Exploring the Experiences of UX Practitioners Using ChatGPT: A Cross-cultural Study

ANONYMOUS AUTHOR(S)

Rapid advancements in artificial intelligence have led to the development of powerful large language models that have the potential to dramatically alter professionals' UX work due to the capabilities of the technology, including automatically identifying and fixing bugs, producing interface text and content, and generating intelligent responses to design questions. However, little is known about the experiences of UX practitioners using ChatGPT during their work. To fill in this gap, we conducted semi-structured interviews with 30 UX practitioners from two cultures: South Korea and the United States. We identified how and why ChatGPT was used to support their UX work in the two cultures. Our study reveals that Korean participants tended to struggle in generating prompts for ChatGPT, while US participants were concerned about sharing confidential information with ChatGPT. From these results, we propose design opportunities for creating tools to address these and other discovered concerns to better support practitioners.

CCS Concepts: • **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Additional Key Words and Phrases: Do, Not, Us, This, Code, Put, the, Correct, Terms, for, Your, Paper

ACM Reference Format:

Anonymous Author(s). 2018. Exploring the Experiences of UX Practitioners Using ChatGPT: A Cross-cultural Study. In . ACM, New York, NY, USA, 15 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Rapid advancements in artificial intelligence (AI) have led to the development of powerful large language models (LLM) such as OpenAI's GPT-4 and Google's Bard. However, the cost of training large language models has been increasing over the years as the number of training data points increases and the models become more complex [29]. There are several methods for reducing costs. One method is prompt adaption (e.g., prompt selection, query concatenation), which allows researchers to reduce the cost of using LLMs by reducing the size of the prompt. Moreover, LLM approximation (e.g., completion cache, model fine-tuning) allows researchers to reduce the cost by utilizing more affordable models when LLM API is expensive to utilize. Additionally, LLM cascades allow researchers to reduce the cost of using LLM by sending a query to LLM APIs sequentially so the query does not have to cycle through every LLM API. Researchers proposed FrugalGPT, which allows researchers to reduce the cost of using LLMs by adaptively using combinations of LLMs for different queries [7]. However, some limitations of FrugalGPT include requiring labeled examples for training the LLM cascade strategy and the LLM cascade learning requiring additional resources.

LiGO (Linear Growth Operator) is another novel technique developed by researchers at MIT to reduce the computational cost of training LLMs by 50% [41]. This method maintains the performance benefits of larger models with reduced computational cost and training time compared to training a large model from scratch. However, while LiGO aims to reduce computational costs, questions remain about whether the models trained with this method achieve the

Manuscript submitted to ACM

same level of performance as models trained conventionally. There might be trade-offs in terms of model quality or accuracy.

Declining costs can lead to wider adaptation of LLMs across many other industrial and academic domains. A wide range of limitations in fields from education to healthcare are being reduced by introducing LLMs. For example, in healthcare, LLMs provide numerous possibilities to elevate patient care, optimize medical workflows, augment the overall healthcare experience, a holistic understanding of each individual while offloading the public health workload, and help mitigate loneliness and emotional burdens [36]. LLMs can also be used as conversational agents for young children and on-call facilitators for educators and caregivers in early childhood education [24]. Moreover, students admire the capabilities of ChatGPT and find it interesting, motivating, and helpful for study and work. They find it easy to use and appreciate its human-like interface that provides well-structured responses and good explanations [38].

However, the use of LLMs has not been extensively examined in the field of human-computer interaction. Examining the use of LLMs in the field of HCI is important for several reasons. First, UX Practitioners (UXP) are the people that take AI out of the lab and into real world products. Although prior Responsible AI (RAI) research in CHI and CSCW has focused on opportunities to intervene in potentially harmful AI development practices, it has, with few exceptions, focused on the work practices of data scientists and machine learning (ML) engineers in the model development process rather than the wider set of practitioners involved in designing for and applying existing models for novel AI applications [42]. As a whole, the UXP population is not currently knowledgeable about the way AI works from a technical perspective. However, they are a crucial element in the process of building human-centered AI and a deeper understanding between design and development must happen to prevent haphazard application of models [25]. Third, UXPs need to deeply understand the technology they are implementing in order to provide good user experiences [9]. One such experience that must be designed are interfaces that utilize large language models.

Additionally, our understanding of LLMs across cultures is also limited. With greater usage in LLMs, the LLM technology that we use may be similar, however the application may be different. LLM-based chatbots used ubiquitously around the world (e.g., ChatGPT) may be nudging users toward adapting similar behaviors across cultures [28]. On the other hand, differing cultural contexts combined divergent user needs may be leading people to use LLM-based chatbots differently across cultures citepark2022cross. Through comparing different cultures, cross-cultural studies allows researchers to find the general truth among the cultures and validate that truth [11].

Prior studies have shown how UXPs use various tools in their UX practices, such as [13, 33, 43]. However, no studies have looked into how UX practitioners use technology such as LLM-based chatbots, such as ChatGPT, in their work and design processes. Prior studies have shown the opportunities for LLM-based chatbots to assist in various fields, such as public health intervention [18], self-diagnosis [37]and education [24] But, those studies mostly investigated how LLM-based chatbots are being used in a field other than HCI. Cross-cultural studies have also looked into addressing the differing experiences on mobile applications and websites [2, 30, 34, 35]. However, none of the cross-cultural studies address the various experiences of LLM-based chatbots.

We conducted a cross-cultural, semi-structured interview study to build an understanding of how UX practitioners are currently using ChatGPT in their processes and their experiences of using ChatGPT in their work as design and HCI professionals. We recruited participants from the United States (N=15) and South Korea (N=15)—representing Western and Eastern cultures respectively—to improve our understanding of cross-cultural nuances in particular. We aimed to answer the following research questions:

- RQ 1. What are the UX practitioners' experiences of using ChatGPT in each culture?

- RQ 2. What are UX practitioners' challenges in using ChatGPT in each culture?

