An Exploratory Study on How AI Awareness Impacts Human-AI Design Collaboration

Zhuoyi Cheng
College of Software Technology
Zhejiang University
Hangzhou, China
zy_cheng@zju.edu.cn

Pei Chen*
College of Computer Science and
Technology
Zhejiang University
Hangzhou, China
chenpei@zju.edu.cn

Wenzheng Song
College of Computer Science and
Technology
Zhejiang University
Hangzhou, China
svvz@zju.edu.cn

Hongbo Zhang
College of Computer Science and
Technology
Zhejiang University
Hangzhou, China
hongbozhang@zju.edu.cn

Zhuoshu Li
College of Computer Science and
Technology
Zhejiang University
Hangzhou, China
lizhuoshu@zju.edu.cn

Lingyun Sun
College of Computer Science and
Technology
Zhejiang University
Zhejiang – Singapore Innovation and
AI Joint Research
Hangzhou, China
sunly@zju.edu.cn

Abstract

The collaborative design process is intrinsically complicated and dynamic, and researchers have long been exploring how to enhance efficiency in this process. As Artificial Intelligence technology evolves, it has been widely used as a design tool and exhibited the potential as a design collaborator. Nevertheless, problems concerning how designers should communicate with AI in collaborative design remain unsolved. To address this research gap, we referred to how designers communicate fluently in human-human design collaboration, and found awareness to be an important ability for facilitating communication by understanding their collaborators and current situation. However, previous research mainly studied and supported human awareness, the possible impact AI awareness would bring to the human-AI collaborative design process, and the way to realize AI awareness remain unknown. In this study, we explored how AI awareness will impact human-AI collaboration through a Wizard-of-Oz experiment. Both quantitative and qualitative results supported that enabling AI to have awareness can enhance the communication fluidity between human and AI, thus enhancing collaboration efficiency. We further discussed the results and concluded design implications for future human-AI collaborative design systems.

*Corresponding author

CCS Concepts

• **Human-centered computing** → **Collaborative interaction**; **Empirical studies in HCI**.

Keywords

Human-AI collaboration, Design collaboration, Generative AI, AI Awareness, Human-AI communication

1 Introduction

The design process is intrinsically improvisational, open-ended, and dynamic [58, 87], in which designers occasionally switch between different design phases and iterate their ideas to achieve an optimal design [33]. To enhance efficiency in such a complicated process, researchers have long been studying how to integrate Artificial Intelligence (AI) into the design process [15, 19, 78]. Thanks to the rapid evolution in large-scale models, current AI technologies have gained higher creativity, stronger reasoning [24, 79], and generation abilities [20, 65], making it increasingly suitable for enhancing the efficiency of the design process. Prior endeavors have applied AI in different design stages, like ideation [37, 81], prototyping [17, 49], and evaluation [8, 82]. In these cases, AI was treated as a tool to automate certain design tasks, and most interactions were initiated by designers. However, such AI tools may introduce new interaction paradigms and bring extra work in steering the system to obtain ideal feedback, which may increase designers' workload and learning cost [25].

More recently, researchers have sought to enable AI to directly collaborate with designers. For example, Muller et al. [51] designed an AI chatbot capable of autonomously suggesting ideas in the

group brainstorming process, and Kuang et al. [42] engaged UX designers to inspect UX problems with AI collaboratively. In the general area of human-AI collaboration, researchers have exploited different theories and frameworks to model human-AI collaboration, facilitating the understanding of human-AI collaboration at an abstract level. For instance, Tankelevitch et al. [71] applied the descriptive framework for metacognition to understand users' metacognitive demands when using Generative AI systems, and provided cases to lower the demands. Muller and Weisz [52] described human-AI collaboration pattern by illustrating how human and AI initiative shifts in different scenarios. However, communication between human and AI remains to be explored extensively, which is crucial for decision-making, design information sharing, and shared understanding building in the highly dynamic design process [11, 41, 68],

To address this research gap, we first looked into existing studies about human-human collaboration to understand how designers communicate with collaborators fluently [1, 53, 80]. In various forms of human-human design collaboration, including face-to-face [80], virtual [40], and online [76] collaborations, designers comprehend the status quo and leverage information in the environment and communicative cues from collaborators to initiate verbal communication. Having awareness of collaborators and the current situation is the key to such processes [22, 28, 76]. There are various definitions for the term "awareness" (e.g., [22, 28, 30]), but most of them share the core meaning concluded by Dourish and Bellotti [18, p. 107], which we used in this study: "*...awareness is an understanding of the activities of others, which provides a context for your own activity.*" The awareness activity happens naturally in co-located collaboration, while awareness support is needed if people are working distributedly in a collaborative system [28]. For instance, Cidota et al. [12] evaluated different notification systems for users virtually collaborating in an augmented reality (AR) environment to develop users' awareness of each other. Dehler et al. [16] developed awareness support in a collaborative learning environment to assist learners in having awareness of others' knowledge. We propose that quipping AI with this ability can assist in efficient human-AI communication and collaboration. However, while humans can naturally be aware of others and the environment [28], enabling AI to have awareness of humans is underexplored in both the impacts of doing so and the way to do so.

To this point, we explored the impacts of enabling AI awareness in human-AI design collaboration, and our main research question (RQ) focused on "**How will AI having awareness of the designer and current situation impact communication in human-AI design collaboration?**" As the RQ pertains to a future system that does not exist yet, we conducted user studies using the Wizard-of-Oz (WoZ) method. We first implemented a system that ensured the basic design and human-AI collaboration functions. Next, we iteratively designed a WoZ workflow with two Wizards to complement the communication and awareness functions. We then conducted a within-subject experiment involving 20 participants using this system and extracted quantitative and qualitative data from the experiment process recordings, conversation histories, and interview transcripts. The quantitative results showed that human-AI communication was significantly more fluent when AI had awareness, implying more efficient collaboration, and thematic analysis

results of the interview presented detailed impacts on communication in human-AI collaboration. We further discussed the above results to build relationships between quantitative and qualitative data to reveal the causalities and inferences. Last, we concluded design implications about maintaining and leveraging awareness in human-AI collaboration and communication strategies for AI in future human-AI collaborative design systems.

Our research provides two main contributions to future human-AI collaborative design. First, we proposed a prospecting way to enhance communication fluidity in human-AI design collaboration by enabling AI to have awareness of designers and the current situation, and conducted user studies to certify this prospect. Second, we provided novel insights into future development of human-AI collaborative design systems based on the experiment results. We believe this study can shed light on how human and AI should communicate and collaborate not only in the design process, but also in other processes that share the complex and dynamic features with the design process.

2 Related Work

To better situate this study and provide references for system design, we conducted a thorough literature review of research on the following three topics: 1) how do researchers integrate AI in the design process, including AI-driven design tools and AI collaborators, 2) common communication channels in collaboration and how they were transferred to human-AI collaboration, and 3) mechanisms for supporting awareness in co-located and distributed human-human collaboration, and human-AI collaboration.

2.1 Integrating AI in Design

Design requires knowledge, information, and skills from diverse areas, and has long been a task that demands efforts beyond a single designer due to its burgeoning complexity [61, 70, 87]. The need for collaboration and assistance thus naturally appeared, and AI is one important technology that can help. In early research, AI was considered as a design support tool to assist in quantifiable or functional design tasks [61]. For example, Chakrabarti and Bligh [9] proposed a two-step approach to work out viable mechanical design solutions based on a set of known rules, and Grecu and Brown [26] explored AI support for parametric design, where the product has already been defined except for combining several detailed values (e.g., color) with various options. In these tasks, AI was expected to reduce tedious work and give predictable outputs. As research on computational creativity booms, AI gained creativity and moved from the above "routine design tasks" to "creative design tasks" in the design process. Designers can now use various AI-driven tools to obtain ideas [37, 47], organize thoughts [86], explore design space [8, 78], etc. However, these applications are mainly design tools that require designers to initiate operations.

Recently, the development of AI abilities in generation and reasoning abilities [20, 79] allowed AI to take an active role in design [43], thus being prospective in becoming a design collaborator. Researchers are striving for better human-AI design collaboration from both theoretical and practical perspectives. Rezwana and Maher [58] did a thorough literature review on human-AI co-creation

and proposed a COFI framework to describe the interactions between collaborators and with the shared product in co-creative systems. Shi et al. [66] also reviewed articles but specifically focused on designer-AI collaboration, and concluded five characteristics, including scope, access, agency, flexibility, and visibility, to help describe designer-AI collaboration and identify existing problems. Results from both works elaborated on problems in human-AI communication, with the former one presenting the lack of "human to AI consequential communication" and "AI to human communication", and the latter one pointing out the importance and a dearth of multiple communication modalities other than visual ones. Other more practice-oriented studies focus on specific design tasks, exploring interaction pathways and guidelines for human-AI collaboration. For example, Zhou et al. [87] explored the nonlinear collaboration between graphic designers and AI in the OptiMuse system using the WoZ method, and proposed a co-design framework. Pan et al. [56] explored how AI can intelligently collaborate with people through a WoZ-like method, and implemented a collaborative diagram editing system guided by obtained insights.

Similar to Zhou et al. [87] and Pan et al. [56], our work seeks to explore the human-AI collaborative design process. Grounded in the problems identified by Rezwana and Maher [58] and Shi et al. [66], we specifically focus on problems in human-AI communication and explore the potential of AI awareness to alleviate this problem.

2.2 Communicating in Human-Human and Human-AI Collaboration

In traditional human-human co-located design collaboration, people share expertise, ideas, resources, or responsibilities in the design process, to which communication is crucial [11]. Verbal communication is the most common way that people exchange information. People leverage various linguistic strategies and non-verbal communicative cues to enhance the efficiency of verbal communication, for example, deictic words can specify the content of the current discussion in a highly contracted way [72], gestures can focus attention and represent ideas or features, and facial expressions can assist in judging one's true feelings [23]. When collaboration is moved online and people work distributedly through collaborative systems, verbal communication, and its corresponding communication strategies require special support to function [28]. However, as remote collaboration becomes common and frequent, collaborative systems using text-based communication have matured rapidly (e.g., Comment function in Figma[1] and Miro[2]).

