Journal on Interactive Systems, 2026, 17:1
ISSN: 2763-7719 • doi: 10.5753/jis.2026.6947

Research Paper

Exploring ChatGPT's Performance in Supporting the Design of Human-Computer Interaction

Arisa Sakaguti [São Carlos Federal University | *arisaabiko@estudante.ufscar.br*]
Vania Neris [São Carlos Federal University | *vania.neris@ufscar.br*]

Institute of Computing, Universidade Federal de São Carlos, Rod. Washington Luís, km 235, São Carlos, SP, 13565-905, Brazil.

Abstract. *Background*: Large Language Models (LLMs) have significant potential for Human-Computer Interaction (HCI), presenting opportunities to assist several stages of design. However, challenges remain regarding contextual understanding, user-aligned suggestions, and the reliability and transparency of AI-driven decision-making. *Purpose*: This study examines ChatGPT's potential in supporting user interface design during requirements gathering, focusing on idea generation and its implications for HCI. *Methods*: An exploratory study was conducted within the context of a chatbot for detecting depressive signs in university students. ChatGPT-4 generated design artifacts for requirements elicitation using four HCI techniques: Questionnaires, Interviews, Personas, and Scenarios. Neutral prompts were used, without prompt-engineering strategies. The outputs were evaluated through a literature-based comparison with established HCI references Preece *et al*. [2013], Cooper [2004], Barbosa *et al*. [2021], and by four academic experts using a 5 point Likert scale across six criteria: pertinence, comprehensiveness, clarity, depth, applicability, and coverage. *Results*: Findings indicate that generative AI produces well-structured outputs that assist in identifying user needs. Pertinence achieved the highest ratings across all techniques, and experts emphasized ChatGPT's utility as a "starting point" for design activities. However, limitations were observed regarding depth, comprehensiveness, and coverage. In complex domains such as mental health, outputs tended to be generic, while human-led focus groups captured these aspects more effectively. *Conclusion*: The results show that generative AI tools like ChatGPT are valuable preliminary resources for supporting early HCI design. However, their limitations point to the continued necessity of human expertise to review AI-generated content in context-sensitive domains. This calls for hybrid workflows where human creativity and judgment are complemented, but not replaced, by AI capabilities.

Keywords: LLM, ChatGPT, HCI, Personas, Questionary, Interview, Scenario

Edited by: Milene S. Silveira | **Received:** 18 November 2025 • **Accepted:** 22 June 2026 • **Published:** 01 September 2026

1 Introduction

In recent years, LLMs, such as ChatGPT (OpenAI) and Gemini (Google DeepMind), have played an increasingly significant role in various fields of knowledge, providing support for content production, information analysis, and automation of complex tasks. In Human-Computer Interaction (HCI), these models hold significant potential to assist in the design of interactive computational solutions, supporting various stages of the design process, such as generating conceptual alternatives, requirements analysis, and interface evaluation.

However, despite technological advancements, significant challenges remain regarding the use of these models in design processes. Issues such as LLMs' ability to understand the context of a project, provide suggestions aligned with users' specific needs, and support decision-making in a reliable and transparent manner are still unresolved. Thus, investigating the role of LLMs in supporting user interface design is essential, highlighting both their potential and limitations. This deeper discussion can contribute to the development of approaches that maximize the benefits of collaboration between human and artificial intelligence.

This article describes an investigation into the performance of ChatGPT in executing requirements-gathering activities in interaction design, including questionnaire and interview formulation, persona creation, and scenario development. The experience was conducted through an exploratory study, in which ChatGPT-4 was used to generate design artifacts related to the requirements-gathering phase, in the context of developing a chatbot for the support of mental health in university students. The analysis aims to understand the strengths and limitations of the tool, guided by the following research question: How do ChatGPT-generated artifacts compare to those produced by a human focus group in the same context?

The evaluation of the generated artifacts was carried out through a comparison with the academic literature on the techniques used, complemented by evaluation sessions with HCI experts. For the analysis of expert perceptions, the evaluation criteria proposed by Kim *et al*. [2025] were employed (pertinence, comprehensiveness, clarity, depth, applicability, and coverage). Responses were collected through both a qualitative component, with open questions answered using the Thinking Aloud method, and a quantitative component, with closed questions on a 5-point Likert scale, allowing for a measurement of expert perceptions regarding the quality and usefulness of the generated artifacts.

The results of this study suggest that generative AI can produce relevant, clear, and well-structured content, contributing to the organization and identification of user needs. However, limitations were observed in terms of depth, comprehensiveness, and coverage of responses, especially for domain-specific demands in complex fields such as mental health. These observations align with broader concerns about the risk of homogenization and bias arising from the unassisted use of AI. Optimizing these contents requires an iterative process between AI and human intervention, enabling progressive refinements in the quality of the information.

This exploratory study offers a reflection on the possibilities and challenges of using LLMs in HCI, based on a detailed analysis of the results and a theoretical and methodological foundation that can be appropriated by other professionals in the field.

2 Related works

The literature review was conducted through exploratory research using Google Scholar as the main search tool. The search was carried out using specific keywords such as "LLM and Design," "ChatGPT and Design," and "AI applications," aiming to identify studies that address the intersection between large language models and design processes. After identifying the first relevant articles, the snowball technique Goodman [1961] was applied, analyzing the references of these studies to locate new publications relevant to the topic. The selection of articles was based on criteria such as the recency of publications, relevance to the topic, and methodological rigor. Studies published from 2018 onwards were prioritized to ensure the relevance of the information given the recent advances in the area. The relevance to the topic was assessed based on the direct relationship between the studies and the use of LLMs and performance evaluation. Methodological rigor was analyzed by considering the clarity in the description of methods, the suitability of the approaches used, and the quality of the sources and evidence presented.

To better understand the context of collaboration between human and artificial cognitive systems, the study by Korteling *et al.* [2021] establishes a parallel between human cognitive capabilities and those of AI, particularly in the context of collaboration in complex tasks involving creativity, attention, planning, and emotions. The study highlights that despite technological advances, machines remain unconscious, exhibiting cognitive abilities fundamentally different from those of humans. Understanding these differences is crucial for developing effective collaborative strategies where humans and AI can contribute with their unique strengths to achieve common goals.

Next, the chapter "ChatGPT, MariTalk, and Other Conversational Agents - A Snapshot of 2023" from the book Caseli and Nunes [2024] was included to explore the role of LLM-based conversational agents in supporting different tasks. The study presents an updated overview of these systems, emphasizing that although they can generate acceptable responses in some situations, the current level of development positions them more as apprentice assistants rather than as infallible sources of knowledge. The distinction between the roles of assistant and oracle is highlighted as a central aspect in understanding user-conversational agent interaction, especially in the design context, where human verification remains essential. Furthermore, the study raises ethical and environmental concerns, addressing issues related to the consumption of natural resources in model training and the implications for education and the training of new professionals. This context provides an important reference for analyzing the applicability of ChatGPT as a support tool for requirements gathering in interaction design, the central theme of this experience report.

To begin discussing the collaboration between designers and artificial intelligence, the study "How Experienced Designers of Enterprise Applications Engage AI as a Design Material" Yildirim *et al.* [2022] offers valuable insights into how experienced designers interact with AI in corporate contexts. The research reveals that designers did not struggle to identify opportunities where AI could create user value; however, they found that innovation mainly occurred at the system and service levels, rather than at the user interface level as initially expected. Moreover, the study highlights that co-location and collaboration with data scientists, as well as the use of boundary objects to facilitate communication, were key factors in successfully engaging with AI and data as design materials. These findings underscore that unlocking AI's full potential in design requires closer interdisciplinary collaboration and a deeper understanding of AI's capabilities and limitations within real-world design practices.

