considered slide news to have the best quality. (3) People regarded NewsRobot as accurate and objective but monotonous, and they were mostly satisfied with its diverse elements. We hope that this work will serve as a step toward a more productive and more inclusive understanding of interfaces for automated news generation systems.

7 DISCUSSION

So far, this dissertation has described four studies that investigated how people interact with algorithm-based systems using AI. Based on the results of each chapter, this chapter discusses AI-based interface design and suggests design principles.

7.1 Human Perception of AI Algorithms

Based on findings from chapter 3, this section discusses the current public awareness of AI and its implications for HCI.

7.1.1 Cognitive Dissonance

According to the theory of cognitive dissonance, when people face an unexpected situation that is inconsistent with their preconceptions, they could experience mental stress or discomfort [225]. We could also see the people's fear of AI through the lens of cognitive dissonance. Through the interviews, we identified that the participants had preconceptions and fixed ideas about AI: (a) AI could be a source of potential danger, and (b) AI agents should help humans. Although these two stereotypes seem to be contradictory, one seeing AI as a potential danger and the other seeing it as a tool, they are connected in terms of control over the technology. The idea that AI could be dangerous to humans can be extended to the idea that it should be controlled so that it can play a beneficial and helpful role for us.

While watching the Google DeepMind Challenge Match, however, people might have faced situations that did not match these notions. The result of the event indicated that humans are no longer superior to AI and cannot control it, which was inconsistent with (b). People might have had difficulty dealing with the possibility of the reversal in position between humans and AI. The participants reported that they felt negative feelings, such as helplessness, disagreeability, depression, a sense of human, and stress. On the contrary, (a) was strengthened. The idea that AI could harm humans provoked people's negative emotions in itself. Thus, it rather reinforced the negative influence of cognitive dissonance caused by (b).

Meanwhile, the negative emotional states attributed to the dissonance show that the fear of AI should not be considered in the view of traditional technophobia, which has focused on the users' basic demographic information and everyday interactions with computers. Users' personal experience with AI is not restricted to real interactions or experiences. Rather, it could be formed from previous media experience and based on their notions and beliefs regarding the technology. In this regard, to understand and neutralize users' technophobia toward AI, we need to include these factors in the theory and practice and discuss ways to reduce dissonance between users' thoughts and real AI-embedded interfaces.

7.1.2 Beyond Technophobia

Two of our most important findings related to certain tendencies behind people's fear of AI: (1) *anthropomorphizing* and (2) *alienating*. They not only anthropomorphized AI as having an equal status with humans but also alienated it, again revealing hostility. While watching AlphaGo's capacities, people regarded it as if it had human-like intelligence. People perceived AI's capacity as being comparable to a human's and interpreted the behaviors of AI by treating it as if it were a rational agent who controls its choice of action [60]. However, at the same time, people also alienated AI by regarding it as different and showed hostility. They tried to find the different aspects of AI and evaluated it with the characteristics of a human being, and they dehumanized it [61] if it was thought to be transcending or differing from such characteristics.

This tendency of anthropomorphizing and alienating AI was not a common phenomenon in their experience of computers, as the participants stated in the interviews. Since they regarded the computer as a tool to complete certain tasks, the computer problem is not related to the computer itself but mainly related to anxiety arising from interactions with it. On the contrary, people viewed AI as a being who almost has a persona. In this sense, the problem does not seem to be a technological issue but similar to communication and relationship problems among humans. In addition, they tried to find its different aspects and then alienated it. Accordingly, the fear of AI may not be a problem of technophobia but an issue similar to xenophobia, the fear of that which is perceived to be foreign or strange.

In this sense, reducing users' fear of AI should be accomplished by understanding the communication among humans rather than simply investigating the superficial problems around the computer interface. Previous studies that proved people show similar social behavior during human-computer interaction by adapting human-human interaction ([64], [65]) also support the need for this viewpoint. In particular, in designing interfaces using AI, reflecting and considering major factors common in human communications and establishing the relationship between AI and users could be crucial for improving the user experience. According to each interface's function and purpose, the relationship and the manner of communication between users and human-like agents (and algorithms) should be set clearly and appropriately.

7.1.3 Toward a New Chapter in Human-Computer Interaction

The Google DeepMind Challenge Match was not just an interesting event in which AI defeated the human Go champion. It was a major milestone marking a new chapter in the history of HCI. We found it to be an opportunity to assess and understand people's view of AI technology and discuss considerations for HCI as we gradually integrate the AI technology within user interfaces.

AI is expected to be used in various devices and services, and users will have more chances to interact with interfaces utilizing the technology. As the term "algorithmic turn [226]" suggests, algorithms will then play an increasingly important role in user interfaces and the experiences surrounding them. Moreover, as algorithms could be cross-linked on various interfaces, it is expected to affect users' holistic experience, such as behavior and lifestyle. This is almost like the "environmental atmospheric" media that Hansen suggested for the twenty-first-century media [227].

In this respect, we suggest to the HCI community the concept of "algorithmic experience" as a new stream of research on user experience in AI-embedded environments. It encompasses diverse aspects of longitudinal user experience with algorithms that are environmentally embedded in various interfaces. This suggests considering long-term relationships with algorithms rather than the simple usability and utility of interfaces. Regarding interfaces, we need to extend their borders to the various devices and services to which algorithms could be applied. In terms of the user, user experience should not be restricted to simple interactions with machines or computers but should be extended to communication and relationship building with an agent. We believe this new concept can help the HCI community to accept and integrate AI into UI and UX design. It calls for a collaborative research effort from the HCI community to study users and help them adapt to a future in which they interact with algorithms as well as interfaces.

7.1.4 Coping with the Potential Danger

Although it is not within the scope of this chapter to address ways of coping with the potential danger of AI, we cannot neglect the gravity of the issue. As shown in the interviews, people revealed their concerns about AI threatening their lives and existence. This shows that the AI problem is not restricted to individuals, and it needs to be addressed as a social problem. Participants insisted that there should be an institutional policy, encompassing law and common ethics, regarding AI. In addition, they argued that sufficient discussions should take precedence when doing so. Recently, it was reported that the world's largest tech companies, all of which are also closely linked to HCI fields, agreed to make a standard of ethics around the creation of AI [228]. This movement signifies the importance of understanding how people perceive and view AI. We believe discussions on building a desirable relationship between humans and AI will play a vital role in the process of devising the standards.

7.2 Users' Interpretation and Evaluation of AI Algorithms

This section discusses lessons learned from the user study in chapter 4 and its implications for user interfaces with AI algorithms that convey subjective information that can be interpreted in diverse ways.

7.2.1 Integrate Diverse Expertise and User Perspectives

Through the user study, we identified that users interpreted AI in different ways according to their group. AI/ML experts tried to find out the characteristics of its training data and learning process based on their knowledge of AI. Photographers looked at it considering the elements of photography and cameras. The non-experts tried to understand it based on their impressions of the photos without relevant prior knowledge. Most notably, contrary to our expectations, AI/ML experts showed the greatest difference from AI and the lowest interpretability and reasonability scores. On the other hand, the photographers showed the smallest difference from AI and the highest interpretability and reasonability scores.

The differences in viewpoints among users according to groups provide implications for algorithmic transparency and fairness. Although advanced AI algorithms are being developed in a wide range of fields, the algorithms that AI/ML experts design might not apply to their activities pertain. Even the best-performing algorithms do not consider the transparency or fairness. Even AI/ML experts themselves may have difficulty interpreting and understanding it. Such algorithms may not be understandable by a wide variety of people. The experts may not be sure of how many or what types of people and groups can agree with the algorithm. For these reasons, many people could feel that AI algorithms are obscure and biased and cannot embrace their own perspectives [84].

