Human Systems Engineering and Design (IHSED 2024), Vol. 158, 2024, 1–9
https://doi.org/10.54941/ahfe1005523

Exploring the Potential of Generative Artificial Intelligence in Human-Centred Design

Susanna Aromaa and Hanna Lammi

VTT Technical Research Centre of Finland Ltd., P. O. Box 1300, 33101 Tampere, Finland

ABSTRACT

This study evaluated the suitability of a generative artificial intelligence (AI) tool for impersonating human test participants in human-centred design (HCD) research. The experiment involved 23 office workers who gave feedback on a telepresence robot concept through an online questionnaire. The same questionnaire was answered by the AI tool (Copilot Microsoft Edge), using the background information of the human participants as an input. The results revealed that the human and AI groups had contradictory answers to the question of benefits of the concept: the human participants did not see much value in the telepresence robot system, while the AI participants gave more positive responses. Otherwise, answers were similar in relation to identifying challenges and improvement ideas. However, the AI tool suggested more improvement ideas than human participants. The findings suggest that even though there is potential in utilising AI in HCD (e.g., inventing ideas) there are still challenges to overcome, for example, in understanding the use context and experiences of users. These findings can be used by researchers and designers when utilising AI in HCD, and AI technology developers who aim to improve AI tools to better mimic human cognitive processes.

Keywords: Human-centred design, Generative artificial intelligence, User study

INTRODUCTION

When developing new technological solutions, it is important to consider human-centred design (HCD). Human-centred design involves understanding the requirements of users, involving them in the design process, and then testing the created designs with them (ISO 9241-210, 2019). In the future, HCD practices may change due to the emergence of generative artificial intelligence (AI) which has opened up new possibilities how to design interactive systems (Schmidt et al., 2024).

Artificial intelligence can be defined as the capability of machines to execute cognitive tasks that are usually done by human brains such as perceiving, reasoning, learning, interacting with the environment, problem solving, and creativity (Chui et al., 2020). Generative AI tools thus have the ability to create novel content, allowing for human-like conversations and interactions (Murphy, 2022).

Many usage possibilities have been identified for AI in HCD, for example, in stakeholder identification, persona and scenario creation, ideation,

question design and simulating user responses (Schmidt et al., 2024). It can be also used for producing design solutions, automatically developing solutions, evaluating the design and understanding the context of use (Stige et al., 2023). Additionally, there has been research efforts focusing on how to use generative AI models in the creation of personas (Salminen et al., 2023).

Li et al. (2024) interviewed user experience designers of how they perceive generative AI tools. They reported that there is a potential for generative AI to assist them in their work even though there are still some concerns (e.g., copyright, human creativity, AI literacy). Hess et al. (2024) agree that opportunities and limitations of using AI in HCD research exists.

Based on current publications related to AI use in HCD, it can be seen that it is in its early stages and, therefore, more research is required to fully learn in which cases generative AI tools offer benefits and where they are not useful.

This experimental study aimed to evaluate the suitability of generative AI tools in impersonating human test participants in HCD research. The goal was to use a visually assisted large language model (LLM) to test its ability to represent real end users in user-based studies. The end goal is not to replace real end users in HCD, but to enhance the possibilities of better acknowledging their needs and expectations during the design process.

MATERIAL AND METHODS

The goal of this study was to explore the potential of generative AI in HCD. Especially, to find out if it could be used to provide similar answers to open questions than human participants. The study was experimental and comparative, and it was executed in Spring 2024. The experiment used a visually assisted LLM, Copilot Microsoft Edge, and involved 23 human participants working in office settings, either on-site or remotely. A concept of a telepresence robot was presented to the participants online (see Table 1), and their feedback was gathered through a questionnaire. The questionnaire included background information on the participants as well as three open-ended questions related to the benefits, challenges, and potential improvements of the telepresence robot concept.

Table 1. Concept description of the robot system.