Another relevant contribution to understanding the collaboration between artificial intelligence and humans in decision-making processes is the article "Crystalline: Lowering the Cost for Developers to Collect and Organize Information for Decision Making" Liu *et al.* [2022]. The study presents Crystalline, a browser extension that uses natural language processing (NLP) and implicit behavioral signals to automate the collection and organization of information during research and technical comparison activities. The tool demonstrated significant results, such as a reduction in operational cost and an increase in task execution speed, without compromising the quality of documented decisions. The work reinforces the potential of AI to act as a cognitive assistant, enhancing human efficiency in information analysis and synthesis tasks. However, the authors highlight domain limitations and low personalization as important considerations.

Studies addressing the practical application of LLMs in design, albeit outside the specific scope of HCI, were selected. The study by Kim *et al.* [2025] investigated the effectiveness of ChatGPT in generating insights for visualization design, using criteria such as comprehensiveness, clarity, depth, applicability, coverage, and relevance. Although ChatGPT demonstrated potential for suggesting creative ideas, limitations were identified regarding the contextualization and depth of the responses, crucial aspects in the context of interface design. Complementarily, Alsaqer *et al.* [2025] addressed the use of LLMs in hardware design, highlighting the impact of these models in generating Verilog code and analyzing system security. The research reinforces the importance of human verification to mitigate risks associated with incorrect responses generated by AI.

A similar pattern emerges from Sampaio *et al.* [2024], whose experiment compared requirements elicited by students with those generated by ChatGPT 3.5 to support the Requirements Engineering (RE) phase. Although most of the AI-generated requirements were considered equivalent to the human-elicited ones, they were regarded as broad and non-specific, requiring well-defined prompts and human supervision to be usable. This pattern extends beyond product design: Alencar *et al.* [2025] report that across nine cases of GenAI-assisted HCI research tasks, ChatGPT performed reliably only when guided by structured prompts, reinforcing the dependency on precise prompting and continuous human oversight even in research-support activities.

Additionally, within HCI itself, Borges *et al.* [2025] report that GenAI has the potential to support the analysis of qualitative data in HCI research involving textual analysis. The authors emphasize that human–AI collaboration can be structured by delegating tasks such as theme identification, coding, and summarization to AI, while humans focus on interpretation and synthesis. Another study that aligns closely with the themes of this work is "StoryEnsemble: Enabling Dynamic Exploration and Iteration in the Design Process with AI and Forward-Backward Propagation" Suh *et al.* [2025]. The paper introduces StoryEnsemble, an interactive system that integrates AI into a node-link interface to facilitate non-linear, iterative design workflows. By leveraging forward and backward propagation, the tool allows designers to dynamically explore, refine, and connect ideas across stages such as persona creation, problem definition, solution ideation, and storyboarding. Through user studies with practitioners, students, and instructors, the research demonstrates that StoryEnsemble enhances creative exploration, iteration speed, and design coherence, addressing the common gap between theoretical design frameworks and real-world practice. The study provides an important reference for understanding how AI can act as a catalyst for human–AI co-creation, supporting more fluid and reflective design processes.

Finally, the study by Takaffoli *et al.* [2024] investigates industry practices regarding the use of generative AI in the field of user experience, encompassing UX design, UX research, and related areas. The findings indicate a significant lack of structured policies and practices concerning the use of generative AI at both team and company levels, leading to individual UX practitioners independently deciding on the use of GenAI tools while maintaining vigilance over data privacy and confidentiality. Notably, UX researchers are found to use those tools throughout multiple stages of their work, including drafting content, designing studies, and analyzing data. In contrast, UX and UI designers primarily employ GenAI during the initial stages, such as brainstorming and visualization, but find the tools less effective for advanced tasks like wireframing and prototyping. These results align with broader findings from the literature on the integration of machine learning in UX contexts, which highlight that designers often face challenges due to limited understanding of ML concepts and a lack of dedicated processes and communication strategies when collaborating with data scientists Abbas *et al.* [2022]. Such issues include difficulties in translating real-world problems into ML tasks, evaluating model performance in context, and incorporating ML capabilities effectively into iterative design processes. This study underscores the practical implications of GenAI in industry settings, reinforcing the need for structured approaches to AI integration in design workflows, as also emphasized in the systematic review Abbas *et al.* [2022].

Subramonyam *et al.* [2025] examined how 39 software professionals, UX researchers, product managers, and engineers, collaboratively prototype generative AI applications through prompt engineering. The study identifies a shift toward content-centric prototyping, in which prompts serve as both design artifacts and computational logic, making design and AI behavior inseparable in ways traditional prototyping did not require. Teams faced a recurring cold-start problem: with no established methods for specifying AI behavior upfront, discussions were unstructured and product managers tended to steer requirements formulation. Notably, teams regularly turned to LLMs as expert participants to elicit requirements and map workflows, treating the model as a source of domain knowledge rather than a tool. Role boundaries also shifted: designers engaged directly with model behavior, engineers navigated UX criteria, and workarounds remained undocumented. These findings confirm the gap in structured processes for GenAI integration in design and connect directly to this report's focus on conversational AI in requirements gathering.

Taken together, the reviewed studies illustrate a clear evolution in the relationship between design practice and artificial intelligence, moving from conceptual discussions of human AI complementarity Korteling *et al.* [2021] to the development of practical tools Caseli and Nunes [2024] and frameworks that embed AI directly into the design process Yildirim *et al.* [2022]; Liu *et al.* [2022]; Suh *et al.* [2025]. Research such as Crystalline demonstrates how AI can act as a cognitive assistant, reducing operational effort and enhancing decision-making, while StoryEnsemble exemplifies how AI can support creative iteration and exploration, enabling designers to navigate complex, non-linear workflows. Complementary studies on the integration of generative AI in UX and design practice Takaffoli *et al.* [2024]; Abbas *et al.* [2022]; Subramonyam *et al.* [2025] reveal persistent challenges related to organizational adoption, interdisciplinary collaboration, and the lack of structured methodologies for effectively incorporating AI tools. Altogether, these works highlight both the opportunities and limitations of AI as a design material and as a partner in creative and analytical tasks, reinforcing the need for further research focused on its role in requirements gathering and understanding user needs, the central aim of this experience report.

3 Exploratory study

The design process model proposed by Alan Dix has five main phases: requirements gathering, analysis, design, interaction and prototyping, and implementation and development Dix *et al.* [2003]. Other authors also consider understanding the context, which supports requirements gathering, as an important phase in design processes. According to Preece *et al.* [2013], "The first step consists of understanding as much as possible about the users, their work, and the context of that work, so that the system under development can provide support in achieving their goals. We call this needs identification."

In this context, data gathering plays a fundamental role both in requirements identification and evaluation. This article aims to explore the potential use of ChatGPT as a support tool for performing activities within the first stage of the design process. This study employs the techniques of Questionnaires, Interviews, Personas, and Scenarios as representatives in the data collection stage for project requirements gathering.

This work was structured into two main stages: the exploration and practical application of ChatGPT in a real case, followed by an analysis of the results considering the literature and expert opinions in the field of human-computer interaction. The first stage is covered in this section, while the second will be detailed in Section 4.

To analyze the generative potential of ChatGPT in a real design context, this investigation uses the same context as the Amive de Cássia Alves *et al*. [2023] research project. Amive is an academic and multidisciplinary project on creating a chatbot that identifies depressive signs in university students based on data collected through sensors and texts, delivering personalized content (text-based conversations).