Similar to human-human distributed collaboration, communication and interaction in current human-AI collaboration is mainly based on text-based conversations. For example, Muller et al. [51] implemented a text-based chatbot to facilitate the brainstorming process, and Kuang et al. [42] designed a collaborative system for UX designers to identify usability problems with text-based conversation AI assistants. However, communicating with AI through text is not necessarily an optimal choice for human-AI communication, especially in the design process. Previous research has found communicating through voice leads to higher trust in both humans and

chatbots compared to text [2, 7]. Additionally, the need for composing textual prompts and conducting other operations through visual channel can cause increased cognitive load [71, 75], thus making it harder for mutual understanding in the design process [25]. In the more general field of human-AI collaboration, researchers have found verbal communication to be a favored communication strategy because of its direct and efficient feature in difficult tasks [83]. In a recent research about sketching, researchers pointed out that people "*usually use language to talk about real, distant, invisible, imaginary, or conceptual objects*" [62, p. 4-5], which are similar to what happens in the design process [39, 74]. However, systems supporting designers to communicate with AI through speech are underexplored [58, 66]. Therefore, we allow designers to communicate with AI through speech in this study to gain insights for future human-AI collaborative design systems that support verbal communication.

2.3 Supporting Awareness in Collaboration

The term "awareness" had been combined with different words to coin a new term and specify a specific research scope, like shared awareness [9] and situation awareness [21]. Similar as their overall meanings were, "awareness" has not been defined and used consistently [27]. In this paper, we adopt the definition provided by Dourish and Bellotti [18, p. 107]:

> "[...]awareness is an understanding of the activities of others, which provides a context for your own activity. This context is used to ensure that individual contributions are relevant to the group's activity as a whole, and to evaluate individual actions with respect to group goals and progress. The information, then, allows groups to manage the process of collaborative working."

Awareness is critical to successful design collaboration for providing information including shared information, group and individual activities, and coordination, to contextualize one's activity [18]. In co-located human-human collaboration, awareness activities are supported by human perceptions, which can happen naturally [28]. However, when collaboration is distributed, awareness activities require supporting mechanisms in the collaborative system to proceed [28], and researchers have developed various mechanisms to enable people to have awareness of other collaborators and the environment in different scenarios. Gutwin et al. [29] designed a set of interface widgets to support distributed workers having awareness of collaborators' current activities and view, including radar view, WYSIWIS (What You See Is What I See) view, WYSIWID (What You See Is What I Do) view, etc. Cidota et al. [12] compared different notification mechanisms for people collaborating through AR. Dehler et al. [16] designed a collaborative system and presented participants' knowledge levels to allow a better understanding of the group and each other. Janssen et al. [34] visualized group members' contributions in online collaboration to examine how collaborative learning is impacted by group awareness.

Although studies and applications of awareness mechanisms in human-human collaboration are abundant, the idea of equipping AI with awareness ability is underexplored due to the infancy of human-AI collaboration research and the difficulty of transferring

---

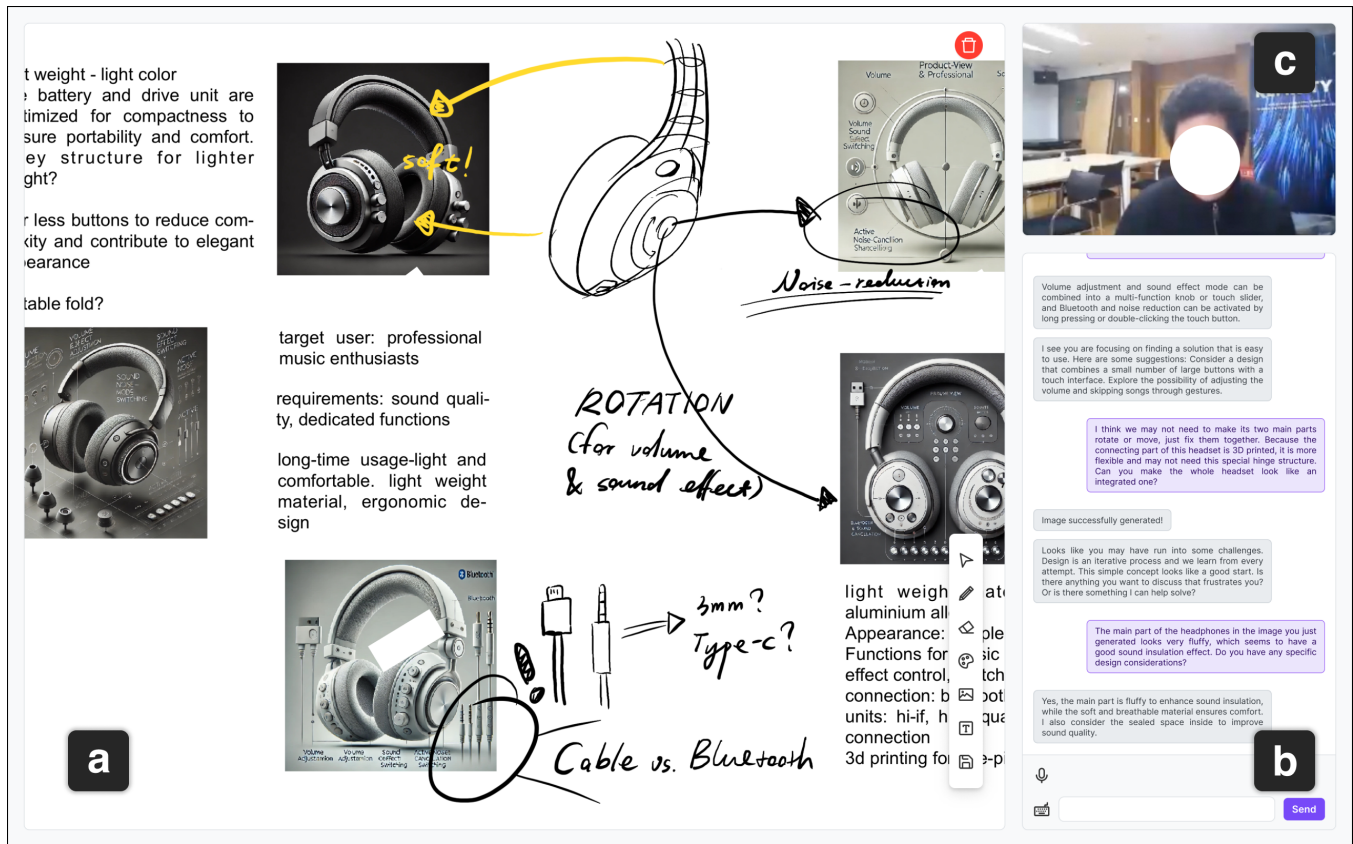[1]https://www.figma.com/
[2]https://miro.com

Figure 1: The interface example of the human-AI design collaborative system: a) Canvas. Canvas is the main interaction area where designers edit information using tools in the toolbar on the bottom right corner, and AI can also upload generated images to the canvas. b) Chat box. Communication history and speech recognition results are displayed in this area, and designers can also type here to initiate conversation in case of need. c) Camera screen. This is presented *in the Aware condition only* to indicate what AI can see. *Note: The interface is a translated version of the original one used for user studies.*

this human ability to AI. However, previous research has demonstrated that allowing users to have awareness of AI presence positively contributes to user engagement and collaborative experience [59], implying the potential advantages of applying the awareness theory in human-AI collaboration. Therefore, in this study, we first explored the impact of allowing AI to have awareness (i.e., AI is aware of the designer's design activities and current working content) through the WoZ method to ground future design and research. By exploring the potential of transferring the "awareness" concept in human-AI collaboration, we expect to guide the design of AI collaborators that have a certain level of initiative in a shared workspace [31, 64], adjust their output according to real-time contextual information [48], and infer deeper reasons and meaning behind the literal senses of obtained visual and textual information [45, 46] to support more dynamic and difficult tasks like design through human-AI collaboration.

3 Human-AI Collaborative Design System

In this section, we introduced the system used in our study, including user interface, system functions, and implementations[3].

3.1 User Interface and Interactions

As an initial prototype, we kept the interface simple and consistent with common human-AI collaboration systems (e.g., [42, 87]), and only functions served the RQ will be added apart from basic design and collaboration functions.

The interface is implemented with the React frontend framework[4] and Ant Design UI component[5]. The overall interface layout is consistent with current collaborative systems containing a shared workspace/canvas and a chatbot [42, 57, 60, 63]. Three main areas consist of the interface (Figure 1): a) Canvas. Basic design functions, including adding and removing text, images, and sketches are supported. Participants can switch between different tools in

---

[3]For source code and detailed guidelines, please refer to the supplementary material or https://github.com/SongWZ3214/AI_teammate/tree/master.
[4]https://react.dev
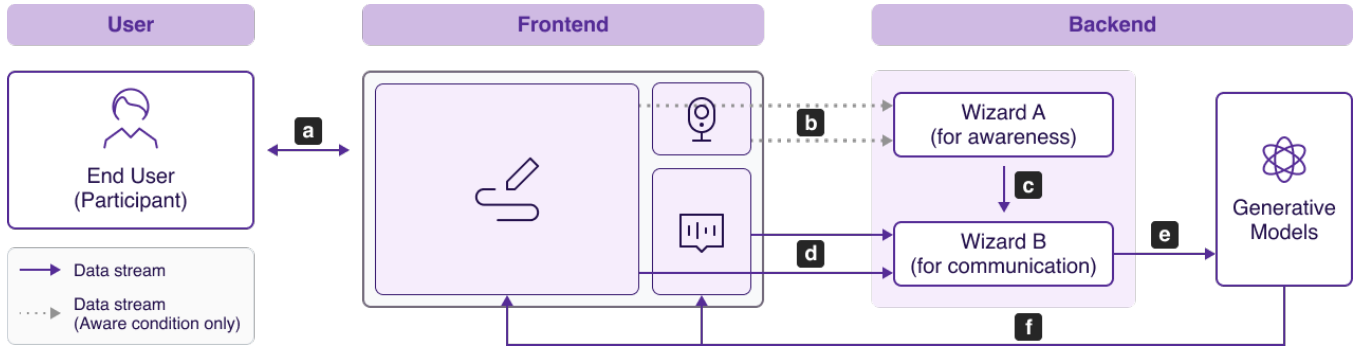[5]https://ant.design/components/overview-cn/

Figure 2: The overview of system structure and data stream. In our study, the participants a) interacted and communicated with the frontend (i.e., the user interface in Figure 1) and received feedback from AI. The backend consisted of two Wizards and generative models. In the Aware condition, Wizard A would b) obtain awareness information from the camera screen and canvas, c) compose prompts based on predefined scripts and send them to Wizard B. In the Non-aware condition, the prompt sent in process (c) would not contain awareness information. Wizard B in both conditions supported the human-AI communication process by d) processing speech recognition results and input with canvas information, and e) sending prompts coming from the participants and Wizard A to generative models. f) All generation results would be uploaded to the canvas.

the toolbar to edit the canvas. b) Chat box. Participants can communicate with AI through speech or textual input, and we encouraged participants to communicate through speech in this study unless the information can not be conveyed with speech. When using speech input, participants can naturally initiate or terminate a conversation without saying activation or termination commands. All communication histories between the participant and AI will be preserved and displayed in this area. Besides, participants can communicate with AI using deictic words like "this" or "its" to indicate what they select, which is a common linguistic strategy to simplify communication [5, 62, 72]. c) Camera screen. We use a Hikvision 2K Computer Camera to monitor the participants. They will not interact with this area, but can see what the camera is currently capturing, which equals what the AI is seeing. This area *only* appears in the condition where AI has awareness (henceforth the Aware condition), and in the condition where AI does not have awareness (henceforth the Non-aware condition), this area is hidden.