The fact that domain experts can better grasp the AI algorithm and narrow the gap with it shows that the views and opinions of field experts are vital for algorithm development and refinement. AI experts should consider communicating with them to gain the necessary perspectives and knowledge in the field. In addition, they should reflect this knowledge in the development of AI algorithms and further interface development and try to ensure the transparency of algorithms from the outset and avoid bias. Differences in viewpoints in the inferences of the general public without expertise also lead us to question what needs to be reflected for real users in algorithm design. To enable understanding among many people, we should obtain feedback from various general users and offset the inequality problem of the algorithm. Furthermore, various populations should be considered in planning and developing algorithms and securing data.

7.2.2 Take Advantage of People's Curiosity about AI Principles

In the absence of any information about AI Mirror's aesthetic score calculation process, users were curious about the algorithm and constantly strived to learn the principles actively through various strategies. Sometimes, people had hypotheses and tested them by taking slightly different photos. Other times, they just explored without a clear hypothesis or direction. Through these strategies, they were finally able to narrow the gap between their thoughts and those of AI. We can think of design implications on both the user side and the AI side.

First, on the user's side, we can consider introducing these factors into the design of tools that help people to understand AI/ML models, which has recently received a great deal of attention [229]. AI can use the strategies people utilized to help them understand its principles. An AI interface needs to prepare and show as many examples as possible so that people can understand the principle as easily as possible. It is also possible to improve the user's understanding in a microscopic manner by preparing several examples and images with similar but clear differences. A macroscopic approach can be presenting users with completely different examples or images to enable new and diverse ideas and expand their thinking. Through these, the public would be able to reduce the difference between their thoughts and those of AI and understand the principles of algorithms easily.

On the other hand, the various strategies and willingness to discover the principles shown by users suggest implications for the AI domain in relation to the production and securing of high-quality data. According to the information gap theory, when people are made aware of a gap between what they know and what they do not know, they become curious and engage in information-seeking behavior to complete their understanding and resolve the uncertainty [230]. This innate desire to satisfy their curiosity can be helpful in gathering information about the way users interpret AI. Through this, we might collect feedback on various use cases and utilize it to improve algorithms. Indeed, curiosity interventions have been shown to improve crowd worker retention without degrading performance [231].

Designing a platform for AI to stimulate users' curiosity and receive various opinions would be useful in securing the large-scale, high-quality data necessary for algorithm refinement. In fact, it is not that users have to reduce their differences with AI. It should actually be the other way around. AI should be designed to learn from users and narrow the gap with them rather than waiting for the users to do so. Such a view might have implications for online learning or reinforcement learning generally, as systems can adaptively learn from user feedback and improve themselves.

7.2.3 Provide AI and Users with Mutual Communication

Finally, we focus on communication between users and AI. Although users tried various ways to understand the AI, they eventually expressed great dissatisfaction with the lack of direct communication with the AI. Users wanted the AI to give them more detailed descriptions directly, but they also wanted to explain their ideas to the AI. Some users even felt negative feelings and sometimes lost confidence. In particular, since aesthetic evaluation is intrinsically highly subjective, the problem of communication due to this difference in interpretation may be even greater.

In this case, we can consider a mixed-initiative approach to help users and intelligent systems collaboratively achieve their goals [100]. Introducing communication channels for users and algorithms in the design of AI-based interfaces could be considered. To resolve users' uncertainties, AI needs to present users with detailed explanations of the reasons for its decisions ([9], [74], [75], [100], [232]). Of course, users should also be able to present their opinions to the AI. The AI should be able to accept a user's opinion, take it as a dataset, and reflect it in the learning process of the algorithm. Rather than a static AI that only presents predetermined results, a dynamic and adaptable AI that responds to users' thoughts and controls should be considered. Through this process, a two-way communication interface could be designed where the user understands the AI and the AI understands the user, refining results.

7.3 How People Build Sequential Actions with AI Algorithms

This section discusses the findings of the study in chapter 5 and its implications for user interfaces in which users and AI collaborate.

7.3.1 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.

7.3.2 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

([211], [212]). *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.

7.3.3 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 [233] 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 [234]. 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.

7.3.4 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 highand 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 ([235], [236]). 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.

7.4 Practical Design of Algorithm-based Systems Using AI

This section discusses lessons learned from the study in chapter 6 and its implications for user interfaces of automated news generation systems.

7.4.1 Provide Selective news with Adaptable Interface

The first thing we discuss is the personalization of news articles. Although many news channels have appeared and offered a variety of news to people, problems related to the bias of news production have been constantly raised [237]. The concentration of news production may not meet the diverse needs of users and may even lead to social problems in which public opinion is biased toward specific issues.

In this regard, NewsRobot showed the possibility and user experience of automated content. NewsRobot's preprogrammed algorithm simultaneously generated multiple player-specific articles, without the extra expense, effort, or time associated with creating articles using human reporters. Users also showed great satisfaction with the selective news. They preferred selective news to general news and rated the former as less boring and more enjoyable, interesting, and pleasing than the latter. They felt that the algorithms produced news tailored to their interests. Some even described that the selective news made them feel their tastes were respected.

However, at the same time, we need to consider and anticipate the adverse effects of selective news consumption. Selective news consumption can make people more biased in other respects. It can separate users from information that does not attract their interest, isolating them in their own thoughts. In particular, in political news rather than sports news, this can cause problems such as filter bubbles and echo chambers ([83], [152]). The survey results also showed that users were aware that selective news articles are less credible than general news articles. If personalization is accelerated and transparency in the personalization process is not ensured and recognizable by the people, selective news is likely to cause many side effects in spite of its many advantages.

In this situation, we need to consider designing adaptable interfaces of automation systems [99]. Adaptable interfaces provide a customization mechanism that relies on the user to perform the adaptation (i.e., user-controlled personalization). Users need to select their news according to their interests, by themselves, and be explicitly informed that the news articles they receive are based on their choices [142]. Fortunately, users also responded that they had a good experience in the news selection process, as it made them feel like they were active news service consumers.

Of course, designing these interfaces requires a lot of attention and effort from interface designers as well as reporters and algorithm engineers, as it is not just an algorithm problem but also an interface problem. In this regard, we urge the designers of the HCI community to address their concerns and efforts in order to prevent the potential risks of such algorithms and to promote better user experiences.

7.4.2 Present Various Multimedia Elements but Not Too Many

The second discussion point is about the style of NewsRobot. The user study showed that participants had better news consumption experiences when more multimedia elements were included. People thought that news articles were more enjoyable, lively, interesting, pleasing, and even informative in order of video, slide, and text news. They expressed high satisfaction by specifically mentioning the various presentation elements of NewsRobot. People appreciated the fact that it could produce those elements more quickly, accurately, and easily than human reporters and provide additional information not previously revealed.

On the other hand, people complained that video news was a bit different from their expectations and that it brought out the awkwardness of automatically generated news content. According to user research studies on software or agents that replace human workers or simulate abilities of humans, people can feel more awkward or uncomfortable as they become more similar to humans ([235], [236]). The more elements that are included in the news, the greater the expectations of people. However, at the same time, the awkwardness could be more prominent, and people's disappointment could be greater [238]. Since people lose confidence in algorithmic reporters more quickly than they do in human reporters after seeing them make the same mistake [239], this problem needs to be solved in algorithm-based news generation.

To overcome these drawbacks while taking advantage of the various presentation elements, we need to consider creating news content that is more perceptible to people rather than creating overly experimental and challenging news content. It is necessary to consider the mental models that people have about each multimedia modality and reflect them in the design of each news article style. This means that we must first understand news media users. Based on this, if the information representation is tailored to each style, it would be helpful for providing news consumers with a better user experience.

7.4.3 Importance of Quality Data and Algorithm Refinement

The last point we discuss is the importance of obtaining quality data and refining algorithms for automatic news generation. In the user study, participants highly appreciated NewsRobot's data-driven news generation algorithm, describing it as accurate, objective, and even trustworthy. However, at the same time, they pointed out its limitations due to its excessive dependence on data, describing it as dull and shallow. Considering that software-generated content is usually perceived as descriptive and boring, although it is also considered objective [151], this could be an inherent problem that automotive journalism needs to address.