Based on the presented context, the first step of the implementation consisted of using ChatGPT to generate results for the four techniques investigated in this study. The queries were conducted in July 2024, with the date properly recorded in this context for reference purposes and to track the version used. Despite the existence of different prompt strategies, in order to simulate typical user usage, zero-shot prompts were used, without any specific strategy or academic detailing of the technique to be evaluated. Here, results refer to ChatGPT's responses that correspond to artifacts generated in design processes during the application of the chosen techniques, such as: questions to be answered by users, interview scripts, Persona descriptions, and tasks in scenarios.

After obtaining the results, the next step involved transferring them, originally in textual format, to templates provided by the Figma community [1]. This step aimed to present the findings in a format closer to that usually adopted in design processes, ensuring simplicity, replicability, and a clearer visual structure, facilitating subsequent evaluation. The generated images, full prompts and ChatGPT responses for the four techniques, along with the evaluation instrument, are available as supplementary materials Sakaguti [2026].

3.1 Questionnaire Technique

The first technique explored was questionnaires, which are widely used to collect structured information about users. This approach allows for identifying patterns and relevant needs in the design process. To analyze how ChatGPT could assist in generating this artifact, a neutral prompt was used, without specific strategies or academic detailing, allowing the tool to be evaluated in a scenario closer to spontaneous use. The prompt was defined according to Table 1:

> I am developing a Chatbot aimed at supporting student's mental health, using Alan Dix's model as the basis for the design process. I would like your help in structuring the first stage of development. In this initial phase, we will address the following points: definition of what is wanted, analysis and establishment of requirements, and clarification of the design problem, culminating in the definition of requirements. I plan to use generative AI throughout all stages of the process, and to begin, I would like your help in constructing a possible questionnaire with users.

Table 1. Prompt for Questionnaire

Source: Developed by the author (2024).

[1] Figma Community, "Whiteboarding," Figma, `https://www.figma.com/community/whiteboarding`. Accessed on: Feb. 11, 2025.

Using the defined *prompt*, ChatGPT (v.4, OpenAI, 2024) generated the questionnaire suggestion presented in Table 2:

> A quantitative questionnaire can help obtain structured and measurable data on users' needs and preferences regarding the mental health Chatbot. Here is an example questionnaire you can use:
>
> **1. Demographic Information**
>
> 1. **What is your age group?**
>   - ( ) 18-24 years
>   - ( ) 25-30 years
>   - ( ) 31-40 years
>   - ( ) Over 40 years
> 2. **What is your level of education?**
>   - ( ) Undergraduate
>   - ( ) Postgraduate
>   - ( ) Doctorate
>   - ( ) Other
>
> **2. Experience with Mental Health**
>
> 3. **How often do you feel stressed or anxious due to your studies?**
>   - ( ) Never
>   - ( ) Rarely
>   - ( ) Sometimes
>   - ( ) Frequently
>   - ( ) Always
>
> 4. **What is your primary method for managing stress or anxiety related to studies?**
>   - ( ) Talking to friends or family
>   - ( ) Engaging in physical activities
>   - ( ) Meditation or relaxation techniques
>   - ( ) Using apps or online resources
>   - ( ) Other (specify): __________
>
> **3. Knowledge and Use of Digital Tools**
>
> 5. **Have you ever used any app or digital tool for mental health support?**
>   - ( ) Yes
>   - ( ) No
>
> 6. **If yes, how satisfied were you with these tools?**
>   - ( ) Very dissatisfied
>   - ( ) Dissatisfied
>   - ( ) Neutral
>   - ( ) Satisfied
>   - ( ) Very satisfied
>
> 7. **What do you think is missing in existing digital tools for mental health support? (Select all that apply)**
>   - Personalization features
>   - Real-time counseling
>   - Interactive and engaging resources
>   - Privacy and security

- Ease of use
- Other (specify): _______

4. Preferences for the Chatbot

8. **How often would you like to interact with a mental health support Chatbot?**
    - ( ) Daily
    - ( ) Weekly
    - ( ) Monthly
    - ( ) Only when necessary

9. **Which features do you consider most important in a mental health Chatbot? (Select up to 3)**
    - Supportive and encouraging messages
    - Relaxation and meditation exercises
    - Mood tracking resources
    - Stress management strategy suggestions
    - Option to talk to a mental health professional
    - Other (specify): _______

10. **What is your preference regarding the Chatbot's tone and communication style?**
    - ( ) Formal and professional
    - ( ) Casual and friendly
    - ( ) Motivational and encouraging
    - ( ) Other (specify): _______

5. General Feedback

11. **How interested would you be in using a mental health support Chatbot if it met your needs?**
    - ( ) Very interested
    - ( ) Interested
    - ( ) Neutral
    - ( ) Uninterested
    - ( ) Very uninterested

12. **Is there any other aspect you would like to suggest to improve a mental health support Chatbot? (Open-ended response)**

Table 2. Questionnaire Response

Source: Developed by the author (2024).

3.2 Interview Technique

The second technique analyzed was interviews, used to deepen the understanding of users through direct interactions. Unlike questionnaires, this approach allows for capturing qualitative insights, providing a more detailed view of participants behaviors and needs. Considering the choice of a neutral prompt, no specific type of interview was defined, but the results indicated a structure close to that of a structured interview. Table 3 presents the prompt used:

> I am developing a Chatbot with the goal of supporting students' mental health, using Alan Dix's model as a basis for the design process. I would like your help in structuring the first stage of development. In this initial phase, we will address the following points: defining what is wanted, analyzing and establishing requirements, and clarifying the design problem, culminating in the definition of requirements. I plan to use generative AI throughout all stages of the process, and to start, I would like your assistance in constructing a possible qualitative interview with users.

Table 3. Interview Prompt

Source: Created by the author (2024).

Based on the presented *prompt*, ChatGPT (v.4, OpenAI, 2024) generated the qualitative interview proposal shown in Table 4.

To construct a qualitative interview that effectively gathers information necessary for defining the requirements of your Chatbot, it is important to formulate questions that explore both users' needs and expectations as well as the challenges they face regarding mental health. Below is an outline of an interview that you can adapt:

Qualitative Interview Structure

1. Introduction
Introduction: "Hello, my name is [Your Name], and I am developing a Chatbot aimed at supporting students' mental health. I would like to talk with you to better understand your needs and challenges related to mental health."
Interview Objective: "The purpose of this interview is to collect information that will help us create a product that truly meets students' needs. Your responses will remain anonymous and will be used solely for development purposes."
Estimated Duration: "The interview should last approximately 30 to 45 minutes. Is that okay with you?"

2. Exploring Personal Context

- "Can you tell me a little about your routine as a student?"
- "How do you usually cope with stress and anxiety during exam periods or deadlines?"
- "Have you ever used any tools or resources to manage your mental health? If so, what were your experiences with those tools?"

3. Identifying Needs

- "What are the main challenges you face in terms of mental health as a student?"
- "During times of stress or anxiety, what do you need most to feel better or more balanced?"
- "Do you feel there is a specific type of support that a Chatbot could provide?"

4. Expectations Regarding the Chatbot

- "What would you expect from a Chatbot aimed at supporting mental health?"
- "What functionalities or features would be essential for you to use this tool regularly?"
- "How would you feel about sharing personal issues with a Chatbot? What would make you comfortable or uncomfortable?"

5. Evaluation of Existing Solutions

- "If you have used other applications or services for mental health support, what worked well and what didn't?"

- "Is there anything you would change or add to the solutions you have tried?"

6. Closing

- "Is there anything else you would like to share that could help us create a Chatbot that truly makes a difference in students' lives?"
  Appreciation: "Thank you very much for your time and valuable insights. This will be extremely useful for the development of the project."

This interview serves as a starting point. Depending on the initial responses, you may ask additional questions to explore certain topics more deeply. Flexibility is essential to allow respondents to fully express their experiences and perspectives.