What if you were able to use the robot system in the picture at your workplace when you are working from home/a remote location? This small, mobile robot system moves around on four wheels and has a robotic arm that can pick up and move objects. It also has a tablet screen that allows you to communicate with your colleagues through a video chat application. You can control the robot remotely. With the robot system, you can move around your workplace, talk to people and handle objects, all from your remote location.	

The collected background information was entered into the AI tool one person at a time and the AI tool was asked to answer the same three questions as if it was the described person. For each persona, the AI tool received three prompts: the person's background information ("I want you to be a following

type of person. Your characteristics are listed below…"), the scenario description ("Please, read the following description of the robot system and imagine if you would use it at work. Then, wait for following instructions…") and open-ended questions ("I would like you to answer to the following three open-ended questions as the person described previously…"). The results from both participant groups were qualitatively analysed (e.g., differences in the content and style were compared). Two human factors researchers with over 20 years of experience in HCD performed analyses individually and then discussed their findings together.

Participants

To create personas for the AI tool, the following background information was gathered from the actual participants. Their average age was 45 years old (range 26-65), and the participants included 12 male, 10 female, and one person who did not reveal their gender. They had a high level of education and they worked at the same research institute in office or lab environments. Most of them dealt with topics related to engineering and social sciences, but other topics were also reported. The majority of them had several years of work experience (e.g., 16 had more than 15 years of experience). Eleven participants had no prior experience with robot systems, while 10 had some experience and two were regular users of robot systems. Regarding technology adoption level (Rogers et al., 2014), most of the participants reported to belong to the early or late majority. None of them identified as laggards. Based on the Big Five personality traits (McCrae & John, 1992), the participants were more open (22) than closed (1), more conscientious (18) than spontaneous (5), more introverted (14) than extroverted (9), more agreeable (19) than hostile (4) and more stable (14) than neurotic (9).

The study setup was consulted with an ethical board member at the research institute, and it was agreed that the evaluation of an ethical committee was not required. However, the ethical review was still conducted to ensure a smooth publication process. According to the ethical committee's decision, the study follows ethical principles and guidelines. For the study participants, the online questionnaire's front page contained a short introduction to the study and links to the information sheet and privacy notice documents. Participants were also informed that by proceeding to the questionnaire, they gave their consent to the study.

RESULTS

This section describes the benefits, challenges and improvement ideas that both human and AI groups suggested. It also explains how both groups formulated their answers.

Human Group

The majority of the participants did not find this type of robot system useful for their work (19/23). They stated that their work involved knowledge work in office settings and the robot would not add value for their work: "*I work at the computer on my desk so I cannot see much use of this system for my*

work". Many said that existing tools such as online meeting systems, emails and chats were sufficient for them. One person could not assess the benefits because they either did not understand the user interface of the system or it was not described clearly enough. Three people favored using the system in their work. They said that when the system is improved further, it could enable natural and creative collaboration. It could improve the viewing of physical objects (e.g., laboratory equipment, machines) and enable them to move and interact in the office space. Although many said that they did not see the advantages of using the system in their work they mentioned situations where it could be helpful, for example, brainstorming sessions, workshops, wandering in the office and meeting people casually, observing things in a laboratory, in a trial or a pilot event, and retrieving a component for someone from a storeroom that only they have access to.

Moving around was seen as one major challenge related to the concept of the robot system. The participants wondered how the system could pass (locked) doors, use elevator, handle different floor materials and heights, and access required areas. Additionally, there might be obstacles in the environment or people moving around which increases the possibility of collisions. Related to navigation, the participants were concerned that using the system might feel slow and frustrating if it required constant control or monitoring while the robot moved: "*… probably I would lose time when navigating the robot in the office and trying to get a feeling of what is around me…*" One concern was that whether the robot system provided any benefit or value or could do anything useful. This value aspect could also have a negative influence on acceptance. The current size of the robot was discussed from different points of views. For example, could it reach the tabletop, whether people could have a conversation with it while standing or maybe people might stumble over it. Security and privacy issues related to (personal) data were seen as another challenge as well as latency in the internet connection. Another concern was how people who work at the office experience the robot. They might find the system odd or feel like it was spying on them or interrupting them in their work. The experience and interaction might differ from seeing someone in person at the office. Additionally, it could be challenging for a person near the robot to tell when it was "occupied" with a human user and when it was idle. There were also other challenges mentioned regarding usability, reliability and cost.