In order to alleviate these problems while taking advantage of providing data-based information, it is necessary to obtain more and better data and refine the news generation algorithm. It is necessary to use diverse data sources, such as image information extraction and social media reactions, and avoid using only refined numerical data. The algorithms should actively interpret the input data and produce various versions of the articles. In addition, they should not only provide calculated values but also help users understand the information by adding more detailed explanations of the meaning.

However, ultimately, it is necessary to introduce state-of-the-art artificial intelligence technologies that can learn and interpret data on their own, and generate sentences and elements based on it, rather than depending on rule-based templates. To do so, stakeholders, such as designers and developers, should continue to search for new technologies related to news generation, such as natural language generation APIs that can express various voices and emotions. Finally, reflecting the opinions of human actors, such as journalists, in the design of algorithms should also be considered to enhance news quality.

7.5 Principles

This paper proposes unique design principles based on the implications of each study.

- **Principle 1** (resulting from the study on perception, RQ1): AI-based systems should be able to mitigate the subtle unpleasantness that a user may have about AI, and interact with the user in a human-like way.
- **Principle 2** (resulting from the study on interpretation and evaluation, RQ2): AI-based systems should be able to present their internal logic, and communicate their reasoning to users when providing information that is automatically calculated by the algorithm.
- **Principle 3** (resulting from the study on sequence of actions, RQ3): AI-based systems should yield to the user's initiative in a series of interactions with the

user, and give only as much instruction as necessary.

- **Principle 4** (resulting from the fourth study on practical applications, RQ4): AI-based systems should be designed to fit users' expectations and mental models of the system. Collaboration between designers and technicians can help elaborate this process.

8 CONCLUSION

Artificial intelligence algorithms are affecting much of our daily lives in numerous areas. this thesis aims to understand how users interact with AI algorithms. Specifically, this dissertation examine algorithm-based human–AI interaction through different stages with various modes of human-computer interaction: The first study investigated how people perceive algorithm-based systems using artificial intelligence, finding that people tend to regard it as a human-like agent, which is distinct from their perceptions of computers. The second study investigated how people interpret and evaluate AI algorithms with a research probe, "AI Mirror," which evaluates the aesthetic scores of given images based on a neural network algorithm. The results revealed that people understand AI algorithms based on their backgrounds and that they want to understand and communicate with the AI algorithm. The third study investigated how people build a sequence of actions with AI algorithms through a mixed method study with a research prototype, "DuetDraw," a drawing tool where users and AI can draw pictures together. The results showed that people want to lead the collaboration while hoping to get appropriate instructions from the AI algorithm. Lastly, a case study on a practical application of the AI was conducted with a research prototype, "NewsRobot," which automatically generates news articles with different content and styles. Findings showed the users' preference for selective news and multimedia news. With these distinct but intertwined studies, this thesis argues that users want to have a more human-like relationship with AI algorithms.

To sum up, artificial intelligence algorithms and users are in a subtle relationship. People tend to not just anthropomorphize but also alienate AI. In particular, given the information automatically computed by the algorithm for the subjective domain, the user wants to know the rationale and wants to communicate with the artificial intelligence. The user does not want to lose control when continuing to interact with artificial intelligence and wants a detailed description of the information they need. Users want to be provided with functionality that meets their expectations rather than artificial intelligence showing off their capabilities.

8.1 Summary of Contributions

The core contribution of this thesis is an increased understanding of AI algorithms in terms of human factors and user experience, investigated through HCI methods. It can be divided into the following detailed contributions.

- **Empirical results on human–AI algorithm interaction:** Through both quantitative and qualitative approaches, this study closely observed the interaction between AI algorithms and users and discovered new aspects of this interaction. It investigated people's fear of AI from various perspectives and identified the confrontational "us vs. them" view between humans and AI, which is distinct from existing views on computers. This work also yielded experimental results showing how users' unique characteristics affected the process of interpreting the outcomes of AI algorithms in terms of strategy, and communication. Furthermore, the study provided insights on the user experience of an automated news generation system.

- **Artifacts:** AI-powered user interfaces were designed for three of the studies, playing a crucial role in understanding the user's interactions: AI Mirror, a user interface that gives aesthetic scores to photographs based on a deep neural network model, DuetDraw, a collaborative drawing application based on neural network technology, and NewsRobot, an automated news generation system that produces multiple news articles considering content and style.
- **Design implications:** This thesis discussed design implications for intelligent user interfaces that are based on AI algorithms. These include implications for interfaces that deliver a variety of interpretable results, which could be utilized by both the AI/ML and HCI communities, interfaces with which users and AI can communicate and cooperate for creative work, and practical interfaces that provide users with information in various forms.
- **Theoretical contribution:** This thesis stresses the importance of AI algorithms and their human factors and user experience in the HCI field and suggests the concept of an expanded user interface and algorithmic experience.

8.2 Future Directions

Understanding human-AI interactions through algorithm-based interfaces provides important opportunities for both the HCI and AI communities.

First, based on the research limitations of each chapter, the following future work can be proposed.

- **Human perception of AI algorithms:** Quantitative research investigating the worldwide reaction to AI by crawling and analyzing data from social network sites could be considered.
- **Users' reasoning about AI algorithms:** Determining a clearer relationship

between variables by carrying out an expanded study with more participants could be considered. A user study in which participants experience the AI system in a real context rather than in a controlled environment could be considered. In addition, improving the research tool to cover various areas of AI rather than limiting it to aesthetic evaluation could be planned. Finally, research that demonstrates the practical effects of the design recommendations that the chapter has proposed could be planned.

- **Human-AI collaboration:** Investigating user experience in a wider variety of interfaces beyond the framework of drawing tools could be planned. As part of the ongoing research of this thesis, there are also plans to improve DuetDraw so that users can use it more flexibly and explore the long-term experience of cooperation with AI.

- **Designing algorithm-based user interfaces:** A comparative study on news articles written by both human journalists and NewsRobot could be planned. Moreover, a user study on an event that is actually happening, measuring the user experience of the automated news generation system in real time could be considered. It could also be planned to expand the user study area to various topics, such as election reports and weather forecasts, to generalize our results on automated journalism. Lastly, improving the automation level of news generation by adopting state-of-the-art techniques could be considered.

In addition, various HCI research topics related to AI can be considered:

- **Designers with Machine Learning:** The study not only presented diverse design guidelines for AI interfaces but also discussed the role of designers in working with algorithms. As AI continues to expand, interaction designers must incorporate this new technology into their product and service designs.

Discussing the practical methods that help designers become accustomed to learning about AI and communicating with stakeholders in the field of AI could be considered for future work.

- **Human-Robot Interaction:** Although the scope of this dissertation is limited to user interaction with intangible algorithms, the results can be extended to tangible AI, or robot interface design studies. This thesis could be extended to study the effects of the existence of robots on user behavior and perhaps inform the design of products that can positively affect people's feelings, such as social robots.
- **Algorithmic Experience:** Algorithms will play an increasingly important role in AI user interfaces. As algorithms become cross-linked and embedded in various interfaces, it is reasonable to expect that algorithms themselves will affect users' holistic experience, including their behaviors and lifestyles. This possibility suggests that research should focus on understanding long-term relationships with algorithms rather than simple usability and utility of interfaces. The human perception of AI algorithms revealed in this thesis can serve as the basis for research into these algorithmic experiences.
- **AI and Crowds:** Although this study focused on the interaction between an individual user and an AI algorithm, the interaction between a large group of users (a crowd) and AI from a macroscopic point of view could be considered. The most critical consideration for AI is to have high-quality data. Crowdsourcing is an excellent way to obtain this, but it is necessary to discuss how to promote crowd participation in various conditions. In particular, investigating whether crowds' behaviors are influenced by AI algorithms' interpretability, or, alternatively, whether crowd behavior could be directed via quality data production, could be considered.

Bibliography

[1] @demishassabis, "#AlphaGo WINS!!!! We landed it on the moon. So proud of the team!! Respect to the amazing Lee Sedol too." Mar-2016.