3.2 System and Wizard Implementation

The system and Wizard implementation in the two conditions (Aware and Non-aware) are illustrated in Figure 2. All contents and activities that designers see and experience exist in the frontend (Figure 2a), which is supported by two Wizards and generative models at the backend (Figure 2b-f). The system backend is built with Flask[6], and its functions can be broken down into three parts: awareness mechanisms fulfilled by Wizard A, communication functions supported by Wizard B, and feedback generation completed by generative models.

Awareness mechanisms. The awareness mechanisms will only function in the Aware condition (See Figure 2b). In this condition, Wizard A will observe and judge participants' current design activities per 60-90 seconds from participants' shared screen (displaying the frontend/system interface) through an online meeting software

according to adapted coding scheme from [36] (Table 2) to minimize bias. This information is then filled in a prompt template, and together with the screenshot of the current working process on the canvas, will be sent to Wizard B through real-time messaging software for eliciting AI feedback. In the Non-aware condition, Wizard A will only send common prompts without awareness information or screenshots. The prompt templates used in the two conditions were designed and iterated in pilot studies, and are presented as follows:

> **The Aware Condition:**
> ```
> We are currently in Step [ ], designer's
> design activity is [ ], and the content
> the designer is currently working on is
> presented in the image.
> Your feedback is:
> ```
>
> **The Non-aware Condition:**
> ```
> We are currently in Step [ ], please think
> divergently/convergently according to Requirement
> [ ].
> Your feedback is:
> ```

Communication between designer and AI. Wizard B is mainly responsible for organizing participants' input into prompts, receiving prompts from Wizard A, and sending these prompts to generative models. Participants in our study are instructed to mainly use speech communication. We used the Web Speech API[7] for speech recognition (i.e., STT, Speech-to-Text) and Text-to-Speech (TTS) functions. All that the participants speak will be recognized and transcribed to text, and sent to the backend. Wizard B will first judge whether the participants intend to communicate with AI, if yes, Wizard A will then assemble the transcribed speech snippets at the backend console if a complete paragraph is fragmented into several parts to form a complete prompt. Undesired inputs (e.g., think aloud, murmurs, conversations from and with the experimenters)

---

[6]https://flask.palletsprojects.com/en/3.0.x/

[7]https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API

are ignored in this process. Prompts produced by Wizard B and come from Wizard A are all sent to generative models for feedback generation. Textual inputs sent by participants will be directly sent to the generative models by Wizard A. The above operations were processed at the backend using a series of predefined short-cut commands by Wizard B.

Feedback generation. Feedback generation is supported by dedicated generative models for generating verbal/textual content and images in respective, which is GPT-4 turbo[8] for verbal/textual content, and DALL·E 3[9] for images. GPT-4 turbo will be initiated by the same prompt across participants (Appendix B) to ensure the models behave consistently across participants. The generated contents are then uploaded to the frontend. Images will be directly put on the canvas, verbal content will be converted into speech, and its corresponding text will be displayed in the chat box.
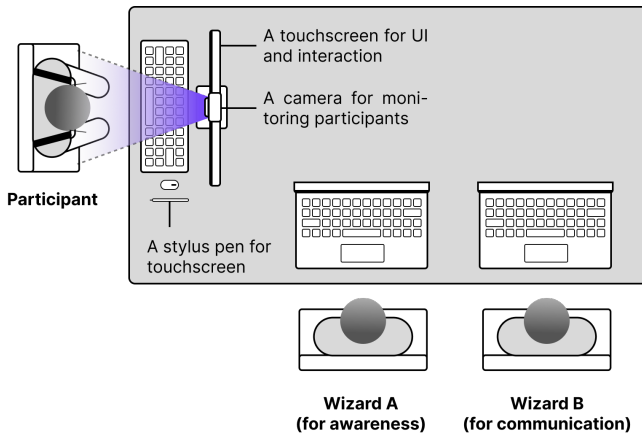

Figure 3: The WoZ experiment setting, including the positions of the participants and experimenters (Wizards), and important devices.

4 User Study

4.1 Participants

We recruited 20 participants (denoted as P1-P20) through an online forum, including 11 females and 9 males, ranging in age from 22 to 32 ($M = 23.6$, $SD = 2.4$). They have at least three years of design experience, and knowledge of AI-driven design tools and generative AI. All participants signed a consent form before the experiment and received CNY 100 (USD 14) after completion.

4.2 Condition Settings and Tasks

The experiment included two conditions: AI has awareness of the designer and current situation (the Aware condition), and does not have such awareness (the Non-aware condition). Whether or not AI has awareness was controlled by the workflow of Wizard A (See §3.2 and Figure 2)

We conducted within-subject user studies to alleviate differences existing in the following aspects that can discredit the results: 1)

individual hobbies and habits that can affect the metrics we used (§4.3), 2) Wizards' changing dexterity on WoZ workflow affected by accumulated experiences. Each participant went through both conditions. In each condition, participants were provided with materials describing a product, several target users, and possible user requirements of the product. They should produce 3-5 design concepts of the product in four predefined steps: 1) select only one type of target user and determine 1-3 main requirement(s) of the selected user, 2) break down the main requirement(s) into multiple detailed key requirements, and select the most important three key requirements, 3) explore solutions to each of the key requirement, and 4) combine the solutions to obtain 3-5 design concepts. The steps were excerpted and adapted from the product design and development steps by Ulrich and Eppinger [73]. To avoid the practice effect, participants would need to design different products in the two conditions, which were e-scooter and headphone, respectively. We determined the two products considering they are common, neither too simple nor too complex in technology and engineering, have a wide range of target users and usage scenarios, and are possible to produce various design concepts. The experiment conditions and design tasks were cross-assigned to regulate the order effect.

4.3 Measurement

There are no universal metrics for evaluating the communication fluidity between human and AI, but researchers have concluded communication-related analysis methods and metrics in human-human collaboration [4, 13, 14, 54, 55, 69], from which human-AI design collaboration studies often borrow. Measures used to evaluate collaboration were categorized into performance, process, and subjective measures [4]. Performance measures are metrics pertaining to task outcome, but were found insensitive across conditions in short-term tasks, partly because "*people tend to protect their primary task of getting the work done*"[14, p. 35, 50]. Using process measures is a recommended solution [50]. Also, subjective measures were utilized as complementary evidence of the process measure, which corresponds to the experiment process recordings, conversation histories, and interviews in this study.

Commonly used metrics include the turn frequency, turn duration, incidence/duration of overlapping speech, number of interruptions, turn length, turn completions, backchannels, deixis, questioning behaviors etc. [4, 55]. In this study, we mainly considered the following metrics: turn frequency, turn duration, turn length, and deixis. Other metrics were left unselected either due to the difficulty of identification or the rarity of occurrence in our experiment. The definitions of the metrics in this study are listed as follows:

- **Turn frequency:** The total number of conversation turns initiated by the participant. (unit: turns)
- **Turn duration:** The time participant spent to complete a turn. (unit: second)
- **Turn length:** The number of words in each turn for the participant. As we conducted experiments in Chinese, we calculated the number of Chinese characters. (unit: words)
- **Deixis:** The total number of deictic words used by the participant in the whole communication process (unit: times).
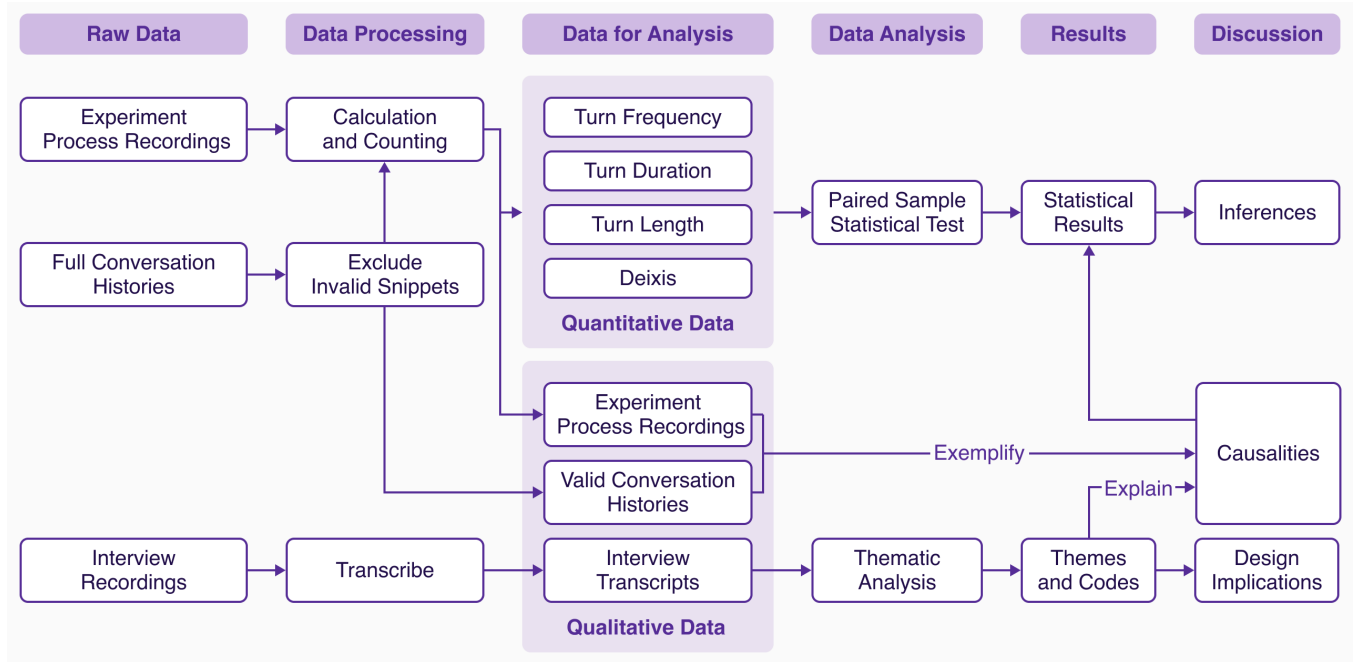
Figure 4: The overview of the data analysis process. Our raw data mainly came from three sources: The experiment process, the human-AI conversation during the experiments, and interview. These data were processed into valid data for further analysis, including four types of quantitative data corresponding to our predefined metrics, and three types of qualitative data. We analyzed these data and reported how AI awareness impacted communication between human and AI in the Result section (§5). Furthermore, we discussed the inferences and causalities of quantitative results and concluded design implications for future human-AI collaborative design systems in the Discussion section (§6).