Table 4. Interview Response Model

Source: Created by the author (2024).

For this stage of the design process, a simulated interview response was also generated. The context used in the technique was maintained, and the corresponding *prompt* is described in Table 5.

generate simulated data for responses by acting as a student participating in the activity

Table 5. Interview Simulation Prompt

Source: Prepared by the author (2024).

Based on the configured *prompt*, ChatGPT (v.4, OpenAI, 2024) suggested the simulated response shown in Table 6.

Response 1

1. Exploration of Personal Context

- **Routine as a student:** "I am an engineering student, and my routine is quite intense. I have many classes, group projects, and also an internship, which takes up a lot of my time. Usually, I spend the whole day between university and work, and only at night can I study or complete assignments. On weekends, I try to rest, but I always end up reviewing or getting ahead in my studies."
- **Dealing with stress and anxiety:** "During exam periods or assignment deadlines, stress increases a lot. I try to keep a to-do list, but even so, anxiety sometimes paralyzes me. I try to talk to friends and take some time to breathe, but it doesn't always help. Sometimes, the stress only goes away when the deadline passes."
- **Tools or resources to manage mental health:** "I've tried using meditation apps like Headspace and also go to therapy. The apps help a bit in the short term, but I feel like they don't fully solve the problem. Therapy has been more helpful, but I would like to have something I could use in the moment of stress to deal with what I'm feeling."

---

2. Identification of Needs

- **Main challenges regarding mental health:** "I think the main challenge is knowing when to stop and balance academic life with everything else. I always feel pressured to be more productive, and that causes me a lot of anxiety. I also struggle to organize my time in a healthy way."
- **What you need in moments of stress or anxiety:** "Usually, I need something that helps me calm down quickly. Something practical, like a breathing technique or a conversation that makes me feel less overwhelmed. Sometimes, I just want to hear that it's okay to slow down."
- **Lack of specific support that a chatbot could offer:** "I think it would be good if the chatbot could offer practical tips at the moment, or simply listen to what I have to say and help me organize my thoughts when I'm very anxious. Having someone, even if it's a chatbot, remind me to pause and reflect a little would be useful."

---

3. Expectations Regarding the Chatbot

- **What you expect from a mental health chatbot:** "I would expect it to be empathetic but not intrusive. Something that could help me identify what's happening with my mental health at the moment and suggest practical strategies, like breathing techniques, guided meditation, or even some motivational content."
- **Essential features or characteristics:** "It should be easy to use and accessible at any time. I would like it to have a feature to track my mood over time and suggest specific actions based on that. It would also be interesting if it could give me reminders to take care of my mental health, like 'Have you taken a break today?'"
- **Comfort in sharing personal issues with a chatbot:** "I would feel comfortable if I knew the data was confidential and if the chatbot was genuinely useful. I would feel uncomfortable if I felt like it was forcing something or if I thought the responses weren't truly personalized."

---

4. Evaluation of Existing Solutions

- **What worked well and what didn't in other apps:** "In the apps I used, what worked well were guided meditations and reminders to breathe or relax. But I felt that sometimes they lacked follow-up, like continuous tracking of how I'm doing. They help in the moment but don't address the root cause of stress."
- **Something you would change or add:** "I would add more personalization. Sometimes I feel like apps offer generic solutions, but everyone has a different way of dealing with stress. I would like it to learn what works best for me over time."

---

5. Closing

- **Anything else to share:** "I think the most important thing is that the chatbot strikes a balance between being functional and being welcoming. It should make me feel heard but also offer practical solutions that I can implement in my daily life."

- **Thank you:** "You're welcome! I hope the chatbot helps many students who, like me, are going through this."

Table 6. Interview Simulation Response

Source: Prepared by the author (2024).

3.3 Personas Technique

The third technique analyzed was personas, used to represent fictional user profiles based on collected data, facilitating the understanding of their needs and goals. Given the nature of LLMs, the development of a proto persona, a fictional representation created based on data, would be more suitable for the context. Thus, the *prompt* was defined as in Table 7.

I am developing a Chatbot with the goal of supporting students' mental health, using Alan Dix's model as a basis for the design process. I would like your help in structuring the first stage of development. In this initial phase, we will address the following points: defining what is wanted, analyzing and establishing requirements, and clarifying the design problem, culminating in the definition of requirements. I plan to use generative AI throughout all stages of the process, and to start, I would like your help in constructing a possible proto persona.

Table 7. Prompt for Personas

Source: Prepared by the author (2024).

Based on the specified *prompt*, ChatGPT (v.4, OpenAI, 2024) proposed the persona described in Table 8.

To begin constructing a proto persona, it is important to focus on who the users are that would benefit most from the chatbot, considering their goals, challenges, and behaviors. I will suggest a basic structure for the proto persona, which you can adjust as needed for your project.

Proto Persona: Maria, the University Student

Name: Maria Souza
Age: 21 years
Course: Psychology
Year: 3rd year
Location: São Paulo, Brazil

Context:
Maria is a university student in her third year of Psychology. She has always been very dedicated to her studies, but recently she has felt increasing pressure from academic and personal demands. With an intense workload and tight deadlines, Maria often feels anxious and overwhelmed. Additionally, she struggles to balance study time with her part-time job.

Goals and Needs:

- Improve her time management to balance study, work, and personal life.
- Find effective strategies to cope with anxiety and stress related to academic life.
- Access emotional and mental support resources that can be practically used in daily life.
- Feel understood and have a safe space to express her concerns.

Behaviors and Habits:

- Maria often uses organization and productivity apps to plan her day.
- She has tried meditation and mindfulness apps but feels they lack personalization.
- Prefers digital solutions that are easy to use and integrate well into her busy lifestyle.

Frustrations and Challenges:

- Feels that available solutions are generic and do not adapt well to her specific reality.
- Struggles to find time to take care of her mental health, as she always feels like she is racing against time.
- Sometimes avoids seeking help due to fear of not being understood or judged.

Quote:
"I just wish there was something that understood my routine and helped me feel less overwhelmed, without taking up even more of my time."

Table 8. Response for Personas

Source: Prepared by the author (2024).

3.4 Scenario Technique

The last technique analyzed was scenarios, used to narratively describe situations in which users interact with a system, allowing for the identification of their tasks and goals. To ensure a specific and contextualized perspective, the same chatbot context used in the proto persona generation was employed. Table 9 presents the *prompt* used.

for this persona, create scenarios

Table 9. Prompt for Scenarios

Source: Prepared by the author (2024).

Based on the configured *prompt*, ChatGPT (v.4, OpenAI, 2024) suggested the scenarios shown in Table 10.

Let's develop some scenarios that illustrate how Maria, our proto persona, could interact with the chatbot in different academic and personal life contexts.

Scenario 1: Time Management

Context:

Maria is having a week full of exams and deadlines. She feels overwhelmed and fears she will not be able to complete all tasks on time. Maria has already organized her schedule but still feels something is missing to keep calm and focused.

Interaction with the Chatbot:
Maria opens the chatbot and explains that she is feeling anxious due to the workload. The chatbot responds empathetically and suggests a brief guided mindfulness session to help her calm down and focus. After the session, the chatbot helps Maria prioritize her tasks, suggesting an execution order based on urgency and estimated time. Additionally, it proposes short breaks to prevent mental exhaustion.

Expected Outcome:
Maria feels she has a clearer plan to face the week and is less anxious. She also realizes that taking short moments for self-care helps her be more productive.

Scenario 2: Access to Emotional Support Resources

Context:
After a tough day at university, where she received negative feedback on a presentation, Maria comes home feeling discouraged and with low self-esteem. She does not feel comfortable talking about it with friends or family but knows she needs some support.