[2] E. Guizzo, "How google's self-driving car works," *IEEE Spectr. Online, Oct.*, vol. 18, 2011.

[3] M. Montemerlo *et al.*, "Junior: The stanford entry in the urban challenge," *J. F. Robot.*, vol. 25, no. 9, pp. 569–597, 2008.

[4] C. C. Bennett and K. Hauser, "Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach," *Artif. Intell. Med.*, vol. 57, no. 1, pp. 9–19, 2013.

[5] S. E. Dilsizian and E. L. Siegel, "Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment," *Curr. Cardiol. Rep.*, vol. 16, no. 1, pp. 1–8, 2014.

[6] S. Gelly and D. Silver, "Monte-Carlo tree search and rapid action value estimation in computer Go," *Artif. Intell.*, vol. 175, no. 11, pp. 1856–1875, 2011.

[7] I. Millington and J. Funge, *Artificial intelligence for games*. CRC Press, 2016.

[8] E. Vargiu and M. Urru, "Exploiting web scraping in a collaborative filtering-based approach to web advertising," *Artif. Intell. Res.*, vol. 2, no. 1, p. p44, 2012.

[9] D. Gunning, "Explainable artificial intelligence (xai)," *Def. Adv. Res. Proj. Agency (DARPA), nd Web*, 2017.

[10] D. L. Poole, A. K. Mackworth, and R. Goebel, *Computational intelligence: a logical approach*, vol. 1. Oxford University Press New York, 1998.

[11] J. Kaplan, "Artificial intelligence: think again," *Commun. ACM*, vol. 60, no. 1, pp. 36–38, 2016.

[12] S. J. Russell, P. Norvig, J. F. Canny, J. M. Malik, and D. D. Edwards, *Artificial intelligence: a modern approach*, vol. 2. Prentice hall Upper Saddle River, 2003.

[13] P. McCorduck, "Machines Who Think , Natick, MA: AK Peters, Ltd," 2004.

[14] X.-B. Jin, T.-L. Su, J.-L. Kong, Y.-T. Bai, B.-B. Miao, and C. Dou, "State-of-the-Art Mobile Intelligence: Enabling Robots to Move Like Humans by Estimating Mobility with Artificial Intelligence," *Appl. Sci.*, vol. 8, no. 3, p. 379, 2018.

[15] D. Crevier, *AI: the tumultuous history of the search for artificial intelligence*. Basic Books, 1993.

[16] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[17] N. R. Council and others, *Funding a revolution: Government support for computing research*. National Academies Press, 1999.

[18] G. F. Luger, *Artificial intelligence: structures and strategies for complex problem solving*. Pearson education, 2005.

[19] J. McDermott, "R1 ('XCON') at age 12: lessons from an elementary school achiever," *Artif. Intell. Perspect.*, p. 241, 1994.

[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," 1985.

[21] J. Grudin, "AI and HCI: Two fields divided by a common focus," *AI Mag.*, vol. 30, no. 4, p. 48, 2009.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] G. E. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

[25] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[26] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, 2013, pp. 6645–6649.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[28] E. Reiter, "Natural language generation," *Handb. Comput. Linguist. Nat. Lang. Process.*, pp. 574–598, 2010.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[30] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in Neural Information Processing Systems*, 2016, pp. 3844–3852.

[31] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in neural information processing systems*, 2009, pp. 1096–1104.

[32] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv Prepr. arXiv1406.1078*, 2014.

[33] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv Prepr. arXiv1409.0473*, 2014.

[34] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Brief. Bioinform.*, vol. 18, no. 5, pp. 851–869, 2017.

[35] D. Shen, G. Wu, and H.-I. Suk, “Deep learning in medical image analysis,” *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.

[36] S. K. Zhou, H. Greenspan, and D. Shen, *Deep learning for medical image analysis*. Academic Press, 2017.

[37] D. Silver *et al.*, “Mastering the game of Go with deep neural networks and tree search,” *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[38] E. Gibney, “Google AI algorithm masters ancient game of Go,” *Nat. News*, vol. 529, no. 7587, p. 445, 2016.

[39] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv Prepr. arXiv1312.6114*, 2013.

[40] I. Goodfellow *et al.*, “Generative adversarial nets,” in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[41] R. High, “The era of cognitive systems: An inside look at IBM Watson and how it works,” *IBM Corp. Redbooks*, 2012.

[42] V. Mnih *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[43] Y. Rogers, H. Sharp, and J. Preece, *Interaction design: beyond human-computer interaction*. John Wiley & Sons, 2011.

[44] J. P. Bigham and J. Seering, “Human-AI Interaction: Introduction,” *Spec. Top. HCI Hum. AI Interact.*, 2018.

[45] H. Lieberman, “User interface goals, AI opportunities,” *AI Mag.*, vol. 30, no. 4, p. 16, 2009.

[46] K. Hornbæk and A. Oulasvirta, “What is interaction?,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 5040–5052.

[47] R. S. Nickerson, J. I. Elkind, and J. R. Carbonell, “Human factors and the design of time sharing computer systems,” *Hum. Factors*, vol. 10, no. 2, pp. 127–133, 1968.

[48] R. D. Pea, “User centered system design: New perspectives on human-computer interaction,” 1986.

[49] D. A. Norman, “Cognitive engineering,” *User centered Syst. Des.*, vol. 31, p. 61, 1986.

[50] D. Norman, *The design of everyday things: Revised and expanded edition*. Constellation, 2013.

[51] J. L. Dyck and J. A.-A. Smither, "Age differences in computer anxiety: The role of computer experience, gender and education," *J. Educ. Comput. Res.*, vol. 10, no. 3, pp. 239–248, 1994.

[52] D. Gilbert, L. Lee-Kelley, and M. Barton, "Technophobia, gender influences and consumer decision-making for technology-related products," *Eur. J. Innov. Manag.*, vol. 6, no. 4, pp. 253–263, 2003.

[53] M. Brosnan and W. Lee, "A cross-cultural comparison of gender differences in computer attitudes and anxieties: The United Kingdom and Hong Kong," *Comput. Human Behav.*, vol. 14, no. 4, pp. 559–577, 1998.

[54] A. S. North and J. M. Noyes, "Gender influences on children's computer attitudes and cognitions," *Comput. Human Behav.*, vol. 18, no. 2, pp. 135–150, 2002.

[55] L. M. Anthony, M. C. Clarke, and S. J. Anderson, "Technophobia and personality subtypes in a sample of South African university students," *Comput. Human Behav.*, vol. 16, no. 1, pp. 31–44, 2000.

[56] R. D. Henderson, F. P. Deane, and M. J. Ward, "Occupational differences in computer-related anxiety: Implications for the implementation of a computerized patient management information system," *Behav. Inf. Technol.*, vol. 14, no. 1, pp. 23–31, 1995.

[57] M. M. Weil and L. D. Rosen, "The psychological impact of technology from a global perspective: A study of technological sophistication and technophobia in university students from twenty-three countries," *Comput. Human Behav.*, vol. 11, no. 1, pp. 95–133, 1995.

[58] G. A. Marcoulides and X.-B. Wang, "A cross-cultural comparison of computer anxiety in college students," *J. Educ. Comput. Res.*, vol. 6, no. 3, pp. 251–263, 1990.

[59] B. R. Duffy, "Anthropomorphism and the social robot," *Rob. Auton. Syst.*, vol. 42, no. 3, pp. 177–190, 2003.

[60] D. C. Dennett, *Kinds of minds: Toward an understanding of consciousness*. Basic Books, 2008.

[61] A. Waytz, N. Epley, and J. T. Cacioppo, "Social Cognition Unbound Insights Into Anthropomorphism and Dehumanization," *Curr. Dir. Psychol. Sci.*, vol. 19, no. 1, pp. 58–62, 2010.

[62] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey, “Anthropomorphic interactions with a robot and robot-like agent,” *Soc. Cogn.*, vol. 26, no. 2, p. 169, 2008.