From our interviews, we surface qualitative findings of what UX tasks UXPs use ChatGPT for, why UX practitioners use ChatGPT, how UX practitioners use ChatGPT, and some of the challenges that they face in using ChatGPT today. Synthesizing these findings, we highlight the similarities and differences between the two cultures. We interpret these findings in relation to prior works and derive design implications. Our study makes the following contributions to the CHI community:

- We provide insights into how people interact with ChatGPT in different cultural contexts.
- We understand the use cases and challenges of leveraging LLM-based chatbots in UX work through surveys and interviews.
- We provide a description of how ChatGPT is actually used in practice.
- We highlight opportunities for AI guidelines and resources to better address UX practitioners' needs.

2 RELATED WORK

This section summarizes previous studies on the experiences of UX practitioners on various tools and diverse groups of people on LLM-based chatbots. After reviewing these studies, we present their limitations and explain the necessity of further research to support the experiences of UX practitioners on LLM-based chatbots.

2.1 User Experiences Among UX Practitioners

2.1.1 Collaborative tools for UX Practitioners. Prior studies explored UX practitioners' experience on collaborative tools for supporting their work [13, 33, 43].While Feng et al. and Williams et al. explored UX practitioners' experience on existing collaborative tools, Remy et al. developed a new prototype for affinity diagramming and evaluated it with the practitioners. For instance, Feng et al. [13] conducted surveys with UX practitioners to understand their collaborative practices and tools. Feng et al. reported various collaborative tools used in five common categories of UX activities [3] and found out that all-in-one collaborative tool called Figma is the most used tool among the five categories. Williams et al. [43] interviewed HCI researchers to understand their online science communication through participatory web channels (e.g., social media). They identified the motivations and challenges of researchers when using social media to promote their own work. On the other hand, Remy et al. [33] specifically focused on exploring UX practitioners' experience for affinity diagramming by developing a web-based affinity diagramming prototype and conducting a user experience study. The participants found both effectiveness and issues regarding the prototype, which lead to authors proposing design suggestions for improving their prototype.

2.1.2 Non-collaborative tools for UX Practitioners. Other four studies investigated the experience of UX practitioners regarding using non-collaborative tools [6, 26, 32, 45]. While three studies [6, 32, 45] explored UX practitioners' experience on existing non-collaborative tools, one study [26] developed a system allowing designers to detect failure patterns early in the design process while exploring the model behavior. For example, Yildirim et al. [45] conducted interviews with UX practitioners to understand how they use guidelines from the People + AI Guidebook when designing AI-enabled products. They found that the practitioners use the guidelines when struggle to design with AI but also for other purposes (e.g., for education, developing internal resources, cross-functional alignment, buy-in within their teams and organizations). Çerçi et al. [6] also conducted interviews with design researchers to learn about their experience on working with Probes [14]. They were able to understand why and how the researchers have been using Probes for their design practice, which revealed the benefits and challenges for utilizing it. Prpa et al. [32] assessed an interview method called a micro-phenomenology based on the interviews with HCI/Design experts, showing both values and difficulties of using it and generation for broader insights through combination with other methods. On the other hand, Moore et al. [26] developed fAIlureNotes, a system that allows designers to explore model behavior and failure patterns prior in the design process. The researchers observed that the system is user-centered, useful, and intuitive via user study with UX experts.

2.2 User Experiences in LLM-Based Chatbots

Prior studies explored UX practitioners' experiences with collaborative and non-collaborative tooling [13, 33, 43]. Here, we review studies done with broader, non-UX practitioner populations and LLM-based chatbots. Three studies [15, 18, 37] explored user experiences with and for patients and clinicians in healthcare settings, seven separate studies [5, 16, 20–22, 24, 39] looked at the experiences of chatbots or ChatGPT in educational and learning contexts. Other studies [27, 46] investigated the user experiences of LLM-powered tools for nonspecific, general audiences.

In the healthcare context, Goktas et al. explored the applications, benefits, and challenges of GPT-4 in clinical settings. They found that while there are advantages to using chatbots in clinical settings, such as increased patient engagement, improved diagnostic accuracy, personalized treatment plans, and the ability to ingest the latest medical literature, it also has drawbacks. Clinicians and AI engineers must be aware of the limitations, which include checks for the accuracy of information. Chatbots are also not a substitute for the critical thinking and judgment skills inherent to human clinicians. Shahsavar & Choudhury conducted a survey-based study to understand the usage of ChatGPT in self-diagnostic health contexts. They found that users are increasingly using ChatGPT for answers to health-related questions and suggest that researchers should investigate any long-term effects of using ChatGPT for health advice as well as explore the integration of ChatGPT into already-established healthcare scenarios. Jo et al. showed that the CareCall LLM-based chatbot offered healthcare workers a holistic understanding of each individual patient in their care. CareCall also offloaded the public health workload while allowing patients to mitigate loneliness and other emotional burdens. However, there are some challenges in using chatbots to perform work that can be better-served by more traditional task-oriented governmental systems.

LLM-based chatbot user experiences have been explored in educational contexts and for learning [5, 16, 20–22, 24, 39]. Shoufan investigated ChatGPT in a learning environment, asking senior students in a computer science program to assess the helpfulness of ChatGPT in completing certain programming-related tasks. ChatGPT did not always give correct answers and did not replace human intelligence, according to the participants in the study. Luo et al. spoke with expert professors from China and the United States to ask their opinion of the role of ChatGPT in early childhood education. ChatGPT was found to be a good conversational companion for young learners as well as an assistive, on-demand tool for educators. Cao explored how LLM-based games could help international students increase their feelings of belonging within first-year computer science courses. It was found that GPT-3-backed games could help scaffold learning by providing adaptive and personalized feedback and reduce anxiety around programming. Leinonen et al. used LLM-based explanations of coding examples and compared these against examples created by the students themselves. It was found that LLM-created code explanations were easier to understand and were more accurate. Lee et al. explored "strategies to elicit desired outputs" from a GPT model through a debate game. The team found that by providing specific instruction to the game or creating case-based prompts improved the accuracy and relevance of ChatGPT's answers. Ho and Lee used chatGPT to help craft educational experiences within the Roblox metaverse. The team found that by injecting chatGPT into the educational process, students' engagement and understanding increased. Lin explored how to use chatGPT as a virtual tutor in self-directed learning environments. Lin found that chatGPT could help adult learners set learning goals, locate resources and monitor performance. However, there are limitations related to providing guidance on building effective prompts, the possibility that learners may become overly reliant on chatGPT, and the uncertainty of policies related to AI usage in learning contexts.