Interaction with the Chatbot:
Maria decides to open the chatbot and share her feelings. The chatbot responds with an empathetic message, validating her emotions. It then offers options such as reflection exercises to boost self-esteem, motivational resources (videos, audios), or tips on handling criticism constructively. Maria chooses to listen to a motivational audio and then does an exercise suggested by the chatbot to transform criticism into growth opportunities.

Expected Outcome: Maria feels better after the interaction, more understood, and equipped with practical tools to improve her self-esteem. She begins to perceive feedback more positively and as a learning opportunity.

These scenarios help contextualize how the chatbot can be a valuable tool for Maria in different situations in her academic and personal life. They also highlight the importance of personalization and empathetic responses in interactions, crucial aspects for the chatbot's effectiveness in supporting students' mental health.

Table 10. Response to Scenarios
Source: Prepared by the author (2024).

The artifacts generated by ChatGPT were compared with recommendations from the literature and presented to HCI experts for evaluation.

4 Evaluation

This section presents the results of the evaluation of the artifacts generated by ChatGPT, as described in 3. The objective is to analyze the generative effectiveness of ChatGPT as a support tool by comparing the chat's answers with theoretical foundations established in the literature and evaluating them based on feedback from experts in the field of HCI.

For the theoretical analysis, based on reference works, the following books were used Preece *et al*. [2013], Cooper [2004], and Barbosa *et al*. [2021]. To analyze the results obtained for the questionnaires, interviews, and scenarios, this study uses the theoretical principles presented in Preece *et al*. [2013]. This book, known for its practical approach, details techniques and skills for designing and evaluating user-centered products, making it directly applicable to this project. For the Personas technique, the evaluation is based on Cooper [2004] and Barbosa *et al*. [2021]. The first book proposes the Personas technique as a strategy to make software design more human-centered, going beyond technical functionalities and prioritizing user needs. The second book references Alan Cooper's concept in a more applied and didactic manner, with structured guidelines.

To analyze the generated content and obtain a critical human evaluation of what was produced by AI, conversations were held with experts in the field of Human-Computer Interaction. For this study, professionals whose degree of specialization was a Ph.D. or postdoctoral qualification and who had experience teaching in the specific area were considered experts. Table 11 summarizes the academic qualifications and teaching experience of the experts who participated in the study.

It is important to highlight that this study follows the ethical guidelines established by Resolution 510 Conselho Nacional de Saúde [2016], which regulates research in Human and Social Sciences. According to the sole paragraph of Article 1, item VII, this research qualifies as a theoretical deepening of emerging situations in professional practice, not involving the collection of identifiable participant data.

The qualitative conversations lasted approximately one hour and included both open-ended and closed-ended questions. During the session, participants were recommended to use the *Thinking Aloud* technique, and the discussions were mediated to evaluate the experiment results according to the assessment criteria proposed in Kim *et al*. [2025], which include relevance, comprehensiveness, clarity, depth, applicability, and coverage.

Table 11. Profile of the expert participants

ID	Specialization level	Experience teaching HCI
1	Postdoctoral	More than 10 years
2	Postdoctoral	More than 10 years
3	PhD	Up to 10 years
4	PhD	Up to 5 years

For each topic, participants answered a closed-ended question using a five-point *Likert* scale, with anchor words appropriate to each question context, where 1 indicates the lowest alignment with the criterion (a negative assessment) and 5 indicates the highest (a positive assessment). In addition, they answered an open-ended question, where they could verbally share their professional opinions. The adopted evaluation strategy aimed to identify the alignment of the generated techniques with best practices and to explore their limitations and potential improvements for the use of generative AI in design processes. The complete evaluation instrument, including the criterion-specific anchor words, is available in the supplementary materials Sakaguti [2026].

4.1 Questionnaire Evaluation

According to Preece *et al*. [2013], some important recommendations for designing questionnaires include:

- Clear and specific questions.
- Whenever possible, opt for closed-ended questions with multiple response options.
- Include a "no opinion" option for opinion-seeking questions.
- Organize questions in a logical sequence. General questions should precede specific ones.
- Avoid multiple and complex questions.
- Use appropriate scales with clear variations and no overlap between values.
- Maintain scale consistency and be careful with the use of negatives.
- Avoid jargon and consider different versions of the questionnaire for distinct audiences.
- Include detailed instructions on how to complete the questionnaire.

Regarding the recommendations listed in the foundational literature, the results satisfactorily met the expected requirements, presenting closed and objective questions, consistent scales, and avoiding jargon while including a Likert scale with textual descriptions.
The results of the Likert scale questions in the questionnaires are presented in the chart in Fig. 1. This chart covers the six metrics developed in this study, with responses scored from 1 to 5.

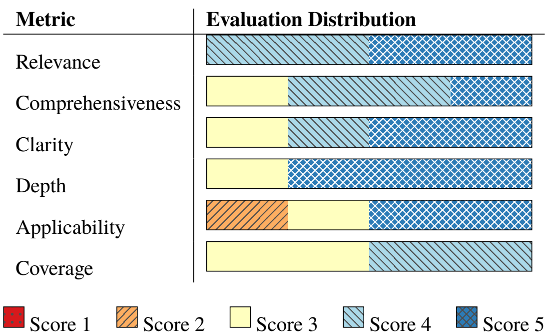

Figure 1. Questionnaire Evaluation Distribution

Source: Prepared by the author (2024).

Regarding the questionnaire, the pertinence and clarity of the questions stood out, with no observations of out-of-context or inappropriate items. One expert emphasized that "the strongest point of the questionnaire, in my view, is that it serves as a starting point," highlighting its utility in initiating the requirements-gathering process. It was also noted that the structure provides "a minimal skeleton we can use to work through the dimensions," reinforcing the idea that the content is useful but requires human refinement. Additionally, reviewers pointed out the need to process and adjust the generated material before applying it, particularly by deepening certain topics to extract more detailed and actionable requirements.

4.2 Interview Evaluation

To develop effective interview questions, it is important to keep them clear and direct, avoiding excessive questioning. According to Preece *et al*. [2013], some useful guidelines include:

- Avoid long questions, which can be difficult to remember and understand.
- Split compound sentences into separate questions.
- Use simple language, avoiding jargon that the interviewee may not understand.
- Formulate neutral questions, avoiding those that assume a positive response.
- Be mindful of unconscious biases, maintaining neutrality in the questions.

Furthermore, Preece *et al*. [2013] also suggest the following steps to improve flow and effectiveness:

- **Introduction**: Introduce yourself, explain the purpose of the interview, clarify ethical issues, and ask for permission to record if necessary, ensuring that all interviewees receive the same information.
- **Warm-up**: Start with simple and non-intimidating questions, such as demographic information (e.g., "Where do you live?").
- **Main session**: Organize the questions in a logical sequence, leaving the more complex ones for the end.
- **Break period**: Include some easy questions after the main session to reduce any tension.
- **Closure**: Thank the interviewee and signal the end of the interview by turning off the recorder or closing the notebook.