Many of the participants said that it is difficult to propose improvement ideas based on the concept description without using the system. Some of the participants suggested raising the tablet so that it would be at eye-level with people at the office to provide more natural interaction: "*I would make the system higher. Now it seems strange, that the picture of the person is so low*.". Three people did not come up with any ideas for improvements because they did not understand the concept and see any clear use case for that. However, they observed that this kind of concept might suit industrial settings better and offer more opportunities there. Two people requested more human like appearance instead of technical looking system. The participants mentioned that the system should be mobile, agile and easily detected at the office, for example, it would make sound while moving around. It could have haptics

and grabbing capabilities, for example, to have two hands so that lifting would be more stable and even heavier loads could be lifted. One person suggested a hologram instead of tablet PC. They highlighted also issues to consider related to latency, control devices, security, privacy and safety. One proposed AI advanced features such as to improve the navigation when the robot is controlled manually and using AI assistance in user interfaces to improve usability. Having a better name for the robot based on its purpose would be nice (e.g., compared to vacuum cleaner). One improvement idea was that the robot could roam around and pop up with colleagues randomly simulating the natural interaction at the office. The participants also wanted to personalise the robot system someway. Another person hoped it would be able to do fine tasks such as handling a pen and writing.

Artificial Intelligence Group

All the AI participants (23/23) thought that the robot system would be beneficial in their work. They used positive expressions such as "*several benefits*", "*immense potential*" and "*significant benefits*". Collaborating was mentioned by several AI persons as a benefit of the concept. Object manipulation and enhanced remote presence were also seen as beneficial. Additionally, they agreed that the robot could improve efficiency, productivity and situation awareness. The AI participants also mentioned other benefits such as exploring the environment, data collection, lab visits and assistance, real-time seminars, security, inclusivity and enhanced mobility. In total, there was 25 different benefits mentioned by AI participants.

Privacy and security issues were brought up by all the AI participants as a key challenge in deploying robot system at the workplace: "*Ensuring privacy—both for remote workers and on-site employees—is critical. Clear guidelines on data handling, video recording, and consent are essential.*". Acceptance issues were mentioned often (e.g., social acceptance, culture, social norms, office dynamics). Another notable challenge mentioned was physical adaptation of the environment (e.g., physical obstacles). The integration of the system with the existing infrastructure and protocols could pose challenges. Cost, network/connectivity and training were also mentioned. Navigation and battery life were also acknowledged. Other challenges that were noted were related to policy and guidelines, cost-benefit analysis, safety measures, technical hurdles, human-robot interaction balance and user interface design.

Artificial intelligence participants suggested more than 30 ways to improve the robot system. The most common ideas were customizable and personalized appearance, intuitive and user-friendly interface, safety features and protocols, and energy efficiency (for example, longer battery life). Adaptive navigation and learning algorithms were also mentioned as areas for improvement. Gesture recognition and gesture interfaces also came up. Privacy features and models, as well as customization and customized access levels, were seen as important. Emotions were seen as a possible improvement from both perspectives: emotive feedback from the robot and emotion

recognition. Another suggestion was to use virtual reality technologies when operating the robot system remotely. Integrating the robot system with other tools such as lab equipment and APIs for Microsoft Office tools were also proposed as a development idea. Advanced sensing and sensors, and enhanced object manipulation, were also considered beneficial. Moreover, adjustable height was seen as important from ergonomics point of view. Furthermore, integrating AI to the system could enhance the interaction with the robot. Modularity, autonomous charging, and enhanced autonomy were also listed.