Studies have also evaluated the user experience of LLM chatbots and general audiences [27, 46]. Park et al. built and evaluated Social Simulacra, a prototyping tool for simulating real-world interactions inside social computing systems, such as online forums. It was found that participants were unable to detect a difference between the Social Simulacra and real online interactions, showing that social computing designers could use the tool to effectively test any proposed design changes. Zamfirescu-Pereira et al. investigated how non-AI-experts can successfully engineer prompts for LLMs. Struggles with creating effective prompts stemmed from over-generalizing of single observations of success or failure and models of human-computer interaction that were rooted in human-human interactions.

2.3 UX Practitioners' Experiences in ChatGPT

2.4 Limitations of prior studies

Based on the prior studies, we found that the experiences of UX practitioners with ChatGPT are still under-examined. Prior studies looked into how UX practitioners use collaborative and non-collaborative tools. However, the focus of those studies was frequently limited to the cases where UX practitioners interacted with tools that were not related to LLM technology. Similarly, prior studies mostly investigated how LLM-based chatbots are being used in a field other than HCI. Our study aims to extend the understanding of UX practitioners' experiences in an LLM-based chatbot by examining it in a cross-cultural context. With a better understanding of their experiences with ChatGPT, we propose design opportunities for technology to support them.

2.5 Characterizing UX Work

The field of UX (user experience) has seen tremendous growth in recent history [12] and is expected to remain on an upward trajectory as more companies hire UX professionals, UX teams continue to grow and specialize, and more countries adapt UX practices. This growth has led researchers to seek scaffolding and frameworks to characterize the work of UX professionals for a variety of purposes. Below, we briefly review two popular frameworks for characterizing UX activities that have been used widely by the field and offer our reasoning for choosing one framework to frame our study.

One design process framework that is widely used is the Double Diamond Design Process, launched by the British Design Council 2003. The Double Diamond contains four phases of UX work: Discover, Define, Develop, and Deliver [1]. The visual is meant to represent the processes of divergent and convergent thinking as products are developed, although the BDC cautions that the process of designing may not be as linear as the arrows suggest. Figure A shows a visual representation of the process.

Perhaps the most described framework for framing a UX workflow is the Design Thinking Process made popular by the Stanford d.school. The Design Thinking Process describes 5 steps: Empathize, Define, Ideate, Prototype, and Test [10]. In the [TABLE B], we describe each step and common activities that would be performed during each step. Due to the popularity of the framework and the distinctness between the different steps, we chose this process to frame the conversations with our participants. Figure B shows a visual representation of the Design Thinking Process.

There is a great amount of elasticity in UX job titles, responsibilities, and situations of UX within the greater organization. Most importantly for our study, the popularity of the Design Thinking Process provides a shared language among UX designers and researchers.

3 METHOD

The goal of the study is to explore the ways in which UX Practitioners use LLM-based chatbots in their practice. To accomplish this, we conducted a survey and semi-structured interviews with UX Practitioners. We understood their HCI practices, motivations, challenges, and needs. Specifically, we heard UX practitioners' reflexive accounts of their work and uncovered explicit expressions of the critical stances and viewpoints held by those who utilized ChatGPT in their UX research and design work. In the following, we provide a more detailed description of our research design. We provide our approach to recruiting participants, the study procedure, and the methods used for data collection and analysis.

3.1 Participants

We recruited participants using social media and snowball sampling [][1]. We contacted an initial set of participants through our personal connections with UX practitioners in industry and academia. We then asked them to share any relevant contacts to grow our set of participants. In parallel, we recruited through online platforms (e.g., Linkedin, Facebook, Twitter, and Slack-based communities). We used a brief screener form to help us target suitable participants. We paid attention to three aspects during recruitment. All participants had to 1) have at least 3 months of professional internship, research, or professional work experience in the field of HCI/UX over the past year, 2) have knowledge of common UX design processes, and 3) have prior experience using ChatGPT when conducting HCI/UX tasks at work. The above process resulted in the recruitment of a total of 30 UX practitioners who have used ChatGPT in their practices. 15 Korean participants were conducted in their native language - Korean - in South Korea. 15 American participants were conducted in English in the United States. All 30 participants were informed of the purpose and nature of the study before the interview and voluntarily agreed to participate.

3.2 Study Procedure

We asked participants to fill out a web-based survey form to collect demographic data before interviewing each participant. The survey aimed to collect participants' demographic information and their ChatGPT experience, including what steps of the design thinking process they use ChatGPT for and which devices they use ChatGPT.

- Basic details about interview participants (e.g., How old are you? What is your gender? Where did you receive your professional training in HCI/UX?)
- ChatGPT usage (e.g., At what stages of the design process have you used ChatGPT? How often do you use ChatGPT? How long do you typically use ChatGPT each time?)
- Patterns of using ChatGPT (e.g., Which device do you primarily use when using ChatGPT? In what ways do you normally access ChatGPT? On what types of subscription plans have you used ChatGPT?)

We conducted semi-structured interviews. The interviewer asked follow-up questions or skipped some questions in the guidelines when appropriate given the content and context of the interview. In each interview, we asked participants about their roles, practices, and workflows on HCI/UX projects to gain background information and context. We asked them about the use of an LLM-based chatbot through a recent case in which they incorporated the chatbot into their work. Some of the interview questions include the following:

- Why do UX practitioners use ChatGPT? (e.g., Why do you use ChatGPT when asking the question that you mentioned earlier? What information do you seek from ChatGPT when asking that question?)
- What HCI/UX tasks do UX practitioners use ChatGPT for? (e.g., What HCI tasks do you perform when using ChatGPT?)
- How do UX practitioners interact with ChatGPT? (e.g., What kind of prompts or questions do you ask ChatGPT? Do you have prompts or questions that you ask ChatGPT frequently when trying to find information? How do you generate the question? Do you have any preferences for how ChatGPT should answer your question?)
- What challenges do UX practitioners face when using ChatGPT? (e.g., What do you do when you don't get the answer that you are looking for? What AI tools, search engines, or other secondary sources of information did you use before ChatGPT to help you with your work?)