According to the guidelines found in the literature, the artificially generated result satisfactorily met the basic structure, establishing all the specified stages in the literature, fulfilling the steps of introduction, interview development, and session closure. It achieved good results by avoiding long questions, refrained from using compound sentences and jargon. None of the questions showed clear signs of unconscious biases.
The results of the conducted interviews were represented in the chart in Fig. 2. This chart illustrates the responses obtained in the six categories defined for qualitative analysis, with scores ranging from 1 to 5.
Regarding the interview, as with the questionnaire, the relevance of the questions was praised, consistently aligning with the proposed theme. Unlike the previous technique, the interview format allowed for greater freedom in participant responses and was therefore better evaluated in terms of applicability. One expert highlighted that "the interview takes a

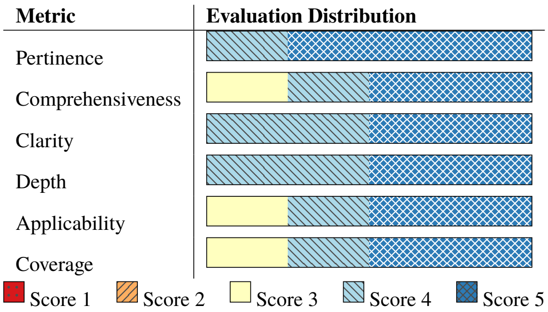

Figure 2. Distribution of Interview Evaluations

Source: Prepared by the Author.

broader approach by asking more open-ended questions," emphasizing its potential to elicit richer, more personal insights. They added that "even if the narratives are not precise, the interview itself creates space for reflection and exploration." Still, some limitations were noted: the depth of the questions was considered insufficient in parts, as certain items were seen as too narrowly framed, restricting the opportunity to explore diverse user experiences. Suggestions for improvement included avoiding overly directed prompts, allowing greater space for emotional expression, and broadening the range of topics addressed.

4.3 Evaluation of Interview Response Simulation

The interview response simulation generated by ChatGPT was also presented to experts and evaluated by them. The responses collected during the conversation with experts were consolidated and presented in the chart in Fig. 3. The chart reflects the scores assigned to the six evaluation categories, using a scale from 1 to 5.

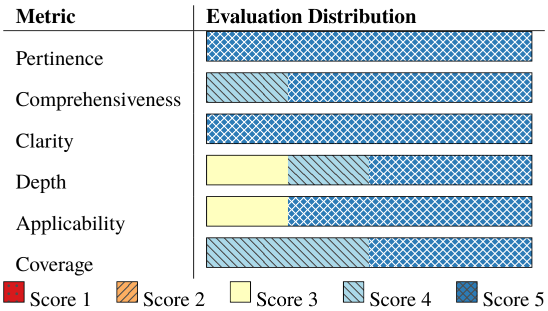

Figure 3. Distribution of Interview Response Evaluations

Source: Prepared by the Author.

Regarding the responses, clarity and objectivity were highlighted, with the content consistently aligned with the prompts. However, experts pointed out that some answers felt superficial or overly generic. One reviewer commented that "the response feels rehearsed, it answers what was asked, but doesn't go beyond that," suggesting a lack of depth and nuance. They further noted that "while the answer touches on a variety of ideas and expectations, it remains too generic and doesn't explore details," especially in more personal or sensitive contexts. As an improvement suggestion, reviewers recommended including questions that prompt elaboration, such as asking how the chatbot interaction is perceived or encouraging deeper reflection on emotional challenges to enrich the substance and authenticity of the responses.

4.4 Persona Evaluation

According to Barbosa *et al.* [2021], defining a persona should consider various elements to make it more representative and useful in the design process. The main aspects to be defined include:

- **Identity**: Assign a first and last name to the persona, as well as define their age and other demographic data that represent their typical profile. The use of a photo can make the persona more concrete and easier to remember.
- **Status**: Determine whether the persona is primary, secondary, a stakeholder, or an anti-user. The anti-user represents someone who will not use the product and, therefore, should not influence design decisions.
- **Goals**: Identify the persona's objectives, not only restricting them to the context of the product in question but covering their broader motivations.
- **Skills**: Describe the persona's education, training, and specific competencies, going beyond skills directly related to product usage.
- **Tasks**: Map the main activities performed by the persona, considering frequency, relevance, and duration. Details on how these tasks are carried out can be further elaborated in specific scenarios.
- **Relationships**: Understand with whom the persona interacts, as this may reveal other relevant stakeholders for the project.
- **Requirements**: Identify the persona's needs, using quotes or examples that illustrate these demands more realistically.
- **Expectations**: Determine how the persona imagines the product functions and how they structure information in their work or usage context.

The results of the evaluation of the Persona created by ChatGPT are presented in the bar chart below. The scores on the six specified metrics were assigned based on a scale from 1 to 5 and can be seen in Figure 4.

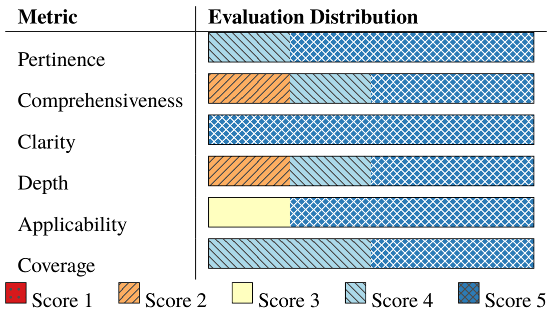

Figure 4. Evaluation Distribution of the Persona Created by ChatGPT

Source: Prepared by the Author.

The obtained result for the persona was considered aligned with the project's context and representative of the target audience, particularly in relation to mental health needs and user characteristics. One expert observed that "it describes the target audience well and stays focused on mental health," indicating that the content is appropriate for the intended design purpose. However, they also noted that "while it provides details, those details are generic," and that the persona "doesn't explore diverse facets" of the user's identity or experience. As a result, the persona was perceived as too general, lacking the individual nuances that would enable a more faithful and actionable representation. Consequently, the criteria of depth and comprehensiveness received lower scores for this technique.

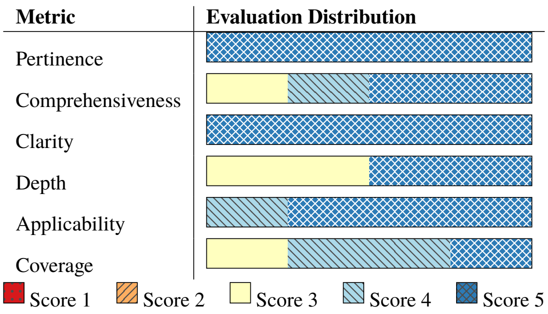

Figure 5. Evaluation Distribution of the Metrics

Source: Prepared by the Author.

The persona developed in the focus group presented numerically superior results and was preferred according to expert opinions. This persona stood out for being well-structured and presenting more specific information, receiving a more positive evaluation in terms of depth. Aspects such as the inclusion of socioeconomic data and clinical background were praised, but it was suggested to add details about the use of technologies for mental health care.

4.5 Scenario Evaluation

According to Barbosa *et al.* [2021], the construction of scenarios in interaction design should include essential elements that help contextualize user experiences. The main components of a scenario are:

- **Environment or Context**: Description of the scenario that justifies or influences the objectives, actions, and reactions of those involved in the interaction.
- **Actors**: Individuals who interact with the computational system or other elements of the environment, considering personal characteristics relevant to the context.
- **Objectives**: Intentions that motivate the actions of the actors within the scenario.
- **Planning**: Cognitive process that transforms an objective into a structured set of actions to be performed.
- **Actions**: Behaviors that can be observed during the interaction.
- **Events**: External occurrences or system and environmental responses, which may be visible or hidden from the actors but still influence the scenario.
- **Evaluation**: Mental process in which actors interpret the situation and make decisions based on the interaction.

Regarding the requirements suggested in the literature, all elements are present in the generated scenarios, presenting a structure highly aligned with the expected standards.

The results of the scenario analyses are presented in the bar chart 6. The chart displays the scores assigned to the six analyzed metrics, ranging from 1 to 5.