The Formatting and Style of Responses

Human participants expressed their own feelings especially when considering how the system could help them in their work, for example, "*Well, I very much doubt that this kind of robot would be something that I would need in my work*" or "*As an information / office worker, I do not see significant benefit in using this kind of robot system. … the capability to physically manipulate objects does not add a lot to the equation, in my opinion*". Overall, the human participants' answers were diverse because they varied in length, content and style. Additionally, the style was often matter-oriented and sometimes quite bold: "*No; the work I do is mainly on the computer…*".

The responses from different AI persons were quite similar with each other and did not show much diversity. The results from the AI tool were presented in a same formal way (see Figure 1). It highlighted the main things with bold text and then elaborated them in more detail. The artificial intelligence tool responded to the questions using two formats: third-person format (e.g., "*Dr. Andersson can actively participate in discussions, attend meetings, and engage with colleagues as if he were physically there*") and the first-person form (e.g., "*The tablet screen facilitates video chats, making communication seamless. I'd feel more connected to my team*"). When listing benefits, the AI tool's answers were enthusiastic and opportunity-focused: "*Having a mobile robot that extends my presence to the workplace while I'm remote would be invaluable*". It also used a descriptive and vivid style: "*imagine remotely navigating our bustling office*".

Benefit in work:

- **Remote Presence**: As she works from home, the robot becomes her eyes and ears in the office. She can virtually attend meetings, observe interactions, and engage with colleagues.
- **Object Handling**: The robotic arm assists her in organizing research materials, retrieving books, and even setting up experiments. It's like having an extra pair of capable hands.
- **Collaboration:** Video chats via the tablet screen foster collaboration. She can discuss findings, brainstorm ideas, and maintain a sense of connection.

Figure 1: Example of the AI tool's formulation of results when describing a robot concept's benefits in work.

DISCUSSION

Artificial intelligence and human groups had contradictory answers when identifying the benefits of the system. The findings indicated that the human participants did not perceive much value in employing the telepresence robot system in their work, since their work was primarily done on a PC and existing online meeting systems were well established. In contrast, the AI participant group had a more positive experience of the concept, with all of them seeing potential benefits in using the telepresence robot system in their work. This implies that the AI tool had difficulties in understanding the context of the participants' work and therefore, provided contradictory results compared to the human responses. It seems that even some background information, such as work location (office, lab) and job content, was provided to the AI tool it did not use the information effectively in its answers. These results show that concerns about reliability of AI-generated data and hallucination raised by Hess et al. (2024) are valid.

Both groups pointed out similar challenges, such as navigation, privacy, safety and security, connectivity, acceptance, cost, etc. However, many of the human participants only mentioned a single challenge in their responses. Instead, the AI participants listed three to seven challenges each. One possible reason is that the human participants gave the first answer that came to their mind and did not think deeply about them. Another factor that may have influenced the responses is that the human participants did not see the value of the concept for their work and thus did not put effort to the subsequent questions.

The number of suggested improvement ideas was higher within the AI group. It was challenging for the human participants to come up with many ideas for improvement, as they did not perceive how the concept could be useful in their own work. On the other hand, the AI group suggested many improvements, such as customizable appearance, a more intuitive interface, and safety features. It can be because the AI tool did not have preconceptions about the idea. This implies that generative AI tools could be useful, for example, in early design phase to identify possible challenges and improvement ideas. This would support current HCD methods. As Schmidt et al. (2024) argue, generative AI will not substitute user studies, but rather will complement the existing methods and tools of HCD researchers and practitioners.

Human participants expressed their feelings more than AI participants. The human participants had more experience-based comments, such as the robot does not provide value in their own work or using it can cause frustration. The artificial intelligence tool's answers were more general. This may have been due to AI tool using occasionally third-person form and not using many adjectives describing its own feelings (e.g., in my opinion, I think, I feel).

Human responses were more personal than AI participants' responses. For instance, some people used concise language, some expressed their views straightforwardly, and some were very polite and constructive. Instead, AI answers were all constructed similarly (see Figure 1). Despite having some

variations, the personalities of the respondents were not apparent within the AI tool. Even those answers which had the first-person format went towards general opinions rather than person's own experience.