Each interview was conducted using the Zoom video conference tool. Each interview lasted approximately 30-45 minutes. At the beginning of the interviews, the participants provided verbal consent to record their interview audio and use their data anonymously for research purposes only. The recordings were deleted when the data analysis was completed. Each participant received a $15 gift card as an appreciation for their time. The study was approved by the Institutional Review Board.

3.3 Data Analysis

We used descriptive statistics to analyze survey responses. We calculated the numbers and percentages of responses to analyze quantitative data from the survey questions. These statistics offered initial insights into the participants' ChatGPT usage patterns, including frequency of use and duration.

We analyzed the qualitative data from the interviews by conducting a thematic analysis. We transcribed the conversation with each interview participant using transcription software and then corrected the generated transcripts by listening to the recorded files. We created interview transcripts which were written separately for each team. Interviews with Korean participants were transcribed by the South Korean team and interviews with American participants were transcribed by the USA team. We removed every piece of information identifying individual participants and assigned new identifications.

- Korean participants were given the following identifications: KR01, KR02
- USA participants were given the following identifications: US01, US02

We used an open coding method to analyze the qualitative data. Researchers from each team independently read all of the transcripts. Two researchers from the South Korea team independently read all of the transcripts and two researchers from the USA team independently read all of the transcripts. For each interview transcript, two researchers highlighted statements that included important or meaningful information by looking at the transcripts independently. The highlighted remarks included sentences that answered our research questions or expressed the participants' impressions and experiences. After reviewing the highlighted statements, the same two researchers marked each statement with an appropriate code independently. To ensure the reliability of the intercoder, we developed a codebook, which includes codes that answer the research questions (e.g., [WHY][1] UP99 uses ChatGPT in the empathize step of the design thinking because XXX). All codes were written in English for both teams. Then, through repeated discussions, the two researchers came to an agreement on which code should be included, removed, or revised.

We merged all codes in Miro. All codes, including the codes from the South Korean participants and American participants, were merged in the Miro Dashboard. We grouped codes into themes for 1) each research question and 2) each step of the design-thinking process. We conducted an affinity diagramming workshop session to identify key insights, themes, and patterns that occurred in the interview data repeatedly. Used the online collaboration tool Miro to group the codes into themes based on the common properties, each of which corresponded to a research question we aimed to answer. The workshop session was conducted with the South Korean team and the USA team via Zoom to group the codes into themes.

4 FINDINGS

4.1 Motivations for using ChatGPT

Six Participants use ChatGPT because participants can get information that they want easily without the need to sift through online resources. Two participants (KR15, US03) from both Korea and US use ChatGPT because participants can get recommendations on a problem. All from Korea, three participants (KR01, KR11, KR12) use ChatGPT because participants can get personas, while two participants (KR13, KR14) use ChatGPT because participants can get knowledge about computers/software. Six Participants from both Korea and US use ChatGPT because participants can perform tasks time-efficiently. Three participants (KR07, US08, US09) use ChatGPT because participants can give their tasks to ChatGPT quickly. Another three participants (KR12, US01, US10) use ChatGPT because participants can get ideas quickly. Four participants from both Korea and US use ChatGPT because participants can analyze and summarize large amounts of information effectively. Four out of five participants (KR14, US05, US09, US10) use ChatGPT because participants can analyze and summarize large amounts of information effectively. Three participants from both Korea and US use ChatGPT because participants can provide insights into ideas that an participants have overlooked, haven't thought of before, or have missed. Two participants (KR11, US01) use ChatGPT because participants can double check if there are any information missing. The other two participants (KR12, US01) use ChatGPT because participants can get ideas that participants haven't thought of.

4.2 Use case of ChatGPT

18 participants used ChatGPT in empathize step of the design thinking process. 12 participants (KR01, KR03, KR11, KR12, US01, US06, US07, US08, US09, UP10, UP11, US15) from both Korea and US use ChatGPT in the empathize step before collecting data. Seven participants (KR07, KR14, UP02, US05, US09, US10, US13) use ChatGPT in the empathize step after creating interview transcripts. Only one US participant (US15) uses ChatGPT in the empathize step to generate research questions before collecting data. Four participants (KR06, US01, US11, US15) from both Korea and US use ChatGPT in the define step of the design thinking process. They all use ChatGPT when defining a problem statement. 13 participants from both US and Korea use ChatGPT in the ideate step of the design thinking process. 11 participants (KR06, KR09, KR11, KR12, KR13, KR15, US01, US03, US06, US10, US12) use ChatGPT in the ideate step to brainstorm ideas and options when generating solutions, while three participants (KR08, US03, US04) use ChatGPT in the ideate step after generating solutions. Eight participants from both Korea and US uses ChatGPT in prototype step of the design thinking process. Two participants (KR11, US08) use ChatGPT in the prototype step to determine what features should be included in the service before building prototypes. Only five Korean participants (KR02, KR05, KR07, KR13, KR14) said they use ChatGPT in the prototype step when coding. Only one US participant (US09) use ChatGPT in the prototype step to generate writing contents for mid to high fidelity mockups. Three participants (KR10, US04, US05)