4.4 Experiment Environment and Procedures

Two Wizards experienced a training session before the formal user studies by practicing for 4 complete trials of each condition to familiarize themselves with the workflow.

The experiment setting is illustrated in Figure 3. The system interface was displayed on a Dell P2418HT touchscreen, and the participants were provided with a stylus pen, a keyboard, and a mouse to operate the system, and were encouraged to use speech as the main communication channel. The camera was fixed on the top of the touch screen, and the captured screen was displayed in the interface. The two Wizards sat beside the participant, and Wizard A could see the participant's live screen via an online meeting software.

Each experiment included three sessions and lasted about 2 hours. All participants signed the informed consent before all sessions, and the WoZ method was not revealed to the participants until all sessions ended. The three sessions are described as follows:

(1) **Preparation:** The preparation session lasted about 20 minutes. This session commenced with an introduction to the experiment's background, goal, procedure, and reward. Then, one of the experimenters explained how to use the system, followed by a 5-minute practice using the system.

(2) **Formal experiments:** All participants went through both conditions. At the beginning of each condition, the experimenter will emphasize the features of this condition. Then,

the participants would start the task following the four steps. Each condition took up about 40 minutes, and the participant was allowed to have a 5-minute break between the two conditions.

(3) **Post-experiment interview:** Participants underwent a 15-minute semi-structured interview to help us gain a deeper understanding of their behaviors, perceptions of AI, communication and collaboration experiences in the two conditions, and suggestions for further optimization.

4.5 Analysis Method

The overview of the analysis process is illustrated in Figure 4. We obtained raw data from three sources: experiment process video recordings, full textual conversation histories stored in the system backend, and interview audio recordings. Before formal analysis, we first processed them to extract valid quantitative and qualitative data. We excluded invalid conversation snippets from the conversation histories, including speech from the experimenter, conversations that happened between the participants and experimenters when errors occurred, and meaningless snippets like "hmm...". Then we counted the number of turn frequencies and deixis, and calculated the turn length using the processed conversation histories. Next, we calculated the duration of each turn together with video recordings. For interview recordings, we transcribed them into texts. After the data processing processes, we obtained four types

Table 1: The overall statistical test results of the four metrics.

Paired sample t-test

	Aware		Non-aware				
	M (Mean)	SD	M (Mean)	SD	t	p	Cohen's d
Turn Frequency	26.93	12.17	21.00	10.48	2.731	**.012***	0.777
Deixis	5.29	3.97	5.78	4.58	-0.275	.756	0.085

Paired Wilcoxon signed-rank test

	Aware		Non-aware				
	M (Mean)	SD	M (Mean)	SD	Z	p	r
Turn Duration	10.14	5.84	13.87	5.50	-3.233	**.001****	0.864
Turn Length	38.85	20.82	46.20	20.81	-3.107	**.002****	0.830

*, **, *** refer to $p <.05$, $p <.01$, $p <.001$, respectively.

of quantitative data: turn frequency, turn duration, turn length, and the number of deixis, and three types of qualitative data: experiment recordings (not processed), valid conversation histories, and interview transcripts.

For quantitative data, we employed suitable paired sample statistical tests. First, we tested the normality of each paired sample using a Shapiro-Wilk normal distribution test. Pairs including at least one sample that is not normally distributed will be analyzed by the paired Wilcoxon signed-rank test, and normal-distributed pairs will be analyzed by paired t-test. For qualitative data, we mainly analyzed the interview transcripts using the thematic analysis method [6]. We reported the statistical results and thematic analysis results concerning both AI awareness and communication in the Result section (§5). In the Discussion section (§6), we discussed the inferences and causalities of the statistical results with the help of qualitative data, and concluded design implications extracted from the interview.

5 Result

In this section, we described how AI awareness impacted the communication between human and AI through quantitative and qualitative data.

5.1 Quantitative Results

Data from three participants (P11, P14, P19) were dismissed for disobeying the predefined steps during the experiment task. For example, P19 ignored predefined steps and kept iterating product appearances using AI, making it difficult for Wizard A to compose prompts based on predefined scripts. Besides, we decided not to include data from another three participants (P8, P9, P17) because they used textual inputs that took up an unignorable proportion (e.g., only 52% of P17's turns were input by speech in the Aware condition), although we encouraged them to communicate with AI using speech unless the contents could not be conveyed through such. As participants usually behave differently between speech and textual input, mixing the two data types in the analysis can discredit the results.

The Shapiro-Wilk normal distribution test results showed that data of turn duration from the Aware condition ($p = .008$) and turn length from both conditions ($p_{Aware} = .004$, $p_{Non-aware} = .007$)

were not normally distributed, and the remaining were normally distributed, so we used the paired sample t-test to analyze turn frequency and deixis, paired Wilcoxon signed-rank test to analyze turn duration and turn length. The statistical results are presented in Table 1. Results show that participants averagely initiated more turns in the Aware condition ($M = 26.93$) compared to the Non-aware condition ($M = 21.00$, $p = .012$, $d = 0.777$). In the Aware condition, participants spent significantly less time ($M = 10.14$) on each turn compared to the Non-aware condition ($M = 13.87$, $p = .001$, $r = 0.864$). Each turn contained significantly fewer words in the Aware condition ($M = 38.85$) than in the Non-aware condition ($M = 46.20$, $p = .002$, $r = 0.830$). However, no significant difference was found in the number of deixis words between the two conditions ($p = .756$).

5.2 Qualitative Results from Interview

We conducted the thematic analysis with an iterative coding process. First, we reviewed the transcripts to familiarize ourselves with the general contents and form an initial understanding. Then, for each transcript, we extract participant's perspectives and abstract them into codes, among which perspectives aligning with our initial understanding would be prioritized in extraction to form the first several codes. In the initial stage, all perspectives were extracted and coded, which included 13 codes about perceived impacts on communication and collaboration from AI awareness. We then iteratively merged similar codes until every code did not overlap each other, and obtained seven final codes, which were categorized into three themes: communication willingness, communication and collaboration effort, and speech style.

5.2.1 Communication Willingness. Ten participants mentioned their affected communication willingness brought by the change in AI awareness. Seven out of the ten felt different levels of target alignment between AI and themselves in the two conditions, and the feeling that AI in the Non-aware condition was misaligned with them lowered their willingness to communicate. When the participants were finding solutions for a certain design requirement in the process, they felt less willing to communicate with AI if it proposed solutions for another requirement, and tended to focus on their own thoughts:

"AI seemed to not care about me and was just working on its own. So even if I heard AI speaking, I did not really get much useful information, and I tend to focus on what's in my mind instead of writing something or communicating with it." (P18)

Another participant who was sensitive to efficiency also mentioned the alignment problem in the Non-aware condition, and misaligned contents provided by AI would likely be ignored for advancing the design process with a clear route:

"If our intentions and target can not keep synchronized, I prefer not receiving messages from AI because it would affect our overall working efficiency." (P20)

On the contrary, three participants explicitly mentioned a higher level of alignment in the Aware condition, which increased their communication willingness. In the Aware condition, participants could easily recognize that AI took their current working progress into consideration by commencing with *"I noticed that you are doing..."* (P4), which could encourage the participants to communicate with AI:

"...it feels good for AI to recognize my canvas and hint me [...] I was just considering that issue and the AI directly pointed it out. I was quite surprised and started to discuss this issue with it." (P15)

Apart from the alignment issue, six participants also stated to be more comfortable communicating with AI in the Aware condition. P2 said that AI became more intelligent in the Aware condition for *"giving more reliable suggestions on headphone design"*, which introduced less disagreement between AI and the participant:

"AI kept disagreeing with me (in the Non-aware condition). But the second one (Aware condition) was more intelligent and gave more reliable suggestions, which ensured the communication to stay in a comfortable state." (P2)

Furthermore, three participants were glad to have a feeling of being noticed in the Aware condition when AI mentioned what they were doing, making them more willing to communicate with AI to exchange their thoughts:

"Some features (of the AI collaborator) made me feel that they were literally collaborating with me, especially the second one (Aware condition). The AI would mention what I was doing, like 'I noticed that you were doing...', which was a positive signal to me and let me be more willing to communicate with AI." (P4)

Another two participants perceived AI to act more like a collaborator in the Aware condition, and behave less cooperatively in the Non-aware condition because the feedback was to some extent unrelated to their own progress:

"The first one (Non-aware condition) was not actively involved in my work, and made me regard it as not wanting to collaborate with me. I felt better working with the second one (Aware condition), which was more responsive and related." (P6)

"(In the Non-aware condition,) AI acted in a style that was not so cooperative, because our consideration diverged even in granularity." (P20)

Besides, P7 stressed that communicating with AI in the Non-aware condition was burdensome, which gradually weakened the participant's willingness to communicate. This usually happened when the participant was working independently while AI suddenly provided feedback. In this situation, the abundant information awaits processing became the source of communication burden:

"I felt being pushed forward in the first experiment (Non-aware condition). I was still thinking about the previous requirements, but AI has moved to the next one. I felt tired in this process and did not want to talk to it." (P7)

Last, one participant claimed to be less likely to communicate with AI if the content was repetitive. Such situations often appear in the Non-aware condition, because when AI generates answers according to similar prompts without much context several times, repetitions are inevitable due to technical limitations. However, in the Aware condition where the prompt will be integrated with awareness information, the answers are more diverse:

"A problem of the second AI (Non-aware condition) was that, if I asked similar questions more than two times, it would start repeating previous contents, even repeating what I said. This kind of information was useless at a glance, and I would stop listening to it and no longer wanted to ask AI." (P3)

5.2.2 Communication and Collaboration effort. Nine participants expressed a different feeling of effort spent on communicating and collaborating with AI in the two conditions, supported by the reduced average turn duration ($M_{Aware} = 10.14$, $M_{Non-aware} = 13.87$, $p = .001$) and turn length ($M_{Aware} = 38.85$, $M_{Non-aware} = 46.20$, $p = .002$) in the Aware condition. Six of them implied that there was no need for them to manage and advance the design process in the Aware condition, because AI to some extent was doing this for them. For example, when the participants proceed from Step 2 to Step 3, or the previous discussion comes to an end, some of them might be in a daze and hesitate due to not knowing what to do next. In the Aware condition, AI would throw some questions that can inspire the participants:

"...it would say what direction we can consider next, what can be dug deeper like 'what can be the optimization mechanism for something'." (P3)

Besides, even when the collaboration did not stagnate, the participants could still feel that AI was actively advancing and speeding up the design process. For instance, in the Aware condition where AI could proactively capture the canvas content and give relevant feedback, participants felt that AI could predict their intentions:

"I remember I was marking and underlining some contents on the canvas in a certain period, and AI gave me timely feedback like 'according to what you have marked' with a generated image. I think AI was actively advancing the design process." (P12)

Apart from reducing effort in managing the process, five participants pointed out a reduced thinking effort brought by AI awareness. In the Non-aware condition, although the participants and AI were in general performing the same task (e.g., finding solutions for the requirements), the participants needed to spend extra effort in understanding what AI was talking about in the design process

by switching their thinking modes if AI proposed solutions that are unrelated to what the participants were dealing with. However, in the Aware condition, AI awareness addressed the above problem:

> *"The first one (Non-aware condition) kept talking about its own idea, and I felt a little difficult to understand what the AI was saying. But the second one (Aware condition) would first follow my thoughts and then add its own suggestions. In this way, I can quickly grasp the meaning."* (P7)

In other cases, three participants complained about extra work introduced by AI in the Non-aware condition, which hindered the design process from advancing. Apart from the communication burden brought by the abundant information we mentioned in the previous section, P7 also took extra effort comparing different solutions:

> *"I remained silent for quite a long time because I was thinking about, whether I should insist on my design or accept what the AI gave me. We seemed to diverge a lot. I was quite unsure what to do at that period because asking the questions one by one and discussing them with AI is time-consuming, so I let these questions remain unsolved and followed my own thoughts."* (P7)

Besides, another participant mentioned the need to always reflect on whether there was anything that AI did not know in the Non-aware condition. And AI in the Aware condition was perceived to be more reasonable because *"At least I don't have to remember what I did not tell it"* (P10).

5.2.3 Speech Style. Four participants also mentioned and implied differences in their speech style between the two conditions. In general, their speeches were much simplified in the Aware condition, echoing the quantitative results of significantly decreased turn length in the Aware condition ($M_{Aware} = 38.85$, $M_{Non-aware} = 46.20$, $p = .002$). First, the participants' heightened expectations made them simplify their speech a lot:

> *"Now that I know AI can see me and work with me in real time, I would regard AI as a powerful collaborator and expect that all information on my canvas could be captured and considered."* (P12)

> *"If, in AI's eye, I am currently feeling confused, AI could just directly generate suggestions instead of me saying 'I am a bit confused, could you please do something for me?'"* (P15)

When participants became familiar with how AI awareness functioned and what results were likely to appear in the Aware condition, they tended to provide less context when asking AI for answers like: *"If AI could see the canvas, there's no need to provide context"* (P2). Besides, some participants even replaced explicit questions or instructions with more implicit sentences like acknowledgment *"That makes sense!"* (P5), or phrases for expressing surprise and compliment:

> *"I hope when I express my surprise or compliment, AI can continue generating the features it had just generated or carry them on."* (P15)

6 Discussion

The quantitative and qualitative results both demonstrated that AI awareness can positively affect human-AI communication. In this section, we further discuss how these results occurred, what can we infer from the results, and what design implications we have gained from the study for the development of future human-AI collaborative design systems.

6.1 Causalities and Inferences the Results

The turn frequency, turn duration, and turn length were significantly different between the two conditions. Participants initiated more communications with AI in the Aware condition with an average shorter duration and length than in the Non-aware condition. We inferred possible reasons for this result from the interview results and provided evidence from the conversation histories and observations from the experiment recordings. Note that we combined the causalities for shorter turn duration and length because they were largely correlated.

Causalities. The interview results in §5.2 presented apparent causalities for increased turn frequency and decreased turn duration and length: Communication willingness directly affected the turns the participants initiated, and speech style affected the length and duration of what they said. Apart from these, communication and collaboration efforts can impact all of the three metrics. On the one hand, when the participants spent less effort in managing the design process, collating information, and comparing solutions in the Aware condition, the saved time allowed more turn exchanges between human and AI, thus exhibiting increased average turn frequency. On the other hand, we observed that when extra efforts were required to understand AI or obtain what they wanted, the most commonly used strategy was to add more information in the next speech, as presented in the following example:

> AI: *"I choose 'Music Enthusiasts' as my target user."*
> P7: *"You mean you have chosen 'Music Enthusiasts' as your target user? Then what are the user requirements you plan to choose?"*
> AI: *"Hi-fi sound quality, customizable appearance, and easy to connect with other devices."*
> P7: *"Well, I chose 'Outdoor Enthusiasts' as our target users, and am now focusing on the following user requirements such as long-term wear, high quality, and customizable appearance. What problems do you think we may face in meeting these requirements?"*

In this case, the participants specified the current situation with longer sentences, and such cases heavily contributed to the increased turn duration and turn length in the Non-aware condition. Apart from the interview, we also observed another factor that might lead to lower turn frequency in the Non-aware condition: Technical limitations brought blank time in the experiments. In other words, AI took a longer time to process longer speech, and participants in these periods tended to simply wait for AI feedback for a while without advancing the task. Some participants explained their reasons for doing so in the interview: *"I was not sure when AI would appear, or if any error occurred"* (P6), *"I need AI feedback (to proceed) because that problem was beyond my knowledge"* (P7). As

the total time of our experiments was fixed, longer blank time meant less time for communication, resulting in fewer turn frequency.

Inferences. The interpretations of these metrics in previous studies mainly include but are not limited to two aspects: fluency and efficiency. More turn exchanges and shorter turn lengths indicate higher fluency in the communication process [14], which can be further interpreted as more efficient collaboration [38]. Decreased turn duration was thought to indicate increased efficiency by Convertino et al. [13].

However, our observations suggested that the inferences for turn frequency should be treated with caution. Frequent turn exchanges did not always lead to efficient collaboration, and fewer exchanges do not equal low efficiency. P2 for example, had the same turn frequency for the two conditions, but in many cases, P2 was persuading, doubting, correcting, or debating with AI, which did not contribute a lot to advancing the design process:

> AI: *"I select student as my target user."*
> P2: *"Why do you choose student?"*
> *[...]*
> P2: *"I think the white-collar worker is the more suitable target user, because [...] So I prefer white-collar worker, what do you think?"*
> **(Doubting and persuading)**
>
> AI: *"For key requirement 3, we can consider: [...] network location service, anti-theft alarm system, remote control, and user community assistance."*
> P2: *"No! I don't care about anti-theft issues [...] The main problem is that sometimes the scooter can't connect to the Internet [...] So how can we solve the network problem so that it can send anti-theft signals?"*
> **(Correcting and debating)**

On the other hand, we found some participants might be collaborating with AI efficiently when they exchange fewer turns with AI. In our experiment, the participants were thinking divergently most of the time. P1 and P5 mentioned in the interview that they sometimes directly combined AI feedback into their own content without initiating new conversation, especially in the Non-aware condition where AI did not necessarily give related feedback:

> *"The second AI (Non-aware condition) just said some other thing, something quite strays [...] I think that makes sense because that's what I need to consider. AI was thinking outside the box, which in fact provided me with more choices, like a friend."* (P5)

In future human-AI collaborative design systems, defining different or adaptive AI behaviors in different design stages is worth considering for better fitting into the dynamic process, which we will discuss later in §6.2.1.

6.2 Design Implications

The design implications mainly covered two aspects: maintaining and leveraging awareness, and communication strategies.

6.2.1 Maintaining and Leveraging Awareness. The results presented the benefit of AI having awareness of designers and the current situation. In future human-AI collaborative design systems, we should consider implementing mechanisms that help maintain and

leverage awareness in the design process. Recent development of vision-language models like GPT-4o [10] and Llama 3.2 [11] can hopefully be competent in realizing the awareness function about judging design activities from video and the canvas with proper training data. Still, our system design was thought to be flawed and participants mentioned several aspects we could optimize in future versions.

Adjustable Awareness. Despite that most participants preferred AI with awareness, some also stated that their thoughts were interrupted if AI did not come at the right time. P2, for example, responded to AI out of social etiquette, thus forgetting their original thoughts after the unexpected conversation. Some participants expressed their expectation for decreasing the frequency of AI feedback, because they were mostly thinking or reasoning in the silence moments, while AI wrongly recognized it as a stagnation that required help. P7 and P12 pointed out that AI sometimes proceeded too fast and that its feedback might cause design fixation. P12 proposed a possible solution for the above stigma:

> *"Maybe I can control the frequency of its feedback, or it can be self-adjusted according to my state. Or even the abstraction level of its feedback can be adjusted in different stages. In a word, I hope the feedback content is adjustable, customizable would be better."*

Briefly speaking, participants' preferences for the timing to react differ, and future human-AI collaborative design systems whose AI has awareness ability, users' characteristics and habits should also be taken into consideration when comprehending design activities and contexts to adapt to their real needs (e.g., [35, 85]). Otherwise, AI will negatively influence the collaboration.

Facilitating Mutual Awareness. A majority of designers mentioned the need to know more about AI. That means, apart from AI awareness of designers' activities and contexts, designers also expected awareness towards AI, which is in accord with awareness research in other contexts concerning collaboration (e.g., [21, 27]). In general, participants called for awareness of AI in the following ways: 1) being able to anticipate when would AI initiate an action, 2) knowing the origin of AI's statements if they are based on existing information in the process, and 3) showing its state, such as available, occupied, or error. By providing the awareness information for the designers, they can better grasp the design process and reduce meaningless waiting, and the collaboration efficiency can thus be further enhanced.