[63] S. Kiesler, L. Sproull, and K. Waters, “A prisoner’s dilemma experiment on cooperation with people and human-like computers.,” *J. Pers. Soc. Psychol.*, vol. 70, no. 1, p. 47, 1996.

[64] C. Nass, J. Steuer, E. Tauber, and H. Reeder, “Anthropomorphism, agency, and ethopoeia: computers as social actors,” in *INTERACT’93 and CHI’93 conference companion on Human factors in computing systems*, 1993, pp. 111–112.

[65] C. Nass, J. Steuer, and E. R. Tauber, “Computers are social actors,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1994, pp. 72–78.

[66] B. Reeves and C. Nass, *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press Cambridge, UK, 1996.

[67] D. Castelvecchi, “Can we open the black box of AI?,” *Nat. News*, vol. 538, no. 7623, p. 20, 2016.

[68] G. Alain and Y. Bengio, “Understanding intermediate layers using linear classifier probes,” *arXiv Prepr. arXiv1610.01644*, 2016.

[69] W. Knight, “The Dark Secret at the Heart of AI,” MIT Technology Review, 2017.

[70] R. Shwartz-Ziv and N. Tishby, “Opening the black box of deep neural networks via information,” *arXiv Prepr. arXiv1703.00810*, 2017.

[71] N. Tishby and N. Zaslavsky, “Deep learning and the information bottleneck principle,” in *Information Theory Workshop (ITW), 2015 IEEE*, 2015, pp. 1–5.

[72] D. Lei, X. Chen, and J. Zhao, “Opening the black box of deep learning,” *arXiv Prepr. arXiv1805.08355*, 2018.

[73] B. Kim *et al.*, “Interpretability beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV),” in *International Conference on Machine Learning*, 2018, pp. 2673–2682.

[74] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, “Trends and trajectories for explainable, accountable and intelligible

systems: An hci research agenda," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, p. 582.

[75] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning," *arXiv Prepr. arXiv1806.00069*, 2018.

[76] M. K. Lee, D. Kusbit, E. Metsky, and L. Dabbish, "Working with machines: The impact of algorithmic and data-driven management on human workers," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 1603–1612.

[77] C. Oh, T. Lee, Y. Kim, S. Park, B. Suh, and others, "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*, 2017, pp. 2523–2534.

[78] M. Eslami *et al.*, "I always assumed that I wasn't really that close to [her]: Reasoning about Invisible Algorithms in News Feeds," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 153–162.

[79] R. Sinha and K. Swearingen, "The role of transparency in recommender systems," in *CHI'02 extended abstracts on Human factors in computing systems*, 2002, pp. 830–831.

[80] E. Rader, K. Cotter, and J. Cho, "Explanations as Mechanisms for Supporting Algorithmic Transparency," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, p. 103.

[81] M. Veale, M. Van Kleek, and R. Binns, "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, p. 440.

[82] J. Otterbacher, J. Bates, and P. Clough, "Competent men and warm women: Gender stereotypes and backlash in image search results," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 6620–6631.

[83] E. Bozdag, "Bias in algorithmic filtering and personalization," *Ethics Inf. Technol.*, vol. 15, no. 3, pp. 209–227, 2013.

[84] A. Woodruff, S. E. Fox, S. Rousso-Schindler, and J. Warshaw, "A Qualitative Exploration of Perceptions of Algorithmic Fairness," in

Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, p. 656.

[85] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 2000, pp. 241–250.

[86] T. Zarsky, "The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making," *Sci. Technol. Hum. Values*, vol. 41, no. 1, pp. 118–132, 2016.

[87] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card, "The cost structure of sensemaking," in *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, 1993, pp. 269–276.

[88] G. Klein, B. Moon, and R. R. Hoffman, "Making sense of sensemaking 2: A macrocognitive model," *IEEE Intell. Syst.*, vol. 21, no. 5, pp. 88–92, 2006.

[89] D. V Day and R. G. Lord, "Expertise and problem categorization: the role of expert processing in organizational sense-making," *J. Manag. Stud.*, vol. 29, no. 1, pp. 35–47, 1992.

[90] C. Naumer, K. Fisher, and B. Dervin, "Sense-Making: a methodological perspective," in *Sensemaking Workshop, CHI'08*, 2008.

[91] B. Dervin, "Project overview: Sense-Making Methodology as dialogic approach to communicating for research and practice," *Sense-making Inf. Conflu. whys hows Coll. Univ. user Satisf. Inf. needs. Phase I Proj. overview, Three-f. Dialogue Proj. state-of-the-art Rev.*, 2006.

[92] E. M. Meyers, K. E. Fisher, and E. Marcoux, "Making sense of an information world: The everyday-life information behavior of preteens," *Libr. Q.*, vol. 79, no. 3, pp. 301–341, 2009.

[93] C. Shah, "Coagmento-a collaborative information seeking, synthesis and sense-making framework," *Integr. demo CSCW*, vol. 2010, 2010.

[94] J. Stasko, C. Görg, and Z. Liu, "Jigsaw: supporting investigative analysis through interactive visualization," *Inf. Vis.*, vol. 7, no. 2, pp. 118–132, 2008.

[95] B. Shneiderman and P. Maes, "Direct manipulation vs. interface agents," *interactions*, vol. 4, no. 6, pp. 42–61, 1997.

[96] H. Lieberman and others, "Letizia: An agent that assists web browsing,"

IJCAI (1), vol. 1995, pp. 924–929, 1995.

[97] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse, "The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users," in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, 1998, pp. 256–265.

[98] C. Ahlberg and B. Shneiderman, "Visual information seeking: Tight coupling of dynamic query filters with starfield displays," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1994, pp. 313–317.

[99] L. Findlater and J. McGrenere, "A comparison of static, adaptive, and adaptable menus," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 89–96.

[100] E. Horvitz, "Principles of mixed-initiative user interfaces," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 1999, pp. 159–166.

[101] J. E. Allen, C. I. Guinn, and E. Horvtz, "Mixed-initiative interaction," *IEEE Intell. Syst. their Appl.*, vol. 14, no. 5, pp. 14–23, 1999.

[102] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[103] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *International conference on machine learning*, 2013, pp. 1139–1147.

[104] A. Graves, "Generating sequences with recurrent neural networks," *arXiv Prepr. arXiv1308.0850*, 2013.

[105] Y. Mann, "AI Duet." 2017.

[106] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *arXiv Prepr. arXiv1511.06349*, 2015.

[107] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 1017–1024.

[108] K. Choi, G. Fazekas, and M. Sandler, "Text-based LSTM networks for automatic music composition," *arXiv Prepr. arXiv1604.05358*, 2016.

[109] A. Huang and R. Wu, “Deep learning for music,” *arXiv Prepr. arXiv1606.04930*, 2016.

[110] A. J. Champandard, “Semantic style transfer and turning two-bit doodles into fine artworks,” *arXiv Prepr. arXiv1603.01768*, 2016.

[111] J. Jongejan, H. Rowley, T. Kawashima, J. Kim, and N. Fox-Gieg, “Quick, Draw!” 2017.

[112] G. C. Lab, “AutoDraw.” 2017.

[113] N. Davis, C.-P. Hsiao, K. Y. Singh, and B. Magerko, “Co-creative drawing agent with object recognition,” in *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2016.

[114] N. Davis, C.-Pi. Hsiao, K. Yashraj Singh, L. Li, and B. Magerko, “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*, 2016, pp. 196–207.

[115] D. Ha and D. Eck, “A Neural Representation of Sketch Drawings,” *arXiv Prepr. arXiv1704.03477*, 2017.

[116] T. Yonetsuji, “Paint Chaniner.” 2017.

[117] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman, “Controlling perceptual factors in neural style transfer,” *arXiv Prepr. arXiv1611.07865*, 2016.

[118] X. Ren, “Rethinking the Relationship between Humans and Computers.,” *IEEE Comput.*, vol. 49, no. 8, pp. 104–108, 2016.

[119] N. Babich, “5 Essential UX Rules for Dialog Design.” 2016.