ID	*Age*	*Gender*	*Frequency*	*Length*	*Subscription Plan*
KR01	35	Male	Once a week	1 to 2 hours	Paid
KR02	28	Female	Once a week	30 minutes to 1 hour	Free
KR03	25	Female	Once a week	Less than 30 minutes	Paid
KR04	25	Female	1-3 days a month	30 minutes to 1 hour	Free
KR05	37	Male	1-3 days a month	Less than 30 minutes	Free
KR06	31	Male	1-3 days a month	2 to 4 hours	Free
KR07	27	Female	Multiple times a day	Less than 30 minutes	Paid
KR08	24	Female	Once a week	30 minutes to 1 hour	Free
KR09	28	Male	Once a week	30 minutes to 1 hour	Free
KR10	29	Male	Once a week	30 minutes to 1 hour	Free
KR11	49	Female	Once a week	30 minutes to 1 hour	Free
KR12	37	Female	1-3 days a month	30 minutes to 1 hour	Free
KR13	25	Male	3-5 days a week	30 minutes to 1 hour	Free
KR14	23	Female	Multiple times a day	Less than 30 minutes	Free
KR15	30	Male	1-2 times a month	Less than 30 minutes	Free
US01	49	Male	1-3 days a month	30 minutes to 1 hour	Free
US02	30	Male	Multiple times a day	Less than 30 minutes	Free
US03	26	Female	Multiple times a day	1 to 2 hours	Free
US04	35	Female	3-5 days a week	More than 4 hours	Free
US05	22	Female	Multiple times a day	Less than 30 minutes	Paid
US06	26	Female	3-5 days a week	Less than 30 minutes	Free
US07	33	Male	1-3 days a month	Less than 30 minutes	Paid
US08	23	Male	Multiple times a day	Less than 30 minutes	Paid
US09	26	Female	Once a week	30 minutes to 1 hour	Free
US10	26	Female	Multiple times a day	1 to 2 hours	Paid
US11	26	Female	3-5 days a week	1 to 2 hours	Free
US12	27	Female	Multiple times a day	30 minutes to 1 hour	Free
US13	48	Female	Once a week	1 to 2 hours	Free
US14	30	Female	1-3 days a month	Less than 30 minutes	Free
US15	27	Female	3-5 days a week	30 minutes to 1 hour	Free

Table 1. The demographic information of Korean and USA participants. Frequency refers to the number of times participants use ChatGPT. Subscription plan refers to whether participants use ChatGPT free plan or ChatGPT Plus.

from both Korea and US use ChatGPT in test step of the design thinking process. They shared that they use ChatGPT when conducting usability testing.

4.3 How ChatGPT is used in each culture

Before receiving responses from ChatGPT, nine participants entered additional information, added specific answer formats, and used directive commands to generate prompts. Both five Korean and US participants (KR02, KR14, US01, US05, US15) entered additional information before receiving responses from ChatGPT. Three participants (KR01, KR12, US05) added specific answer formats before receiving responses from ChatGPT. Another three participants (KR01, KR11, US02) used directive commands before receiving responses from ChatGPT. During receiving responses from ChatGPT, seven participants from both Kore and US asked the same question multiple times, modified the initial prompt, and narrowed down the scope of questions to receive a desired response. Four participants (KR05, KR06, KR12 US12) narrowed down the scope of questions during receiving responeses from ChatGPT. Three participants (KR09, KR11, US05) modified the initial prompt during receiving responses from ChatGPT. Only one Korean participant (KR06) asked the same question multiple times during receiving responses from ChatGPT. After receiving responses from ChatGPT, only two Korean participants (KR05, KR14) either used other search engines or edited on their own to utilize the response to their work. They used other search engines after receiving responses from ChatGPT.

4.4 Challenges UX practitioners face when using ChatGPT

11 participants mentioned they struggle using ChatGPT as they do not think information provided by ChatGPT is not trustworthy. Eight participants (KR02, KR04, KR12, KR13, US04, US08, US12, US14) from both Korea and US said they could not trust the information provided by ChatGPT since they could not verify that infotmation. Only three Korean participants (KR10, KR13, KR14) struggled to use ChatGPT because they could not trust the references provided by ChatGPT. Only one US participant (US11) struggled to use ChatGPT. The participant said he/she could not trust the information provided by ChatGPT since its capabilities are not trustworthy. Four participants all from Korea said they could not generate prompts well. Three participants (KR03, KR06, KR09) stated that they did not know how to start generating prompts as there was no specific guidance about it. One participant (KR12) stated that he/she did not know how to modify prompts in order to get satisfying responses. Three participants (US01, US04, US09) only from US said they could not share confidential information with ChatGPT. They recalled that they could not share confidential information with ChatGPT as they did not know what it would do with their information. 14 participants had difficulties using ChatGPT as they could not get desired information. Eight participants (KR02, KR03, KR10, US05, US06, US12, US13, US15) from both Korea and US could not get desired information as they had to revise ChatGPT's responses before they could use. Six Korea and US participants (KR01, KR03, KR05, KR09, KR12, US12) said they could not get desired responses as the responses did not include exact information they needed. Two participants (KR11, KR14) from Korea said they could not get desired responses as they could not gain new ideas from the responses.

5 DISCUSSION

Our findings shed light on how and why UXPs experience and interact with ChatGPT. However, the findings of this study revealed specific needs and challenges that UXPs experienced. In this section, we describe particular challenges for UXP when using ChatGPT and present design implications for technology to address the challenges.

5.1 Cross-cultural similarities

5.1.1 Participants heavily utilized ChatGPT during the empathize and prototype steps. In our study, we found that both US and Korean participants heavily utilized ChatGPT during the empathize and prototype phases. Our research sought to explore whether participants faced greater challenges in the empathize or prototype stages. Beverland et al. During the design process, it is common to move forward to the next step before returning to previous stages [4]. However, empathize and prototype are the two phases that participants return to most frequently. Thus, it could be inferred that ChatGPT is not used frequently because the empathize and prototype stages pose the greatest challenges during the iteration of the design thinking process.

5.1.2 Participants had difficulty finding the desired information when using ChatGPT.. Both U.S. and Korean participants had difficulty finding the desired information when using ChatGPT, raising the question of why HCI practitioners from both countries face this challenge. Zheng et al. analyzed ChatGPT failures in answering complex open-domain questions and categorized them into four key areas: comprehension, factuality, specificity, and inference [47]. It is likely that both U.S. and Korean participants had difficulty finding information using ChatGPT because of the possibility of errors in providing accurate answers that meet users' expectations.