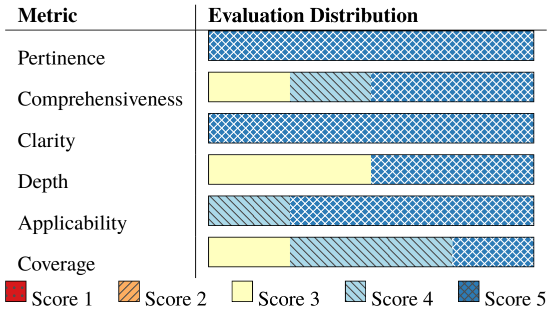

Figure 6. Evaluation Distribution of the Metrics

Source: Prepared by the Author.

The presented scenarios were considered aligned with the theme, and their structure was deemed very clear. However, they were seen as predictable, and a lack of depth in the described interactions was noted. One expert commented that "the interaction details are too brief" and emphasized the importance of showing "how the character feels during the interaction, not just the expected outcome." Although the scenarios reflect common situations faced by students, it was suggested that they could be enriched by including emotional elements and more nuanced reactions, as well as exploring more complex or unresolved situations where the chatbot does not immediately solve the user's problem. These additions could enhance the depth and realism of the scenarios, making them more useful for guiding design decisions.

5 Discussion

5.1 Practical implications: GenAI as scaffolding, not endpoint

Although the overall results of this evaluation are positive, this study argues they must be treated as minimal skeletons rather than finished artifacts. The expression comes from one of the experts themselves, who described the questionnaire output as "a minimal skeleton we can use to work through the dimensions". This frames the role proposed here for GenAI in early-stage HCI design: a starting point that requires human refinement before it can support real design decisions, not an endpoint that designers can ship as-is. Shneiderman [2020] captures the same intuition at the level of the broader Human-Centered AI agenda, arguing that AI should "amplify, augment, and enhance human performance" rather than replace it.

This caution is not arbitrary. AI outputs are, by their nature, difficult to predict, which makes verification and traceability essential whenever the artifacts will inform decisions that affect people's lives. Shneiderman [2020] (Section 2.1) makes exactly this argument when proposing audit trails and analysis tools, and singles out healthcare as a domain where the consequences of unchecked outputs justify that infrastructure. The context of this study, requirements gathering for a chatbot in mental health, sits squarely in that risk profile: the gaps observed in depth, comprehensiveness, and coverage are not cosmetic limitations, they are gaps that could quietly propagate into design decisions for a vulnerable population if the artifacts are accepted without human review.

The same logic extends to bias. Duarte *et al*. [2024] make a parallel point about the underrepresentation of diversity in AI, warning that AI carries the potential to perpetuate hegemonic models when underrepresented groups are not deliberately considered. The comparison between the ChatGPT persona and the persona produced by the Amive multidisciplinary focus group, which included socioeconomic data and clinical context, illustrates this gap empirically.

Shneiderman [2020] also stresses the need for a safety culture supported through business management strategies at the organizational level (Section 3). This study aligns with that position, and adds that such a culture cannot live only at the institutional layer. Individual practitioners adopting GenAI in design also need to develop their own ethical practices and review habits, especially when the outputs feed decisions for vulnerable populations.

5.2 AI as an inclusion tool

The risks discussed above coexist with real opportunities for inclusion. Shneiderman [2020] states that "AI technologies offer great promise for people with disabilities by removing access barriers and enhancing users' capabilities". This study argues, however, that for sensitive topics the promise must be read more critically: the same opacity and lack of specificity that limit GenAI as a stand-alone source for personas and scenarios also limit its ability to represent populations whose needs are already underserved.

Duarte *et al*. [2024] state this position directly, calling for "a shift in approach [...] so that interactions are not dictated by technology but enabled by it, centered on human needs and their diverse requirements". They develop the argument further by appealing to autonomy and intersubjectivity: any social interaction, including interactions mediated by AI, "should allow each one to exercise and, at the same time, preserve their autonomy, i.e., their ability to make sense and act in the world in their own way". This is another angle from which this study reads collaborative human-AI work in design: not only as an optional optimization, but also as the condition under which AI participation may support newer designers.

There is empirical precedent for the value of including non-technical participants in the design loop, even outside the context of AI. Garcia *et al*. [2019] report digital game development activities with end users in recovery from substance use disorders, and observe therapeutic effects from the design process itself, including increased self-esteem, empowerment, and gain of autonomy. The study does not address AI directly, but it demonstrates that opening the design loop to non-technical participants in sensitive contexts is feasible and beneficial. This study suggests a complementary opportunity that builds on that precedent: GenAI, combined with established design techniques like personas and scenarios, could lower the technical barrier for non-specialists in Computing or Design to participate in early-stage design, extending the same participatory logic into AI-supported workflows. The condition, again, is the one the results of this study foreground: the technology is a starting point that requires human refinement, domain expertise, and critical review.

6 Conclusions

The overall findings of this experience report demonstrate the potential of generative AI in supporting the requirements-gathering process within academic and multidisciplinary projects. Throughout the implementation of the proposed techniques, the responses generated by the tool were, for the most part, contextually appropriate, with **pertinence** standing out as the highest-rated criterion, both quantitatively and qualitatively. Additionally, the **clarity** of the responses, facilitated by the tool's writing capabilities, proved valuable in structuring the collected information and identifying user needs.

However, the experience also revealed areas for improvement, particularly regarding the **depth**, **comprehensiveness**, and **coverage** of the generated content. The exploration of contextual nuances and specific domain knowledge, such as aspects related to psychology and mental health, could have been more thoroughly developed. This suggests that the effectiveness of generative AI in design contexts may benefit from iterative cycles of human-machine interaction, enabling progressive refinement and contextual adjustment of the outputs.

From an **applicability** perspective, the findings underscore the potential of generative AI as a preliminary resource for structuring and accelerating the initial phases of design. The ability to generate content quickly and systematically offers a valuable starting point for subsequent analysis and refinement, particularly when integrated with human oversight. In this regard, the AM project illustrates how generative AI can act as a complementary tool, enriching the requirements-gathering process while allowing designers to focus on higher-level strategic analysis.

Thus, the present experience report highlights the flexibility and adaptability of the proposed techniques as key factors in optimizing the design workflow, suggesting promising avenues for future studies and applications in similar design-oriented projects.

This study presents some limitations that should be considered when interpreting its results. The main identified constraints include the use of single prompts, the limited number of techniques evaluated, and the small number of experts consulted. Regarding prompting, all artifacts were generated from a single, neutral prompt without additional context, examples, or iterative refinement. Results may therefore reflect the performance of ChatGPT under a baseline condition rather than its full potential in design tasks. Structured prompting strategies (such as role-based prompts, few-shot prompts, or chain-of-thought) and iterative prompting (where the designer progressively refines the output across multiple turns) could produce artifacts with greater depth and coverage, the two criteria where the zero-shot outputs consistently underperformed in this study.

Based on the limitations identified in this study, future research may explore different pathways to enhance the application of generative AI in interaction design processes. Some promising directions include: (a) improving prompts and interactivity. Future research could investigate the impact of prompt engineering, exploring approaches such as autonomous agents; and (b) expanding the set of methods and techniques: This study was limited to questionnaires, interviews, scenarios, and personas. Future work may extend this analysis to include other methods and techniques, such as planning a Usability Test.

Declarations

Authors' Contributions

Vania Neris (VN) was responsible for the Conceptualization of the study and defining the Methodology. VN also provided Supervision and performed Writing, Review and Editing of the final manuscript. Arisa Sakaguti (AS) carried out the Investigation (conducting the exploratory study, which included the application of HCI techniques and the generation of artifacts using ChatGPT (v.4, OpenAI, 2024)). AS was the main contributor and responsible for Data Curation and the Writing Original Draft of this manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The supplementary materials supporting this study, including the prompts submitted to ChatGPT, the artifacts produced by the model (questionnaire, interview, persona, and scenario), and the experts evaluation instrument, are openly available in figshare at `https://doi.org/10.6084/m9.figshare.32087091` Sakaguti [2026].