[120] W. Knight, “The Dark Secret at the Heart of AI. MIT Technology Review.” 2016.

[121] U. Farooq, J. Grudin, B. Shneiderman, P. Maes, and X. Ren, “Human Computer Integration versus Powerful Tools,” in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2017, pp. 1277–1282.

[122] G. Tecuci, M. Boicu, and M. T. Cox, “Seven aspects of mixed-initiative reasoning: An introduction to this special issue on mixed-initiative assistants,” *AI Mag.*, vol. 28, no. 2, p. 11, 2007.

[123] J. Pavlik, “The impact of technology on journalism,” *Journal. Stud.*, vol. 1, no. 2, pp. 229–237, 2000.

[124] H. Hardt, “Newsworkers, technology, and journalism history,” *Crit. Stud. Media Commun.*, vol. 7, no. 4, pp. 346–365, 1990.

[125] L. Tsui, “Rethinking journalism through technology,” in *The Changing Faces of Journalism*, Routledge, 2009, pp. 63–65.

[126] A. Graefe, “Guide to automated journalism,” 2016.

[127] K. N. Dörr, “Mapping the field of algorithmic journalism,” *Digit. Journal.*, vol. 4, no. 6, pp. 700–722, 2016.

[128] T. Montal and Z. Reich, “I, robot. You, journalist. Who is the author? Authorship, bylines and full disclosure in automated journalism,” *Digit. Journal.*, vol. 5, no. 7, pp. 829–849, 2017.

[129] N. Science, “Narrative Science | Natural Language Generation Technology.” 2010.

[130] A. I. INC., “Automated Insights - Natural Language Generation.” 2007.

[131] @earthquakesLA, “LA QuakeBot.” May-2009.

[132] M. Carlson, “Automated journalism: A posthuman future for digital news?,” in *The Routledge companion to digital journalism studies*, Routledge, 2016, pp. 226–234.

[133] T. Lokot and N. Diakopoulos, “News bots: Automating news and information dissemination on Twitter,” *Digit. Journal.*, vol. 4, no. 6, pp. 682–699, 2016.

[134] N. Thurman, K. Dörr, and J. Kunert, “When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism’s capabilities and consequences,” *Digit. Journal.*, vol. 5, no. 10, pp. 1240–1259, 2017.

[135] K. McKeown and D. R. Radev, “Generating summaries of multiple news articles,” in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, 1995, pp. 74–82.

[136] J. W. Boumans and D. Trilling, “Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars,” *Digit. Journal.*, vol. 4, no. 1,

pp. 8–23, 2016.

[137] C. W. Anderson, “Towards a sociology of computational and algorithmic journalism,” *New media Soc.*, vol. 15, no. 7, pp. 1005–1021, 2013.

[138] N. S. Cohen, “From pink slips to pink slime: Transforming media labor in a digital age,” *Commun. Rev.*, vol. 18, no. 2, pp. 98–122, 2015.

[139] M. Carlson, “The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority,” *Digit. Journal.*, vol. 3, no. 3, pp. 416–431, 2015.

[140] A. Van Dalen, “The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists,” *Journal. Pract.*, vol. 6, no. 5–6, pp. 648–658, 2012.

[141] J. Jung, H. Song, Y. Kim, H. Im, and S. Oh, “Intrusion of software robots into journalism: The public’s and journalists’ perceptions of news written by algorithms and human journalists,” *Comput. Human Behav.*, vol. 71, pp. 291–298, 2017.

[142] N. Diakopoulos, “Accountability in algorithmic decision making,” *Commun. ACM*, vol. 59, no. 2, pp. 56–62, 2016.

[143] N. Diakopoulos, “Algorithmic-Accountability: the investigation of Black Boxes,” *Tow Cent. Digit. Journal.*, 2014.

[144] N. Diakopoulos, “Algorithmic accountability: Journalistic investigation of computational power structures,” *Digit. Journal.*, vol. 3, no. 3, pp. 398–415, 2015.

[145] E. Adar, C. Gearig, A. Balasubramanian, and J. Hullman, “PersaLog: Personalization of News Article Content,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 3188–3200.

[146] P. Tolmie *et al.*, “Supporting the use of user generated content in journalistic practice,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 3632–3644.

[147] N. Maiden *et al.*, “Making the News: Digital Creativity Support for Journalists,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, p. 475.

[148] D. Kim, “An Algorithmic Approach to Personalized and Interactive News Generation,” Seoul National University, 2017.

[149] D. Kim and J. Lee, “Designing an Algorithm-Driven Text Generation System for Personalized and Interactive News Reading,” *Int. J. Human--Computer Interact.*, pp. 1–14, 2018.

[150] A. Graefea, M. Haima, B. Haarmannb, and H.-B. Brosius, “Perception of Automated Computer-Generated News: Credibility, Expertise, and Readability.”

[151] C. Clerwall, “Enter the robot journalist: Users’ perceptions of automated content,” *Journal. Pract.*, vol. 8, no. 5, pp. 519–531, 2014.

[152] S. Flaxman, S. Goel, and J. M. Rao, “Filter bubbles, echo chambers, and online news consumption,” *Public Opin. Q.*, vol. 80, no. S1, pp. 298–320, 2016.

[153] A. Tuzhilin, “Personalization: The state of the art and future directions,” *Bus. Comput.*, vol. 3, no. 3, pp. 3–43, 2009.

[154] R. K. Chellappa and S. Shivendu, “Mechanism design for ‘free’ but ‘no free disposal’ services: The economics of personalization under privacy concerns,” *Manage. Sci.*, vol. 56, no. 10, pp. 1766–1780, 2010.

[155] G. Petriglieri, J. D. Wood, and J. L. Petriglieri, “Up close and personal: Building foundations for leaders’ development through the personalization of management learning,” *Acad. Manag. Learn. Educ.*, vol. 10, no. 3, pp. 430–450, 2011.

[156] B. P. S. Murthi and S. Sarkar, “The role of the management sciences in research on personalization,” *Manage. Sci.*, vol. 49, no. 10, pp. 1344–1362, 2003.

[157] H. Xu, X. R. Luo, J. M. Carroll, and M. B. Rosson, “The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing,” *Decis. Support Syst.*, vol. 51, no. 1, pp. 42–52, 2011.

[158] J. Vesanen and M. Raulas, “Building bridges for personalization: a process model for marketing,” *J. Interact. Mark.*, vol. 20, no. 1, pp. 5–20, 2006.

[159] H. Fan and M. S. Poole, “What is personalization? Perspectives on the design and implementation of personalization in information systems,” *J. Organ. Comput. Electron. Commer.*, vol. 16, no. 3–4, pp. 179–202, 2006.

[160] B. Mobasher, R. Cooley, and J. Srivastava, “Automatic personalization based on web usage mining,” *Commun. ACM*, vol. 43, no. 8, pp. 142–

151, 2000.

[161] M. A. Beam, “Automating the news: How personalized news recommender system design choices impact news reception,” *Communic. Res.*, vol. 41, no. 8, pp. 1019–1041, 2014.

[162] P. Brusilovski, A. Kobsa, and W. Nejdl, *The adaptive web: methods and strategies of web personalization*, vol. 4321. Springer Science & Business Media, 2007.

[163] B. Mobasher, “Data mining for web personalization,” in *The adaptive web*, Springer, 2007, pp. 90–135.

[164] M. Macdonald, “Rethinking personalization in current affairs journalism,” *Tabloid tales Glob. debates over media Stand.*, pp. 251–266, 2000.

[165] N. Thurman and S. Schifferes, “The future of personalization at news websites: lessons from a longitudinal study,” *Journal. Stud.*, vol. 13, no. 5–6, pp. 775–790, 2012.

[166] B. K. Rimer and M. W. Kreuter, “Advancing tailored health communication: A persuasion and message effects perspective,” *J. Commun.*, vol. 56, pp. S184--S201, 2006.