Another question that arises is why participants would continue to use ChatGPT to find information despite their expressed difficulties. This question leads us to explore the broader reasons for using ChatGPT beyond information retrieval. Huang and Tan found that ChatGPT serves as a powerful tool to assist scientists in writing review articles more efficiently, increasing proficiency, speeding up the writing process, and saving time [17]. Thus, it is plausible that both US and Korean participants continue to use ChatGPT because it supports their various writing tasks, even in the face of information retrieval challenges.

5.1.3 Participants had difficulty visualizing data while utilizing ChatGPT.. We discovered in our research that both US and KR participants had difficulty visualizing data while utilizing ChatGPT. We questioned the rationale behind UX professionals utilizing visual input or output for ChatGPT. Pop and schricker found that UX designers utilize ChatGPT in conjunction with AI tools such as Midjourney [31]. As a result, it is plausible that US and KR participants require the visual input of ChatGPT to work alongside other visual AI tools.

5.2 Cross-cultural differences

The observation that Korean participants had more difficulty generating prompts when using ChatGPT raises the question of what specific difficulties they encountered in this process. It is worth noting that ChatGPT has been shown to have a performance bias toward English, particularly in tasks requiring advanced reasoning skills [19]. In addition, it has been found that ChatGPT can perform better with English prompts, even when the task and input texts are intended for other languages [19]. Therefore, it is conceivable that Korean participants may have had more difficulty generating prompts because they were attempting to do so in English, which is not their native language, likely due to ChatGPT's documented preference for English prompts.

5.3 Design Opportunities

The challenges articulated by our participants suggest multiple design opportunities for using ChatGPT, from before getting responses from ChatGPT to after getting responses from ChatGPT. Understanding how users of different cultures perceive ChatGPT usage (e.g., ways in which they can share, challenges they face) has implications for how platforms ought to be designed for them, and the prioritization of design features to improve the experience of specific user groups. Given the significant role user experience plays in interactions with ChatGPT, and the desire for more and easier ChatGPT use in both cultures, these implications can help inform the design of features that facilitate greater connections.

5.3.1 Indicating the sources of the references. To improve UX practitioners' user experience, we suggest a feature that provides them with sources of references. This feature's primary purpose is to help users get references to improve the validity and reliability of the sources provided by ChatGPT. Getting specific, accurate references on the information that the user needs would enhance trust in the information that ChatGPT provides, which would lead to higher frequency in using ChatGPT. Liu et al. conducted an audit with four popular LLMs, including Bing Chat, NeevaAI, perplexity.ai, and YouChat. The authors found that "on average, a mere 51.5% of generated sentences are fully supported by citations and only 74.5% of citations support their associated sentence" [23]. Dao and Le, in their analysis involving Vietnamese students within a wide range of school subjects, found that BingChat demonstrated superior performance compared to ChatGPT in terms of student comprehension, reasoning, and generation of creative and informative text. Notably, BingChat employs features, such as hyperlinks and citations within responses, which helps amplify learning that is already taking place [8]. It would lead to an increase in the number of ChatGPT users.

5.3.2 Providing visual outputs to the users. Participants struggled to gain information regarding visual data, such as images and UI design. For example, KR participants found it difficult to analyze visual components, such as screen layouts and labeled components, when using ChatGPT. This difficulty led to KR participants highlighting the current limitation of ChatGPT in understanding visual and complex components compared to its proficiency in processing text-based data. In the previous studies, Wu et al. suggested a system called Visual ChatGPT, where it allows the user to interact with ChatGPT by sending and receiving images [44]. Therefore, we believe that a features that provides visual input and ouput may provide personalized support that fulfills the needs of individual UX practitioners who need visual aids for their UW work, such as prototyping a UI design and creating a high fidelity wireframe for a website design.

5.3.3 Assisting in generating prompts when using ChatGPT.. Participants struggled with creating prompts, thus getting poor-quality responses. KR participants, in particular, found it difficult to generate prompts when using CHatGPT. We suggest the need for prompt writing support. Although ChatGPT appears decent at understanding its users, some participants have reported poor experiences when it fails to comprehend their requests. This often led them to rephrase their query, which some found problematic and cumbersome [40]. While Microsoft, one of the main technology companies, argued that ChatGPT is useful in its ability to refine prompts for "more details, clarity, and ideas," it may be experienced as frustrating when users expect ChatGPT to offer the desired response on the first shot. To avoid frustrating user experiences due to a need for rephrasing, Skjuve et al. proposed prompt-writing support, which may be a relevant area of future research – in terms of intelligent support for prompt-writing and optimizing the process of prompt refinement. For example, by having mechanisms that detect the use of inefficient prompts and suggest alternatives based on assumptions regarding user intent.

6 LIMITATIONS AND FUTURE WORK

As an exploration, we conducted an interview-based study that would enable us to capture rich user data with various anecdotes and complex nuances of personal experiences in using ChatGPT. However, it has some limitations.

The first limitation of the study was the small number of participants who completed the interview. We interviewed a total of 30 valid participants. Thus, the findings of our study may not be representative of other UX practitioners who did not participate in our study. Conducting the study with larger populations would be necessary. Next, we only looked into the experiences of UX practitioners in two cultures. Therefore, we would also need to explore the experiences of UX practitioners in other cultures as well. Moreover, we only focused on UX practitioners' experience using ChatGPT. We would also need to explore the experiences of people in different fields other than UX.

Also, our data collection methods (e.g., survey and interview) have a few limitations. For instance, we did not ask about UX practitioners' experience using ChatGPT during other processes or models. Studying how practitioners utilize ChatGPT not only in the design thinking process but also in different processes or models would be crucial for understanding their experience with ChatGPT from a broader perspective. Moreover, we explored UX practitioners' experience in using ChatGPT only. Our future study should include experience with other LLM-based chatbots. Additionally, we did not actually observe how participants use ChatGPT in their design thinking process. For our future work, we should use other data collection methods that would allow us to explore their use in depth.