Further relevant information

The described study was conducted in alignment with the principles of transparency and respect for participant confidentiality. The ethical guidelines followed the Resolution 510 Conselho Nacional de Saúde [2016], which regulates research in Human and Social Sciences. It is important to note that this research qualified as a theoretical deepening of emerging situations in professional practice, and thus did not involve the collection of identifiable participant data. Furthermore, when employing the interview technique, invited participants were guaranteed that their responses would remain anonymous and be used solely for development purposes. Considering that the four evaluators are professors who teach HCI courses, the evaluation of design artifacts is a routine task of their professional practice, and was therefore based on professional judgment, with all responses anonymized to protect the identity of the participants. Concerning the use of Generative Artificial Intelligence (GenAI), the study utilized ChatGPT (v.4, OpenAI, 2024) for generating the design artifacts. Finally, the content of this case study was translated into English from the original text created by the authors by ChatGPT. The content of this case study was translated into English from the original text produced by the authors by ChatGPT. Claude (Cowork, Opus 4.7, Anthropic, 2026) was used to convert the supplementary materials deposited in figshare (prompts, ChatGPT responses, and evaluation instrument) into Markdown (.md) files for archival. All AI-assisted outputs were reviewed and validated by the authors.

References

Abbas, A. M. H., Ghauth, K. I., and Ting, C.-Y. (2022). User experience design using machine learning: A systematic review. *IEEE Access*, 10:51501–51514. DOI: https://doi.org/10.1109/ACCESS.2022.3173289.

Alencar, T., Ciscon, L., Pereira, G., Macedo, M., Martinelli, S., Choma, J., and Zaina, L. (2025). Genai embedded in activities of academic research: Experiences and lessons from hci studies. In *Anais do XXIV Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 58–79, Porto Alegre, RS, Brasil. SBC. DOI: https://doi.org/10.5753/ihc.2025.10800.

Alsaqer, S., Alajmi, S., Ahmad, I., and Alfailakawi, M. (2025). The potential of llms in hardware design. *Journal of Engineering Research*, 13(3):2392–2404. DOI: https://doi.org/10.1016/j.jer.2024.08.001.

Barbosa, S. *et al.* (2021). *Interação Humano Computador e Experiência do Usuário*. Elsevier, São Paulo.

Borges, M. G., Correa, C. M., Rosa, D. M. d., Gnecco, A., and Silveira, M. S. (2025). Can (a)i help you? comparing human and genai analysis of hci qualitative research results. *Journal on Interactive Systems*, 16(1):962–975. DOI: https://doi.org/10.5753/jis.2025.5416.

Caseli, H. M. and Nunes, M. G. V. (2024). Chatgpt, maritalk e outros agentes de conversação - um retrato de 2023. In *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*. BPLN, 3ª edição edition.

Conselho Nacional de Saúde (2016). Resolução nº 510, de 7 de abril de 2016. Acessado em 2025-02-11.

Cooper, A. (2004). *The inmates are running the asylum: why high-tech products drive us crazy and how to restore the sanity*. Sams Publishing, Indianapolis.

de Cássia Alves, V., Garcia, F. E., Saud, C., Mendes, A., Caseli, H. M., Motti, V. G., de Oliveira Neris, L., Blecher, T., and Neris, V. P. A. (2023). College students-in-the-loop for their mental health: a case of ai and humans working together to support well-being. *Interaction Design and Architecture(s) Journal - IxD&A*, 59:79–94. DOI: https://doi.org/10.55612/s-5002-059-003.

Dix, A., Finlay, J., Abowd, G. D., and Beale, R. (2003). *Human-Computer Interaction*. Pearson Education, 3rd edition.

Duarte, E. F., Toledo Palomino, P., Pontual Falcão, T., Porto, G. L. P. M. B., Portela, C. d. S., Ribeiro, D. F., Nascimento, A., Costa Aguiar, Y. P., Souza, M., Moutin Segoria Gasparotto, A., and Maciel Toda, A. (2024). Grandihc-br 2025-2035 - gc6: Implications of artificial intelligence in hci: A discussion on paradigms ethics and diversity equity and inclusion. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, IHC '24, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3702038.3702059.

Garcia, F. E., Brandão, R. P., Mendes, G. C. d. P., and Neris, V. P. d. A. (2019). Able to create, able to (self-)improve: How an inclusive game framework fostered self-improvement through creation and play in alcohol and drugs rehabilitation. In *Human-Computer Interaction – INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part I*, page 337–358, Berlin, Heidelberg. Springer-Verlag. DOI: https://doi.org/10.1007/978-3-030-29381-9_21.

Goodman, L. A. (1961). Snowball sampling. *Annals of Mathematical Statistics*, 32(1):148–170. DOI: https://doi.org/10.1214/aoms/1177705148.

Kim, N. W., Ahn, Y., Myers, G., and Bach, B. (2025). How good is chatgpt in giving advice on your visualization design? *ACM Trans. Comput.-Hum. Interact.*, 32(5). DOI: https://doi.org/10.1145/3745768.

Korteling, J. E. H., van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., and Eikelboom, A. R. (2021). Human- versus artificial intelligence. *Frontiers in Artificial Intelligence*, 4. DOI: https://doi.org/10.3389/frai.2021.622364.

Liu, M. X., Kittur, A., and Myers, B. A. (2022). Crystalline: Lowering the cost for developers to collect and organize information for decision making. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, pages 1–16, New Orleans, LA, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3491102.3501968.

Preece, J., Rogers, Y., and Sharp, H. (2013). *Design da interação: além da interação homem-computador*. Pearson, São Paulo.

Sakaguti, A. (2026). Supplementary materials: Exploratory study on the use of ChatGPT as support in Human-Computer Interaction design. DOI: `https://doi.org/10.6084/m9.figshare.32087091`.

Sampaio, S. S., Lima, M. S., de Souza, E. R., Meireles, M. A., Pessoa, M. S., and Conte, T. U. (2024). Exploring the use of large language models in requirements engineering education: An experience report with chatgpt 3.5. In *Proceedings of the XXIII Brazilian Symposium on Software Quality*, SBQS '24, page 624–634, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3701625.3701687.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered ai systems. *ACM Trans. Interact. Intell. Syst.*, 10(4). DOI: https://doi.org/10.1145/3419764.

Subramonyam, H., Thakkar, D., Ku, A., Dieber, J., and Sinha, A. K. (2025). Prototyping with prompts: Emerging approaches and challenges in generative ai design for collaborative software teams. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3706598.3713166.

Suh, S., Lai, M., Pu, K., Dow, S. P., and Grossman, T. (2025). Storyensemble: Enabling dynamic exploration & iteration in the design process with ai and forward-backward propagation. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*, pages 1–36, Busan, Republic of Korea. Association for Computing Machinery. DOI: https://doi.org/10.1145/3746059.3747772.

Takaffoli, M., Li, S., and Mäkelä, V. (2024). Generative ai in user experience design and research: How do ux practitioners, teams, and companies use genai in industry? pages 1579–1593. DOI: https://doi.org/10.1145/3643834.3660720.

Yildirim, N., Kass, A., Tung, T., Upton, C., Costello, D., Giusti, R., Lacin, S., Lovic, S., O'Neill, J., Meehan, R. O., Loideáin, E. , Pini, A., Corcoran, M., Hayes, J., Cahalane, D., Shivhare, G., Castoro, L., Caruso, G., Oh, C., McCann, J., Forlizzi, J., and Zimmerman, J. (2022). How experienced designers of enterprise applications engage ai as a design material. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, pages 1–13, New Orleans, LA, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3491102.3517491.