The artificial intelligence tool's formulation of the results may ease data analysis. Neither of the groups was told how to answer the questions. However, the style AI group adopted was beneficial for the data analysis phase. The formulation of the answers made it easier to identify, for example, the main challenges and count how many times they were mentioned. User experience designers agree that generative AI has the potential to enhance efficiency in their work (Li et al., 2024).

This study was experimental in nature. The study was conducted from the HCD researchers' perspective and how easy it would be for them to utilise generative AI tools as test participants. This means that the prompting techniques and LLMs were not optimized for their best possible performance during the research. To improve the AI tool's responses, especially regarding the knowledge of experience and context, more extensive background material and better prompting techniques should be provided to make the AI tool better represent a certain type of user personality. It should also be noted that generative AI tools are constantly evolving, and this study was conducted with the version available at the time.

CONCLUSION

This paper presents one of the first studies on how to use generative artificial intelligence (AI) tools in human-centred design (HCD). The goal of this study was to find out if AI tools could be used in a qualitative evaluation to provide realistic results. Twenty-three human participants evaluated a robot concept by answering three open-ended questions (benefits, challenges, improvements) online in Spring 2024. The same setup was replicated with large a language model (LLM). Results were analysed qualitatively.

This study shows that there are possibilities and shortcomings when using AI tools in HCD. Artificial intelligence tools can be beneficial in identifying and listing many aspects related to the topics such as challenges and improvement ideas. Another benefit was the structured way it presented results which could make data analysis more efficient. However, there were still some notable shortcomings. The AI tool struggled especially in understanding the context of use and personality characteristics of the users. Therefore, the results for identifying benefits of this robot system were incorrect if compared to the responses of real human participants.

To fully replicate human cognitive processes, AI tools would require more fine tuning and improved and iterative prompt techniques. This way, they could better understand the context of use and the individual differences in personality, and how they influence the responses of the participants.

These findings highlight the differences between the groups of human and AI participants and the potential for using AI tools in HCD research. The findings can be used by researchers and technology developers when considering the use of AI tools in HCD.

ACKNOWLEDGMENT

This project has received funding from the VTT Technical Research Centre of Finland Ltd. The authors are grateful to all researchers and study participants who have contributed to and supported the work presented in this publication.

REFERENCES

Chui, M., Kamalnath, V. and McCarthy, B. (2020). An Executive's Guide to AI. McKinsey. Retrieved 19.1.2024 from Website: https://www.mckinsey.com/capabilities/quantumblack/our-insights/an-executives-guide-to-ai

Hess, A., Immich, T., Tamanini, J., Biedenbach, M. and Koch, M. (2024) "Opportunities and Limitations of AI in Human-Centered Design a Research Preview", In International Working Conference on Requirements Engineering: Foundation for Software Quality. Cham: Springer Nature Switzerland.

ISO 9241–210 (2019). Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems.

Li, J., Cao, H., Lin, L., Hou, Y., Zhu, R. and El Ali, A. (2024). "User experience design professionals' perceptions of generative artificial intelligence", proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.

McCrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of personality Volume 60 No. 2.

Murphy, K. P. (2022). Probabilistic machine learning: An introduction. MIT press.

Rogers, E. M., Singhal, A. and Quinlan, M. M. (2014). "Diffusion of innovations", in: An integrated approach to communication theory and research, Routledge. pp. 432–448.

Salminen, J., Jung, S.-G., Almerekhi, H., Cambria, E. and Jansen, B. (2023). "How Can Natural Language Processing and Generative AI Address Grand Challenges of Quantitative User Personas?", International Conference on Human-Computer Interaction. Cham: Springer Nature Switzerland.

Schmidt, A., Elagroudy, P., Draxler, F., Kreuter, F. and Welsch, R. (2024). Simulating the Human in HCD with ChatGPT: Redesigning Interaction Design with AI. Interactions Volume 31 No. 1.

Stige, Å., Zamani, E. D., Mikalef, P. and Zhu, Y. (2023). Artificial intelligence (AI) for user experience (UX) design: A systematic literature review and future research agenda. Information Technology & People.