[167] A. J. Roberto, J. L. Krieger, and M. A. Beam, “Enhancing web-based kidney disease prevention messages for Hispanics using targeting and tailoring,” *J. Health Commun.*, vol. 14, no. 6, pp. 525–540, 2009.

[168] M. Prensky, “What can you learn from a cell phone? Almost anything!,” *Innov. J. Online Educ.*, vol. 1, no. 5, 2005.

[169] O. Westlund, “Mobile news: A review and model of journalism in an age of mobile media,” *Digit. Journal.*, vol. 1, no. 1, pp. 6–26, 2013.

[170] S. S. Sundar, “Multimedia effects on processing and perception of online news: A study of picture, audio, and video downloads,” *Journal. Mass Commun. Q.*, vol. 77, no. 3, pp. 480–499, 2000.

[171] M. G. Petersen, “Evaluating usability of multimedia interfaces,” *HCI MSc Proj. Report, Dep. Comput. Sci. Queen Mary Westf. Coll.*, 1996.

[172] P. Johnson and F. Nemetz, “Towards principles for the design and evaluation of multimedia systems,” in *People and Computers XIII*, Springer, 1998, pp. 255–271.

[173] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.

[174] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science (80-.).*, vol. 350, no. 6266, pp. 1332–1338, 2015.

[175] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

[176] R. Cellan-Jones, "Stephen Hawking warns artificial intelligence could end mankind," *BBC News*, vol. 2, 2014.

[177] R. Kurzweil, *The singularity is near: When humans transcend biology*. Penguin, 2005.

[178] B. S. Association, "One in three believe that the rise of artificial intelligence is a threat to humanity." 2015.

[179] S. Kulviwat, I. I. Bruner, C. Gordon, A. Kumar, S. A. Nasco, and T. Clark, "Toward a unified theory of consumer acceptance technology," *Psychol. Mark.*, vol. 24, no. 12, pp. 1059–1084, 2007.

[180] T.-C. Chien, "Factors Influencing Computer Anxiety and Its Impact on E-Learning Effectiveness: A Review of Literature.," *Online Submiss.*, 2008.

[181] DeepMind, "AlphaGo." 2016.

[182] B. G. Glaser and A. L. Strauss, *The discovery of grounded theory: Strategies for qualitative research*. Transaction publishers, 2009.

[183] O. Workshop, "Reframer." 2016.

[184] C. Nass, J. Steuer, E. Tauber, and H. Reeder, "Anthropomorphism, agency, and ethopoeia: computers as social actors," in *INTERACT'93 and CHI'93 conference companion on Human factors in computing systems*, 1993, pp. 111–112.

[185] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[186] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in

Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[187] D. Cire\csAn, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” *Neural networks*, vol. 32, pp. 333–338, 2012.

[188] I. Sutskever, O. Vinyals, and Q. V Le, “Sequence to sequence learning with neural networks,” in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[189] G. Mesnil *et al.*, “Using recurrent neural networks for slot filling in spoken language understanding,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 3, pp. 530–539, 2015.

[190] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, “Draw: A recurrent neural network for image generation,” *arXiv Prepr. arXiv1502.04623*, 2015.

[191] G. W. Smith and F. F. Leymarie, “The Machine as Artist: An Introduction,” in *Arts*, 2017, vol. 6, no. 2, p. 5.

[192] LabintheWild, “Augury Standalone.” 2018.

[193] H. Talebi and P. Milanfar, “Nima: Neural image assessment,” *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, 2018.

[194] M. W. Van Someren, Y. F. Barnard, and J. A. C. Sandberg, “The think aloud method: a practical approach to modelling cognitive,” 1994.

[195] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006.

[196] C. Chen, A. Seff, A. Kornhauser, and J. Xiao, “Deepdriving: Learning affordance for direct perception in autonomous driving,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2722–2730.

[197] M. Bojarski *et al.*, “End to end learning for self-driving cars,” *arXiv Prepr. arXiv1604.07316*, 2016.

[198] H. Greenspan, B. van Ginneken, and R. M. Summers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[199] Y. Xu *et al.*, "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*, 2014, pp. 1626–1630.

[200] A. Kittur *et al.*, "The future of crowd work," in *Proceedings of the 2013 conference on Computer supported cooperative work*, 2013, pp. 1301–1318.

[201] E. Wenger, *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann, 2014.

[202] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *arXiv Prepr. arXiv1508.06576*, 2015.

[203] M. Ren, R. Kiros, and R. Zemel, "Exploring models and data for image question answering," in *Advances in neural information processing systems*, 2015, pp. 2953–2961.

[204] T. Winograd, "Shifting viewpoints: Artificial intelligence and human--computer interaction," *Artif. Intell.*, vol. 170, no. 18, pp. 1256–1258, 2006.

[205] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell, "Explaining the user experience of recommender systems," *User Model. User-adapt. Interact.*, vol. 22, no. 4–5, pp. 441–504, 2012.

[206] A. Purington, J. G. Taft, S. Sannon, N. N. Bazarova, and S. H. Taylor, "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*, 2017, pp. 2853–2859.

[207] A. Xu, Z. Liu, Y. Guo, V. Sinha, and R. Akkiraju, "A New Chatbot for Customer Service on Social Media," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 3506–3510.

[208] L. McCarthy, "p5.js." 2017.

[209] W. Albert and T. Tullis, *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.

[210] A. M. Lund, "Measuring Usability with the USE Questionnaire12," *Usability interface*, vol. 8, no. 2, pp. 3–6, 2001.

[211] M. Hartmann, “Challenges in Developing User-Adaptive Intelligent User Interfaces.,” in *LWA*, 2009, p. ABIS--6.

[212] A. D. Jameson, “Understanding and dealing with usability side effects of intelligent processing,” *AI Mag.*, vol. 30, no. 4, p. 23, 2009.

[213] C. Lewis and J. Rieman, “Task-centered user interface design,” *A Pract. Introd.*, 1993.

[214] J. Collier, “Photography in anthropology: a report on two experiments,” *Am. Anthropol.*, vol. 59, no. 5, pp. 843–859, 1957.

[215] B. Glaser, *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017.

[216] R. Hartson and P. S. Pyla, *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier, 2012.

[217] P. S. Earle, D. C. Bowden, and M. Guy, “Twitter earthquake detection: earthquake monitoring in a social world,” *Ann. Geophys.*, vol. 54, no. 6, 2012.

[218] I. O. Committee, “PyeongChang Olympics | Next Winter Games in Korea.” 2018.

[219] C. Scanlan, “Birth of the inverted pyramid: A child of technology, commerce and history,” *Poynter. Retrieved from poynter. org*, 2003.

[220] M. Errico, J. April, A. Asch, L. Khalfani, M. Smith, and X. Ybarra, “The evolution of the summary news lead,” *Media Hist. Monogr.*, vol. 1, no. 1, 1997.

[221] M. Schudson, “The sociology of news production,” *Soc. meanings news A text-reader*, vol. 722, 1997.

[222] G. Hart, “The Five Ws of Online Help,” *TECHWR-L, Retrieved April*, vol. 30, 2012.

[223] N. Developers, “Clova Speech Synthesis (CSS) - Naver Developers.” 2017.

[224] S. S. Sundar, “Exploring receivers’ criteria for perception of print and online news,” *Journal. Mass Commun. Q.*, vol. 76, no. 2, pp. 373–386, 1999.

[225] L. Festinger, *A theory of cognitive dissonance*, vol. 2. Stanford university

press, 1962.

[226] W. Uricchio, “The algorithmic turn: Photosynth, augmented reality and the changing implications of the image,” *Vis. Stud.*, vol. 26, no. 1, pp. 25–35, 2011.

[227] M. B. N. Hansen, *Feed-forward: On the Future of Twenty-first-century Media*. University of Chicago Press, 2015.

[228] J. Markoff, “How Tech Giants Are Devising Real Ethics for Artificial Intelligence,” *New York Times*, 2016.

[229] G. P. R. (PAIR), “What-If Tool.” 2018.