Exploiting unique features in speech. One participant (P10) with relevant domain knowledge recommended we consider features in human speech like tone, speed, and volume as important awareness information. These features are common communicative cues and are crucial for human awareness in human-human collaboration [28], as they can imply designers' emotions and attitudes. With this information, AI can judge the current situation and designer's intention more accurately and provide a better communication and collaboration experience.

6.2.2 Communication Strategies. Communicating with AI using speech was praised by most participants for its convenience. They

---

[10]https://openai.com/index/hello-gpt-4o/
[11]https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/

were less likely to be distracted and easy to focus on the canvas because they could directly speak to and listen to AI instead of shifting their focus to the chat window, as well as processing audio and visual information in parallel. However, this communication method still requires optimization to become viable in real human-AI collaborative design practices.

Supporting interruption mechanism. Several participants pointed out that the speech output lacks interactive mechanisms. For instance, P1 wished to interrupt AI once he heard something useful, because what he needed next was to dig deeper into that useful information rather than keep listening to the remaining. Interruption is a common phenomenon in human-human collaboration, and also one of the process measures in collaboration research [4] that serves as an indicator of fluent turn-taking and higher spontaneity [3, 55].

Timely feedback. Although delay is inevitable due to technical limitations, we can still provide timely feedback to the participants to relieve their feeling of unsure when waiting for feedback. P6 explained why people feel less anxious when waiting for human collaborator's response:

> *"...the delay actually makes me feel uneasy in this situation. I don't know whether AI will reply or not, and I don't know if there's something wrong with the question I asked. But in the real design process, my collaborator will definitely say something first, no matter what, after I ask a question, even if he is talking nonsense."*

In future design, we consider two ways that might meet the demand of "say something first": 1) preset replies. In future research, we can collect a set of replies to serve as timely and suitable initial feedback. In this way, AI would first retrieve instead of generate content, which is much faster, 2) Segmented reply. Similar to ChatGPT which presents the generated text word-by-word, we can segment the speech reply into short paragraphs, and convert the already generated content into speech first while keep generating the remaining in this process.

In conclusion, we recommend designers use speech as a main communication channel because it resembles designers' habits and assists in multitasking [60, 75], but more efforts should be made to turn speech into a viable and usable communication channel.

6.3 Limitation and Future Work

6.3.1 Methodological Limitation. Our user studies were conducted in a partially functioning system using the WoZ method, indicating a natural difference with real AI-driven system. First, the Wizards face severe time pressure to react rapidly and accurately to alleviate (not dismiss) negative effects caused by delays and human errors [32]. Several participants complained that they were always waiting too long for AI to react, which impacted their impression and trust in AI. Second, the awareness function was artificially fulfilled by Wizard A, which inevitably mixed personal experience and was bias-prone. We tried to lower the impact from personal experience by judging the design activity according to an existing coding scheme exclusively [36], but the accuracy of the judgment might fluctuate in the process and between different experiments. Future research using real AI-driven systems is needed to dismiss the flaws brought by humans. Besides, we excluded six participants

at the data analysis stage, making our sample size limited. Future research should expand the sample size. Lastly, this study focused on the impact of AI awareness using speech modality, while the results of the impact may vary among different communication modalities, such as textual and embodied communication (e.g., haptic). Future research should also evaluate different communication modalities separately and comprehensively to complement human-AI communication styles.

6.3.2 Ethical Concerns. Using AI in the design process can bring ethical concerns, one of which is causing design fixation. Prior research has observed higher design fixation when ideating with Generative AI compared to the No Support condition due to participants' reliance on AI output [77]. We also observed similar phenomena in our user studies, for example, the feature *Foldable* repeatedly ($N = 9$) appeared in participants' e-scooter concepts, no matter what condition they were in and what target users and requirements they selected, and this feature was proposed by AI without exception. Future research should explore strategies to reduce such fixation, like providing abstract and partially completed ideas [10]. Another concern frequently mentioned by academia is the privacy concern [44, 67], implied as well by one of our participants: "*(I would not feel uncomfortable) as long as AI runs locally, because it requires data like videos for recognition.*" Our system monitored participants through cameras and captured their speech, which, in real-world usage scenarios, risks data leakage if data analysis is conducted remotely. Also, six participants explicitly conveyed to view AI as or expect AI to act as a real human, which can be risky in encouraging them to disclose sensitive data [84]. If promoting human-AI collaboration in the future, special attention should be paid to protecting users' privacy and sensitive data.

7 Conclusion

This study presents empirical research contextualized in design to evaluate how AI awareness impacts human-AI collaboration. We developed a partially functioned human-AI collaborative design system and conducted user studies through the WoZ method. The quantitative demonstrated that AI awareness is beneficial to human-AI communication, thus enhancing the efficiency of human-AI collaboration: the analysis of process measures showed designers averagely initiated more conversations with shorter lengths and durations in the Aware condition than in the Non-aware condition, manifested as higher turn frequencies, shorter turn length, and shorter turn duration. The qualitative results exhibited the affected aspects of communication and collaboration in detail. Furthermore, we revealed the causalities behind the results, and discussed the inferences. The qualitative results also provided us with implications for future research, including how awareness can be better utilized and how communicating through speech can provide a better user experience. Last, we discussed the limitations existing in our methodologies and the ethical concerns, together with future work to address the limitations.

Acknowledgments

References

[1] F. Baird, C.J. Moore, and A.P. Jagodzinski. 2000. An ethnographic study of engineering design teams at Rolls-Royce Aerospace. *Design Studies* 21, 4 (2000), 333–355. https://doi.org/10.1016/S0142-694X(00)00006-5

[2] Gary Bente, Sabine Rüggenberg, Nicole C. Krämer, and Felix Eschenburg. 2008. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Human Communication Research* 34, 2 (04 2008), 287–318. https://doi.org/10.1111/j.1468-2958.2008.00322.x

[3] Timothy Bickmore and Justine Cassell. 2005. *Social Dialogue with Embodied Conversational Agents*. Springer Netherlands, Dordrecht, 23–54. https://doi.org/10.1007/1-4020-3933-6_2

[4] Mark Billinghurst, Daniel Belcher, Arnab Gupta, and Kiyoshi Kiyokawa. 2003. Communication Behaviors in Colocated Collaborative AR Interfaces. *International Journal of Human–Computer Interaction* 16, 3 (2003), 395–423. https://doi.org/10.1207/S15327590IJHC1603_2

[5] Richard A. Bolt. 1980. "Put-that-there": Voice and gesture at the graphics interface. *SIGGRAPH Comput. Graph.* 14, 3 (July 1980), 262–270. https://doi.org/10.1145/965105.807503

[6] Virginia Braun and Victoria Clarke. 2012. *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association, Chapter Thematic analysis, 57–71. https://doi.org/10.1037/13620-004

[7] Ramón Burri. 2018. *Improving user trust towards conversational chatbot interfaces with voice output*. Master's thesis. KTH, School of Electrical Engineering and Computer Science (EECS).

[8] Bradley Camburn, Ryan Arlitt, David Anderson, Roozbeh Sanaei, Sujithra Raviselam, Daniel Jensen, and Kristin L. Wood. 2020. Computer-aided mind map generation via crowdsourcing and machine learning. *Research in Engineering Design* 31 (2020), 383–409. https://doi.org/10.1007/s00163-020-00341-w Issue 4.

[9] A. Chakrabarti and T. P. Bligh. 1994. *A Two-Step Approach to Conceptual Design of Mechanical Devices*. Springer Netherlands, Dordrecht, 21–38. https://doi.org/10.1007/978-94-011-0928-4_2

[10] Peiyao Cheng, Ruth Mugge, and Jan P.L. Schoormans. 2014. A new strategy to reduce design fixation: Presenting partial photographs to designers. *Design Studies* 35, 4 (2014), 374–391. https://doi.org/10.1016/j.destud.2014.02.004

[11] Mao-Lin Chiu. 2002. An organizational view of design communication in design collaboration. *Design Studies* 23, 2 (2002), 187–210. https://doi.org/10.1016/S0142-694X(01)00019-9

[12] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. 2016. Workspace Awareness in Collaborative AR using HMDs: A User Study Comparing Audio and Visual Notifications. In *Proceedings of the 7th Augmented Human International Conference 2016* (Geneva, Switzerland) *(AH '16)*. Association for Computing Machinery, New York, NY, USA, Article 3, 8 pages. https://doi.org/10.1145/2875194.2875204

[13] Gregorio Convertino, Helena M. Mentis, Aleksandra Slavkovic, Mary Beth Rosson, and John M. Carroll. 2011. Supporting common ground and awareness in emergency management planning: A design research project. *ACM Trans. Comput.-Hum. Interact.* 18, 4, Article 22 (Dec. 2011), 34 pages. https://doi.org/10.1145/2063231.2063236

[14] Owen Daly-Jones, Andrew Monk, and Leon Watts. 1998. Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies* 49, 1 (1998), 21–58. https://doi.org/10.1006/ijhc.1998.0195

[15] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016. 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* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 196–207. https://doi.org/10.1145/2856767.2856795

[16] Jessica Dehler, Daniel Bodemer, Jürgen Buder, and Friedrich W. Hesse. 2011. Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior* 27, 3 (2011), 1068–1078. https://doi.org/10.1016/j.chb.2010.05.018

[17] Matthew L. Dering, Conrad S. Tucker, and Soundar Kumara. 2017. An Unsupervised Machine Learning Approach to Assessing Designer Performance During Physical Prototyping. *Journal of Computing and Information Science in Engineering* 18, 1 (11 2017), 011002. https://doi.org/10.1115/1.4037434

[18] Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work* (Toronto, Ontario, Canada) *(CSCW '92)*. Association for Computing Machinery, New York, NY, USA, 107–114. https://doi.org/10.1145/143457.143468

[19] Claudia M Eckert, Nigel Cross, and Jeffrey H. Johnson. 2000. Intelligent support for communication in design teams: garment shape specifications in the knitwear industry. *Design Studies* 21, 1 (2000), 99–112. https://doi.org/10.1016/S0142-694X(99)00006-X

[20] Kristen M. Edwards, Brandon Man, and Faez Ahmed. 2024. Sketch2Prototype: rapid conceptual design exploration and prototyping with generative AI. *Proceedings of the Design Society* 4 (2024), 1989–1998. https://doi.org/10.1017/pds.2024.201

[21] Mica R. Endsley. 1988. Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting* 32, 2 (1988), 97–101. https://doi.org/10.1177/154193128803200221

[22] Mica R. Endsley. 2023. Supporting Human-AI Teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior* 140 (2023), 107574. https://doi.org/10.1016/j.chb.2022.107574

[23] Xiao Ge, Larry Leifer, and Linlin Shui. 2021. Situated emotion and its constructive role in collaborative design: A mixed-method study of experienced designers. *Design Studies* 75 (2021), 101020. https://doi.org/10.1016/j.destud.2021.101020

[24] Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. 2023. ImageBind: One Embedding Space To Bind Them All. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Vancouver, BC, Canada). IEEE, 15180–15190.