Lastly, our data analysis methods have some limitations. We did not compare participants who use the ChatGPT free plan and those who use the paid plan. In the future, we aim to compare UX practitioners' experience with ChatGPT based on their plans. We also did not compare the experience of practitioners in academia and industry. Thus, we should compare the experience of practitioners in academia and industry. In addition, we did not compare practitioners' experience by age. Comparing their experience by age would be necessary for future work. Also, we did not compare the experience of practitioners by their HCI backgrounds (e.g., computer science, design). Therefore, understanding the experience of practitioners by their HCI backgrounds would be necessary. Finally, we did not analyze UX practitioners' experience in depth. In the future study, we should consider the cultural context behind the practitioners and analyze their experience more in-depth.

7 CONCLUSIONS

We conducted surveys and semi-structured interviews with 30 participants from two cultures, South Korea and the United States. Taking a user-centered approach of understanding UX practitioners, we uncovered motivations for using chatgpt, how UXPs use Chatgpt, and challenges perceived by UXPs when using ChatGPT. Through this investigation, we found similarities and differences between the two cultures in sharing music. Participants from both cultures heavily utilized ChatGPT during the empathize and prototype steps. Korean participants had more difficulty generating prompts when using ChatGPT than American participants Combining our findings, we suggest several design opportunities for features that can help UX practitioners to use generative AI systems effectively. Prior studies have reported on UX practitioners' experience of LLM-based chatbots. To further extend these studies, our study highlights experience of UX practitioners from two cultures regarding using ChatGPT in design thinking process. By exploring different aspects of the usage of ChatGPT by UX practitioners in their design work, we provide insights into how people interact with ChatGPT in different cultural contexts. We encourage researchers in the CHI community to conduct further studies on the experience on LLM-based chatbots and challenges UX practitioners face.

REFERENCES

[1] [n. d.]. The Double Diamond - Design Council. https://www.designcouncil.org.uk/our-resources/the-double-diamond/

[2] Aino Ahtinen, Shruti Ramiah, Jan Blom, and Minna Isomursu. 2008. Design of Mobile Wellness Applications: Identifying Cross-Cultural Factors. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat* (Cairns, Australia) *(OZCHI '08)*. Association for Computing Machinery, New York, NY, USA, 164–171. https://doi.org/10.1145/1517744.1517798

[3] Nick Babich. 2020. The UX Design Process: Everything You Need to Know. *Adobe Xd Ideas* (2020).

[4] Michael B. Beverland, Sarah J. S. Wilner, and Pietro Micheli. 2015. Reconciling the tension between consistency and relevance: design thinking as a mechanism for brand ambidexterity. *Journal of the Academy of Marketing Science* 43, 5 (01 Sep 2015), 589–609. https://doi.org/10.1007/s11747-015-0443-8

[5] Chen Cao. 2023. Scaffolding CS1 Courses with a Large Language Model-Powered Intelligent Tutoring System. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 229–232. https://doi.org/10.1145/3581754.3584111

[6] Sena Çerçi, Marta E. Cecchinato, and John Vines. 2021. How Design Researchers Interpret Probes: Understanding the Critical Intentions of a Designerly Approach to Research. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 624, 15 pages. https://doi.org/10.1145/3411764.3445328

[7] Lingjiao Chen, Matei Zaharia, and James Zou. 2023. FrugalGPT: How to Use Large Language Models While Reducing Cost and Improving Performance. arXiv:2305.05176 [cs.LG]

[8] Xuan-Quy Dao and Ngoc-Bich Le. 2023. ChatGPT is Good but Bing Chat is Better for Vietnamese Students. arXiv:2307.08272 [cs.CL]

[9] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 278–288. https://doi.org/10.1145/3025453.3025739

[10] Stanford d.school. 2012. A Design Thinking Process. https://web.stanford.edu/class/me113/d_thinking.html

[11] C.R. Ember. 2009. *Cross-Cultural Research Methods*. AltaMira Press. https://books.google.com/books?id=bcClscT2CIAC

[12] World Leaders in Research-Based User Experience. [n. d.]. A 100-Year View of User Experience (by Jakob Nielsen). https://www.nngroup.com/articles/100-years-ux/

[13] K. J. Kevin Feng, Tony W Li, and Amy X. Zhang. 2023. Understanding Collaborative Practices and Tools of Professional UX Practitioners in Software Organizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 764, 20 pages. https://doi.org/10.1145/3544548.3581273

[14] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: Cultural Probes. *Interactions* 6, 1 (jan 1999), 21–29. https://doi.org/10.1145/291224.291235

[15] Polat Goktas, Gul Karakaya, Ali Fuat Kalyoncu, and Ebru Damadoglu. 2023. Artificial Intelligence Chatbots in Allergy and Immunology Practice: Where Have We Been and Where Are We Going? *The Journal of Allergy and Clinical Immunology: In Practice* 11, 9 (2023), 2697–2700. https://doi.org/10.1016/j.jaip.2023.05.042

[16] Won Ho and Daehyun Lee. 2023. Enhancing Engineering Education in the Roblox Metaverse: Utilizing chatGPT for Game Development for Electrical Machine Course. *International Journal on Advanced Science, Engineering & Information Technology* 13, 3 (2023).

[17] Jingshan Huang and Ming Tan. 2023. The role of ChatGPT in scientific communication: writing better scientific review articles. *Am J Cancer Res* 13, 4 (April 2023), 1148–1154.

[18] Eunkyung Jo, Daniel A. Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 18, 16 pages. https://doi.org/10.1145/3544548.3581503

[19] Viet Dac Lai, Nghia Trung Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui, and Thien Huu Nguyen. 2023. ChatGPT Beyond English: Towards a Comprehensive Evaluation of Large Language Models in Multilingual Learning. arXiv:2304.05613 [cs.CL]

[20] Eun-young Lee, Ngagaba Gogo Dae il, Gi-hong An, Sungchul Lee, and Kiho Lim. 2023. ChatGPT-Based Debate Game Application Utilizing Prompt Engineering. In *Proceedings of the 2023 International Conference on Research in Adaptive and Convergent Systems* (Gdansk, Poland) *(RACS '23)*. Association for Computing Machinery, New York, NY, USA, Article 29, 6 pages. https://doi.org/10.1145/3599957.3606244

[21] Juho Leinonen, Paul Denny, Stephen MacNeil, Sami Sarsa, Seth Bernstein, Joanne Kim, Andrew Tran, and Arto Hellas. 2023. Comparing Code Explanations Created by Students and Large Language Models. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1* (Turku, Finland) *(ITiCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 124–130. https://doi.org/10.1145/3587102.3588785

[22] Xi Lin. 2023. Exploring the Role of ChatGPT as a Facilitator for Motivating Self-Directed Learning Among Adult Learners. *Adult Learning* (2023), 10451595231184928.