[230] G. Loewenstein, “The psychology of curiosity: A review and reinterpretation.,” *Psychol. Bull.*, vol. 116, no. 1, p. 75, 1994.

[231] E. Law, M. Yin, J. Goh, K. Chen, M. A. Terry, and K. Z. Gajos, “Curiosity killed the cat, but makes crowdwork better,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 4098–4110.

[232] C. Oh, J. Song, J. Choi, S. Kim, S. Lee, and B. Suh, “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*, 2018, p. 649.

[233] T. Leong, S. Howard, and F. Vetere, “Choice: abidcating or exercising?,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 715–724.

[234] B. Shneiderman *et al.*, “Creativity support tools: Report from a US National Science Foundation sponsored workshop,” *Int. J. Hum. Comput. Interact.*, vol. 20, no. 2, pp. 61–77, 2006.

[235] L. Swartz, “Why people hate the paperclip: Labels, appearance, behavior, and social responses to user interface agents,” Stanford University Palo Alto, CA, 2003.

[236] S. N. K. Watt, “Artificial societies and psychological agents,” in *Software Agents and Soft Computing Towards Enhancing Machine Intelligence*, Springer, 1997, pp. 27–41.

[237] E. L. Cohen, “Online journalism as market-driven journalism,” *J. Broadcast. Electron. Media*, vol. 46, no. 4, pp. 532–548, 2002.

[238] D. S. Chung, “Profits and perils: Online news producers’ perceptions of interactivity and uses of interactive features,” *Convergence*, vol. 13, no. 1, pp. 43–61, 2007.

[239] B. J. Dietvorst, J. P. Simmons, and C. Massey, “Algorithm aversion: People erroneously avoid algorithms after seeing them err.,” *J. Exp. Psychol. Gen.*, vol. 144, no. 1, p. 114, 2015.

논 문 초 록

컴퓨팅 파워의 개선, 인터넷과 소셜미디어, 모바일 디바이스 등의 보급을 통한 수많은 데이터의 축적, 딥러닝을 비롯한 기계학습 알고리즘의 발전으로 인공지능 기술이 어느때보다 더욱 큰 성과를 보이고 있다. 음성 인식, 컴퓨터 비전, 자연어 처리 등의 분야에서 인공지능은 이미 인간에 필적하거나 혹은 인간을 뛰어넘는 성능을 보이고 있으며, 자율주행, 로봇, 의료서비스 등의 다양한 분야에 적용되어 우리의 삶에 많은 변화를 가져올 것으로 기대된다.

하지만 알고리즘 측면에서의 기술적인 발전에 비해 인공지능의 인간공학적 요소와 사용자 경험에 대한 관심과 논의가 상대적으로 부족한 편이다. 이에 이 연구는 인간컴퓨터상호작용의 관점에서 인공지능과 사용자가 상호작용 하는 방식에 대해 다층적이고 통합적으로 이해하는 것을 목표로 하고 이를 통해 인공지능 기반의 사용자 인터페이스 디자인을 위한 함의점을 도출하는 것을 목표로 한다. 특히 이 논문은 인공지능 기술을 이용한 알고리즘 기반의 시스템과 사용자의 상호작용에 주목하고, 이를 대상으로 인지, 해석 및 평가, 지속적인 인터랙션, 실용적인 어플리케이션을 주제로 한 네 단계의 연구를 기획하고 진행하였다.

첫번째 연구는 인공지능 알고리즘에 대한 사람들의 선험적 인식을 조사하였다. 연령과 성별, 직업의 다양성을 고려하여 인구통계학적 대표성을 갖는 참가자를 모집하였으며, 이들을 대상으로 인공지능 인식에 대한 정성적 방식의 조사를 진행하였다. 조사 결과 사람들이 인공지능 알고리즘에 대해 갖는 선입견과 고정관념을

확인할 수 있었으며, 사람들이 인공지능을 의인화 할 뿐만 아니라 타자화 하는 경향이 있음을 확인할 수 있었다. 또한 인공지능 알고리즘과 사용자의 관계에서 지속적이고 전체적인 경험이 중요함을 확인하였다.

두번째 연구는 인공지능 알고리즘에 대한 사용자의 해석과 평가에 관한 것이다. 이를 위해 이미지의 미적 점수를 계산해주는 신경망 기반의 알고리즘이 구현된 AI Mirror라는 연구 프로토타입을 제작하였으며, 인공지능/기계학습 분야의 전문가, 사진전문가, 일반인으로 구분된 세 집단의 사용자를 모집하여 실험을 진행하였다. 사용자는 저마다 다른 배경 지식을 반영해 인공지능 알고리즘을 해석하고 평가하는 경향을 보였다. 사진전문가 집단이 알고리즘을 가장 높은 정도로 해석하였으며 합리적이라고 여긴 반면, 인공지능/기계학습 전문가 집단은 가장 낮은 정도로 알고리즘을 해석하고 평가했다. 사용자는 다양한 전략을 통해 인공지능 알고리즘의 원리를 추론하고자 하였으며 이를 통해 인공지능 알고리즘과의 차이를 좁혀갈 수 있었다. 또한 사용자는 인공지능 알고리즘과 쌍방 소통을 통해 의견을 교환하고자 하는 니즈를 표출하였다.

세번째 연구는 인공지능 알고리즘과 사용자가 공동의 목표를 두고 지속적인 인터랙션을 이어가는 과정에 대한 이해를 목표로 하였다. 사용자가 일부 그린 물체를 완성하고 스케치에 색칠을 자동으로 완성해주는 신경망 기반의 알고리즘 API를 이용하여 DuetDraw라는 리서치 프로토타입을 제작하였고, 정량 및 정성적 방법으로 이에 대한 사용자 평가를 진행하였다. 사용자 평가 결과 사용자는 인공지능 알고리즘과의 협업 과정에서 인공지능으로부터 단순한 피드백 보다는 자세한 설명을 제공받기를 원했으며, 알고리즘과의 관계에서 항상 주도적인 위치에 있고자 하였다. 인공지능과의 인터랙션은 과업 수행에 대한 사용자의 예측가능성, 이해도, 통제력을 낮추는 경향이 있었지만, 사용자에게 상대적으로 높은 사용성을 제공하였을 뿐만 아니라 사용자가 전반적으로 만족스러운 경험을 할 수 있도록 하였다.

끝으로, 네번째 연구는 보다 실용적인 어플리케이션을 제작하여 이에 대한 사용자 인터랙션을 이해하고자 하였으며, 이에 최근 큰 각광을 받고 있는 로봇저널리즘 기술을 구현한 NewsRobot을 제작하였다. NewsRobot은 2018 평창동계올림픽의 주요 경기 결과를 자동으로 수집하고 요약하며, 내용과 형식을 각각 종합뉴스-선택뉴스, 텍스트-카드-동영상으로 달리하여 뉴스를 생성한다. 정량 및 정성적 방법의 사용자 평가 결과, 선택뉴스가 종합뉴스에 비해 낮은 신뢰도를 보였음에도 불구하고 선택뉴스에 대한 사용자의 높은 선호도를 확인할 수 있었다. 또한 멀티미디어 모달리티가 높아질수록 사용자의 뉴스에 대한 만족도가 높아지지만 사용자의 기대수준에 어긋난 경우 오히려 낮은 평가를 받는 것을 확인하였다. 사용자는 알고리즘이 자동으로 생성한 뉴스에 대해 정확하고 객관적이라고 평가하였으며, 빠른 뉴스 생성 속도와 다양한 정보 시각화 요소에 대해서도 만족감을 드러냈다.

본 연구는 이 네 가지 연구의 결과들을 바탕으로 인간-인공지능 상호작용에 대한 다양한 시사점들을 도출하였으며, 인공지능을 이용한 알고리즘 기반의 시스템의 사용자 인터페이스 디자인을 위한 함의점들을 제안한다.

주요어: 인공지능, 인간-인공지능 인터랙션, 인공지능을 이용한 알고리즘 기반의 시스템, 인간-컴퓨터 상호작용, 알고리즘 경험

학 번: 2012-22466