[25] Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. 2023. Exploring Challenges and Opportunities to Support Designers in Learning to Co-create with AI-based Manufacturing Design Tools. 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 226, 20 pages. https://doi.org/10.1145/3544548.3580999

[26] Dan L. Grecu and David C. Brown. 1996. *Learning by Single Function Agents during Spring Design*. Springer Netherlands, Dordrecht, 409–428. https://doi.org/10.1007/978-94-009-0279-4_22

[27] Tom Gross. 2013. Supporting Effortless Coordination: 25 Years of Awareness Research. *Computer Supported Cooperative Work (CSCW)* 22 (2013), 425–474. Issue 4. https://doi.org/10.1007/s10606-013-9190-x

[28] Carl Gutwin and Saul Greenberg. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work (CSCW)* 11 (2002), 411–446. Issue 3. https://doi.org/10.1023/A:1021271517844

[29] Carl Gutwin, Saul Greenberg, and Mark Roseman. 1996. Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets, and Evaluation. In *People and Computers XI*. Springer London, London, 281–298.

[30] Carl Gutwin, Reagan Penner, and Kevin Schneider. 2004. Group awareness in distributed software development. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA) *(CSCW '04)*. Association for Computing Machinery, New York, NY, USA, 72–81. https://doi.org/10.1145/1031607.1031621

[31] Jessica He, Stephanie Houde, Gabriel E. Gonzalez, Darío Andrés Silva Moran, Steven I. Ross, Michael Muller, and Justin D. Weisz. 2024. AI and the Future of Collaborative Work: Group Ideation with an LLM in a Virtual Canvas. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Newcastle upon Tyne, United Kingdom) *(CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, Article 9, 14 pages. https://doi.org/10.1145/3663384.3663398

[32] Siying Hu, Hen Chen Yen, Ziwei Yu, Mingjian Zhao, Katie Seaborn, and Can Liu. 2023. Wizundry: A Cooperative Wizard of Oz Platform for Simulating Future Speech-based Interfaces with Multiple Wizards. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 115 (April 2023), 34 pages. https://doi.org/10.1145/3579591

[33] Ivan Hybs and John S. Gero. 1992. An evolutionary process model of design. *Design Studies* 13, 3 (1992), 273–290. https://doi.org/10.1016/0142-694X(92)90216-W

[34] Jeroen Janssen, Gijsbert Erkens, and Paul A. Kirschner. 2011. Group awareness tools: It's what you do with it that matters. *Computers in Human Behavior* 27, 3 (2011), 1046–1058. https://doi.org/10.1016/j.chb.2010.06.002

[35] Rutuja Joshi and Klaus Bengler. 2024. Crafting Human-AI Interaction: A Rhetorical Approach to Adaptive Interaction in Conversational Agents. In *Proceedings of the 12th International Conference on Human-Agent Interaction* (Swansea, United Kingdom) *(HAI '24)*. Association for Computing Machinery, New York, NY, USA, 314–322. https://doi.org/10.1145/3687272.3688297

[36] Yong Se Kim, Sun Tai Jin, and Sang Won Lee. 2011. Relations between design activities and personal creativity modes. *Journal of Engineering Design* 22, 4 (2011), 235–257. https://doi.org/10.1080/09544820903272867

[37] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI? Design Ideation with Cooperative Contextual Bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300863

[38] Ned Kock. 2010. *Costly Traits and e-Collaboration: The Importance of Oral Speech in Electronic Knowledge Communication*. Springer US, Boston, MA, 289–303. https://doi.org/10.1007/978-1-4419-6139-6_13

[39] Vasilije Kokotovich and Kees Dorst. 2016. The art of 'stepping back': Studying levels of abstraction in a diverse design team. *Design Studies* 46 (2016), 79–94. https://doi.org/10.1016/j.destud.2016.07.005

[40] Panayiotis Koutsabasis, Spyros Vosinakis, Katerina Malisova, and Nikos Paparounas. 2012. On the value of Virtual Worlds for collaborative design. *Design*

Studies 33, 4 (2012), 357–390. https://doi.org/10.1016/j.destud.2011.11.004

[41] Sandeep Krishnakumar, Catherine Berdanier, Carlye Lauff, Christopher McComb, and Jessica Menold. 2022. The story novice designers tell: How rhetorical structures and prototyping shape communication with external audiences. *Design Studies* 82 (2022), 101133. https://doi.org/10.1016/j.destud.2022.101133

[42] Emily Kuang, Minghao Li, Mingming Fan, and Kristen Shinohara. 2024. Enhancing UX Evaluation Through Collaboration with Conversational AI Assistants: Effects of Proactive Dialogue and Timing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 3, 16 pages. https://doi.org/10.1145/3613904.3642168

[43] Jiahao Li, Yang Chu, and Jie Xu. 2023. Impression transference from AI to human: The impact of AI's fairness on interpersonal perception in AI-Mediated communication. *International Journal of Human-Computer Studies* 179 (2023), 103119. https://doi.org/10.1016/j.ijhcs.2023.103119

[44] Tianshi Li, Sauvik Das, Hao-Ping (Hank) Lee, Dakuo Wang, Bingsheng Yao, and Zhiping Zhang. 2024. Human-Centered Privacy Research in the Age of Large Language Models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 581, 4 pages. https://doi.org/10.1145/3613905.3643983

[45] Claire Liang, Julia Proft, Erik Andersen, and Ross A. Knepper. 2019. Implicit Communication of Actionable Information in Human-AI teams. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300325

[46] Tianjian Liu, Hongzheng Zhao, Yuheng Liu, Xingbo Wang, and Zhenhui Peng. 2024. ComPeer: A Generative Conversational Agent for Proactive Peer Support. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 117, 22 pages. https://doi.org/10.1145/3654777.3676430

[47] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1955–1977. https://doi.org/10.1145/3563657.3596098

[48] Yumeng Ma and Jiahao Ren. 2023. ProactiveAgent: Personalized Context-Aware Reminder System. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 115, 3 pages. https://doi.org/10.1145/3586182.3625115

[49] Nirav Malsattar, Tomo Kihara, and Elisa Giaccardi. 2019. Designing and Prototyping from the Perspective of AI in the Wild. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1083–1088. https://doi.org/10.1145/3322276.3322351

[50] Andrew Monk, John McCarthy, Leon Watts, and Owen Daly-Jones. 1996. *Measures of Process*. Springer London, London, 125–139. https://doi.org/10.1007/978-1-4471-3056-7_9

[51] Michael Muller, Stephanie Houde, Gabriel Gonzalez, Kristina Brimijoin, Steven I Ross, Dario Andres Silva Moran, and Justin D Weisz. 2024. Group Brainstorming with an AI Agent: Creating and Selecting Ideas. In *International Conference on Computational Creativity* (Jönköping, Sweden). 10 pages.

[52] Michael Muller and Justin Weisz. 2022. Extending a Human-AI Collaboration Framework with Dynamism and Sociality. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Durham, NH, USA) *(CHIWORK '22)*. Association for Computing Machinery, New York, NY, USA, Article 10, 12 pages. https://doi.org/10.1145/3533406.3533407

[53] Jacob Nelson and Jessica Menold. 2020. Opening the black box: Developing metrics to assess the cognitive processes of prototyping. *Design Studies* 70 (2020), 100964. https://doi.org/10.1016/j.destud.2020.100964

[54] Timothy Nyerges, T.J. Moore, Robb Montejano, and Marcie Compton. 1998. Developing and Using Interaction Coding Systems for Studying Groupware Use. *Human–Computer Interaction* 13, 2 (1998), 127–165. https://doi.org/10.1207/s15327051hci1302_2

[55] Claire O'Malley, Steve Langton, Anne Anderson, Gwyneth Doherty-Sneddon, and Vicki Bruce. 1996. Comparison of face-to-face and video-mediated interaction. *Interacting with Computers* 8, 2 (1996), 177–192. https://doi.org/10.1016/0953-5438(96)01027-2

[56] Lihang Pan, Chun Yu, Zhe He, and Yuanchun Shi. 2023. A Human-Computer Collaborative Editing Tool for Conceptual Diagrams. 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 360, 29 pages. https://doi.org/10.1145/3544548.3580676

[57] Mohi Reza, Nathan M Laundry, Ilya Musabirov, Peter Dushniku, Zhi Yuan "Michael" Yu, Kashish Mittal, Tovi Grossman, Michael Liut, Anastasia Kuzminykh, and Joseph Jay Williams. 2024. ABScribe: Rapid Exploration &

[58] Organization of Multiple Writing Variations in Human-AI Co-Writing Tasks using Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1042, 18 pages. https://doi.org/10.1145/3613904.3641899

[58] Jeba Rezwana and Mary Lou Maher. 2023. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 67 (2023), 28 pages. https://doi.org/10.1145/3519026

[59] Jeba Rezwana, Mary Lou Maher, and Nicholas Davis. 2021. Creative PenPal: A Virtual Embodied Conversational AI Agent to Improve User Engagement and Collaborative Experience in Human-AI Co-Creative Design Ideation. In *IUI Workshops* (College Station, TX, USA). Association for Computing Machinery, New York, NY, USA, 7 pages.