[23] Nelson F. Liu, Tianyi Zhang, and Percy Liang. 2023. Evaluating Verifiability in Generative Search Engines. arXiv:2304.09848 [cs.CL]

[24] Wenwei Luo, Huihua He, Jin Liu, Ilene R. Berson, Michael J. Berson, Yisu Zhou, and Hui Li. 2023. Aladdin's Genie or Pandora's Box for Early Childhood Education? Experts Chat on the Roles, Challenges, and Developments of ChatGPT. *Early Education and Development* 0, 0 (2023), 1–18. https://doi.org/10.1080/10409289.2023.2214181 arXiv:https://doi.org/10.1080/10409289.2023.2214181

[25] Steven Moore, Q. Vera Liao, and Hariharan Subramonyam. 2023. FAIlureNotes: Supporting Designers in Understanding the Limits of AI Models for Computer Vision Tasks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 10, 19 pages. https://doi.org/10.1145/3544548.3581242

[26] Steven Moore, Q. Vera Liao, and Hariharan Subramonyam. 2023. FAIlureNotes: Supporting Designers in Understanding the Limits of AI Models for Computer Vision Tasks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 10, 19 pages. https://doi.org/10.1145/3544548.3581242

[27] Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2022. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 74, 18 pages. https://doi.org/10.1145/3526113.3545616

[28] So Yeon Park, Kyung Yun Lee, and Jin Ha Lee. 2022. Cross-Cultural Exploration of Music Sharing. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 383 (nov 2022), 28 pages. https://doi.org/10.1145/3555108

[29] David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon Emissions and Large Neural Network Training. arXiv:2104.10350 [cs.LG]

[30] Helen Petrie, Tanja Walsh, Olufunmilayo Odutola, and Lei Ang. 2015. Cross-Cultural Issues in Working with Users in the Design of Interactive Systems. *ICoRD'15–Research into Design Across Boundaries Volume* (2015), 515.

[31] Mira Pop and Max Schricker. 2023. AI as a tool and its influence on the User Experience design process: A study on the usability of human-made vs more-than-human-made prototypes. , 80 pages.

[32] Mirjana Prpa, Sarah Fdili-Alaoui, Thecla Schiphorst, and Philippe Pasquier. 2020. Articulating Experience: Reflections from Experts Applying Micro-Phenomenology to Design Research in HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376664

[33] Christian Remy, Gunnar Harboe, Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2021. Challenges and Opportunities in the Design of Digital Distributed Affinity Diagramming Tools. In *Proceedings of the 32nd European Conference on Cognitive Ergonomics* (Siena, Italy) *(ECCE '21)*. Association for Computing Machinery, New York, NY, USA, Article 11, 5 pages. https://doi.org/10.1145/3452853.3452871

[34] Vidya Sarangapani, Ahmed Kharrufa, David Leat, and Pete Wright. 2019. Fostering Deep Learning in Cross-Cultural Education through Use of Content-Creation Tools. In *Proceedings of the 10th Indian Conference on Human-Computer Interaction* (Hyderabad, India) *(IndiaHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 7, 11 pages. https://doi.org/10.1145/3364183.3364184

[35] Anna Karen Schmitz, Thomas Mandl, and Christa Womser-Hacker. 2008. CULTURAL DIFFERENCES BETWEEN TAIWANESE AND GERMAN WEB USER. (2008).

[36] Yeganeh Shahsavar and Avishek Choudhury. 2023. User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study. *JMIR Hum Factors* 10 (17 May 2023), e47564. https://doi.org/10.2196/47564
[37] Yeganeh Shahsavar and Avishek Choudhury. 2023. User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study. *JMIR Hum Factors* 10 (17 May 2023), e47564. https://doi.org/10.2196/47564
[38] Abdulhadi Shoufan. 2023. Exploring Students' Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey. *IEEE Access* 11 (2023), 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224
[39] Abdulhadi Shoufan. 2023. Exploring Students' Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey. *IEEE Access* 11 (2023), 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224
[40] Marita Skjuve, Asbjørn Følstad, and Petter Bae Brandtzaeg. 2023. The User Experience of ChatGPT: Findings from a Questionnaire Study of Early Users. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 10 pages. https://doi.org/10.1145/3571884.3597144

[41] Peihao Wang, Rameswar Panda, Lucas Torroba Hennigen, Philip Greengard, Leonid Karlinsky, Rogerio Feris, David Daniel Cox, Zhangyang Wang, and Yoon Kim. 2023. Learning to Grow Pretrained Models for Efficient Transformer Training. arXiv:2303.00980 [cs.LG]

[42] Qiaosi Wang, Michael Madaio, Shaun Kane, Shivani Kapania, Michael Terry, and Lauren Wilcox. 2023. Designing Responsible AI: Adaptations of UX Practice to Meet Responsible AI Challenges. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 249, 16 pages. https://doi.org/10.1145/3544548.3581278

[43] Spencer Williams, Ridley Jones, Katharina Reinecke, and Gary Hsieh. 2022. An HCI Research Agenda for Online Science Communication. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 490 (nov 2022), 22 pages. https://doi.org/10.1145/3555591

[44] Chenfei Wu, Shengming Yin, Weizhen Qi, Xiaodong Wang, Zecheng Tang, and Nan Duan. 2023. Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models. arXiv:2303.04671 [cs.CV]

[45] Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People + AI Guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 356, 13 pages. https://doi.org/10.1145/3544548.3580900

[46] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388

[47] Shen Zheng, Jie Huang, and Kevin Chen-Chuan Chang. 2023. Why Does ChatGPT Fall Short in Providing Truthful Answers? arXiv:2304.10513 [cs.CL]

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009