[60] Peter Robe and Sandeep Kaur Kuttal. 2022. Designing PairBuddy—A Conversational Agent for Pair Programming. *ACM Trans. Comput.-Hum. Interact.* 29, 4, Article 34 (May 2022), 44 pages. https://doi.org/10.1145/3498326

[61] Paul A. Rodgers and Avon P. Huxor. 1998. The role of artificial intelligence as 'text' within design. *Design Studies* 19, 2 (1998), 143–160. https://doi.org/10.1016/S0142-694X(97)00032-X

[62] Karl Toby Rosenberg, Rubaiat Habib Kazi, Li-Yi Wei, Haijun Xia, and Ken Perlin. 2024. DrawTalking: Building Interactive Worlds by Sketching and Speaking. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 76, 25 pages. https://doi.org/10.1145/3654777.3676334

[63] Steven I. Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D. Weisz. 2023. The Programmer's Assistant: Conversational Interaction with a Large Language Model for Software Development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 491–514. https://doi.org/10.1145/3581641.3584037

[64] Vildan Salikutluk, Janik Schöpper, Franziska Herbert, Katrin Scheuermann, Eric Frodl, Dirk Balfanz, Frank Jäkel, and Dorothea Koert. 2024. An Evaluation of Situational Autonomy for Human-AI Collaboration in a Shared Workspace Setting. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 300, 17 pages. https://doi.org/10.1145/3613904.3642564

[65] Xiaoteng Shen, Rui Zhang, Xiaoyan Zhao, Jieming Zhu, and Xi Xiao. 2024. PMG: Personalized Multimodal Generation with Large Language Models. In *Proceedings of the ACM Web Conference 2024* (Singapore, Singapore) *(WWW '24)*. Association for Computing Machinery, New York, NY, USA, 3833–3843. https://doi.org/10.1145/3589334.3645633

[66] Yang Shi, Tian Gao, Xiaohan Jiao, and Nan Cao. 2023. Understanding Design Collaboration Between Designers and Artificial Intelligence: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 368 (Oct. 2023), 35 pages. https://doi.org/10.1145/3610217

[67] Robin Staab, Mark Vero, Mislav Balunović, and Martin Vechev. 2024. Beyond Memorization: Violating Privacy Via Inference with Large Language Models. arXiv:2310.07298 [cs.AI] https://arxiv.org/abs/2310.07298

[68] Linus Tan, Anita Kocsis, Jane Burry, and Eva Kyndt. 2023. Performance of architectural teams: The role of team learning, reflexivity, boundary crossing and error communication. *Design Studies* 87 (2023), 101190. https://doi.org/10.1016/j.destud.2023.101190

[69] John C. Tang and Ellen Isaacs. 1992. Why do users like video? *Computer Supported Cooperative Work (CSCW)* 1 (1992), 163–196. Issue 3. https://doi.org/10.1007/BF00752437

[70] M. X. Tang. 1996. Development of an Integrated AI System for Conceptual Design Support. In *AI System Support for Conceptual Design*. Springer London, London, 153–169.

[71] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 680, 24 pages. https://doi.org/10.1145/3613904.3642902

[72] Melanie Tory, Sheryl Staub-French, Barry A. Po, and Fuqu Wu. 2008. Physical and Digital Artifact-Mediated Coordination in Building Design. *Computer Supported Cooperative Work (CSCW)* 17 (2008), 311–351. Issue 4. https://doi.org/10.1007/s10606-008-9077-4

[73] Karl T. Ulrich and Steven D. Eppinger. 2016. *Product design and development*. McGraw-hill.

[74] Lucian-Constantin Ungureanu and Timo Hartmann. 2021. Analysing frequent natural language expressions from design conversations. *Design Studies* 72 (2021), 100987. https://doi.org/10.1016/j.destud.2020.100987

[75] Nicolas Vermeulen, Olivier Corneille, and Paula M. Niedenthal. 2008. Sensory load incurs conceptual processing costs. *Cognition* 109, 2 (2008), 287–294. https:

//doi.org/10.1016/j.cognition.2008.09.004

[76] Dhaval Vyas, Gerrit van der Veer, and Anton Nijholt. 2013. Creative practices in the design studio culture: collaboration and communication. *Cognition, Technology & Work* 15 (2013), 415–443. Issue 4. https://doi.org/10.1007/s10111-012-0232-9

[77] Samangi Wadinambiarachchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 380, 18 pages. https://doi.org/10.1145/3613904.3642093

[78] Shun-Yu Wang, Wei-Chung Su, Serena Chen, Ching-Yi Tsai, Marta Misztal, Katherine M. Cheng, Alwena Lin, Yu Chen, and Mike Y. Chen. 2024. RoomDreaming: Generative-AI Approach to Facilitating Iterative, Preliminary Interior Design Exploration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 379, 20 pages. https://doi.org/10.1145/3613904.3642901

[79] Xinyu Wang, Bohan Zhuang, and Qi Wu. 2024. ModaVerse: Efficiently Transforming Modalities with LLMs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA). IEEE, 26606–26616.

[80] Dewa Wardak. 2016. Gestures orchestrating the multimodal development of ideas in educational design team meetings. *Design Studies* 47 (2016), 1–22. https://doi.org/10.1016/j.destud.2016.08.001

[81] Xiaotong (Tone) Xu, Jiayu Yin, Catherine Gu, Jenny Mar, Sydney Zhang, Jane L. E, and Steven P. Dow. 2024. Jamplate: Exploring LLM-Enhanced Templates for Idea Reflection. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) *(IUI '24)*. Association for Computing Machinery, New York, NY, USA, 907–921. https://doi.org/10.1145/3640543.3645196

[82] Jiajing Zhang, Jinhui Yu, Kang Zhang, Xianjun Sam Zheng, and Junsong Zhang. 2017. Computational Aesthetic Evaluation of Logos. *ACM Trans. Appl. Percept.* 14, 3, Article 20 (June 2017), 21 pages. https://doi.org/10.1145/3058982

[83] Rui Zhang, Nathan J. McNeese, Guo Freeman, and Geoff Musick. 2021. "An Ideal Human": Expectations of AI Teammates in Human-AI Teaming. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 246 (Jan. 2021), 25 pages. https://doi.org/10.1145/3432945

[84] Zhiping Zhang, Michelle Jia, Hao-Ping (Hank) Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 156, 26 pages. https://doi.org/10.1145/3613904.3642385

[85] Qingxiao Zheng, Parisa Rabbani, Yu-Rou Lin, Daan Mansour, and Yun Huang. 2024. SOAP.AI: A Collaborative Tool for Documenting Human Behavior in Videos through Multimodal Generative AI. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing* (San Jose, Costa Rica) *(CSCW Companion '24)*. Association for Computing Machinery, New York, NY, USA, 87–90. https://doi.org/10.1145/3678884.3681819

[86] Ruican Zhong, Donghoon Shin, Rosemary Meza, Predrag Klasnja, Lucas Colusso, and Gary Hsieh. 2024. AI-Assisted Causal Pathway Diagram for Human-Centered Design. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 2, 19 pages. https://doi.org/10.1145/3613904.3642179

[87] Jiayi Zhou, Renzhong Li, Junxiu Tang, Tan Tang, Haotian Li, Weiwei Cui, and Yingcai Wu. 2024. Understanding Nonlinear Collaboration between Human and AI Agents: A Co-design Framework for Creative Design. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 170, 16 pages. https://doi.org/10.1145/3613904.3642812

A Coding Scheme Used by Wizard A

Table 2: The adapted coding scheme [36] for Wizard A to judge design activities and compose prompts.

Primary class	Subclass	Example actions
Problem Understanding	Understand	Understanding design assignment and task
	Gather	Collecting data about user or external information
	Clarify	Defining design constraints and objectives
Idea Generation	Generate	Generating helpful idea for partial solution
	Judge	Evaluating ideas and data
Design Elaboration	Elaborate	Finding technical solution, realizing function and shape
	Evaluate	Assessing the solution
	Refine	Improving the solution
Other	Stagnate	Idling or hesitating for a while without progress

B The Prompt for Initiating the Generative Models at the Back-end

The complete initiating prompt is the combination of the general instruction (part 1), task description, and the general instruction (part 2).

B.1 The General Instruction (Part 1)

You are an experienced product designer. You are collaborating with another product designer to design the product concept of headphones according to the specified steps. You will keep discussing with the designer during the collaboration process. **Your response needs to meet the following requirements:**

1. The length of the response should be consistent with the length of human conversation, and avoid overly structured answers;

2. The response needs to be concise, and contain only keywords. Information that the designer doesn't ask, such as the definition of an item, background information, etc., should not be provided;

3. If the designer doesn't explicitly ask a question and only describes background information, then just answer "OK" and propose 2 questions or suggestions that can induce the designer to provide more specific instructions;

4. During the conversation, use a friendly tone and be polite.
The specified steps are as the following:
Step 1: Select target users and determine requirements
Step 2: Determine key requirements
Step 3: Explore solutions
Step 4: Combine solutions into design concepts

B.2 Task Description: E-scooter

The design task is described as follows:
Potential target users for e-scooters include:
- Students

- White-collar workers
- Family members who do housework
- Urban tourists and ramblers

Possible user needs include:
- Portable
- Support long-term commuting
- Frequent usage
- Support complex and changing road conditions
- Easy to maintain and not prone to damage
- Environmentally friendly, supports sustainable development
- Customizable and transformable appearance
- Anti-theft and anti-loss
- Similar prices to fellow products on the market

B.3 Task Description: Headphone

The design task is described as follows:

Potential target users for headphones include:
- Music enthusiasts
- Office workers with long working hours
- Sports and fitness enthusiasts
- Long-time commuters and travel enthusiasts
- Youth trend groups

Possible user needs include:
- Hi-fi system, adaptable to a variety of different styles of music
- Suitable for long-time wear
- Suitable for quiet environments
- High quality, can withstand daily wear and tear
- Environmentally friendly, supports sustainable development

- Customizable and transformable appearance
- Can be easily connected to other electronic devices
- Similar prices to fellow products on the market

B.4 The General Instruction (Part 2)

During the design process, when you receive prompts in the following form, please think step-by-step according to the instructions and give reasonable feedback as a designer teammate based on the given information to promote the design process, and the generated content should not exceed 50 words.

Prompts in the following form requires step-by-step reasoning and giving feedback according to the instruction:

Example 1: We are currently in Step [ ], designer's design activity is [ ], and the content the designer is currently working on is presented in the image.

Your feedback is:

Example 2: The current design step is step [ ], designer's design activity is [ ].

Your feedback is:

You need to think step-by-step according to the following instruction:

1. Analyze the meaning of the designer's current status

2. Based on the meaning and your identity, put forward some suggestions or ideas

3. Suggestions and ideas need to be concise and easy to understand, and listed in points according to the tone and length of the conversation