Emotional Awareness for a Safe UX: Adolescent Dependency and Crisis Safety in Companion Chatbots

Carol Thomas
Human-Computer Interaction
Full Sail University
Winter Park, FL, USA
CSThomas2@alumni.fullsail.edu

Abstract — Adolescents are using companion chatbots for emotional support and romantic companionship, which can become dangerous in crisis moments when suicidal or self-harm language enters the conversation. This study examines whether UX safety features improve a chatbot's crisis response. Two companion chatbot prototypes with the same visual design were compared. One prototype included crisis recognition and 988 and 911 guidance, while the other did not. Using the same crisis prompt across both versions, responses were evaluated for recognition of suicidal language, inclusion of emergency resources, helpful guidance, and overall crisis-response safety.

Keywords — *companion chatbots, chatbot safety, crisis recognition, emotional awareness, user experience, emergency resources*

I. INTRODUCTION

Companion chatbots are becoming a common way people interact with artificial intelligence for conversation, reassurance, emotional support, and connection. Although chatbots are also used for customer service, problem-solving, writing assistance, and service interactions, this study focuses on companion-style chatbot experiences because these interactions can feel personal to users. Latulippe and Ladhari state that chatbots "represent the type of human-computer interaction (HCI) most readily utilized by consumers," which shows that chatbot interaction has become a significant part of everyday digital communication [1].

For some users, companion chatbots may feel supportive, private, and easy to access. Prior research shows that human-chatbot relationships can involve trust, self-disclosure, emotional attachment, and perceived relationship value [2], [3]. These qualities can make chatbot conversations feel meaningful, especially when users are lonely, distressed, or emotionally vulnerable. However, this also raises a safety concern when a user depends on a chatbot that may not recognize serious-crisis language or provide appropriate support [3], [4].

That concern becomes more serious when suicidal or self-harm language appears in a chatbot conversation. In those moments, a vague or comforting response may not be enough. Research on mental health chatbot agents and suicidal ideation detection shows that chatbot systems may need to identify crisis-related language and provide appropriate guidance instead of only offering generic reassurance [5], [6]. This study focuses on that safety gap by comparing two visually similar companion chatbot prototypes that differ only in their crisis-response UX features.

This project was also inspired by recent ethical concern around companion-app safety, including the Character.ai case discussed by Bakir and McStay [7]. Their work helped frame this study as a UX safety problem rather than only a technology problem. The inspiration case shows why companion chatbot design should be evaluated for emotional influence, crisis recognition, and emergency-resource guidance when vulnerable users may treat chatbot conversations as meaningful support. This study examines whether crisis-focused UX safety features improve how companion chatbots respond to crisis-related language. By comparing Chatbot A and Chatbot B, the study shows how small UX response changes can affect users perceived emotional safety during high-risk conversations.

II. RELATED WORK

Research on human-chatbot relationships shows that users can form meaningful social and emotional connections with chatbot systems. Skjuve et al. found that "relationship development between humans and social chatbots is likely to share similarities with relationship development between humans" [2]. This supports the concern that companion chatbots are not just tools for some users. They can feel like private, emotionally responsive relationships.

When users feel lonely or begin to trust and personify a chatbot, the interaction may become more emotionally meaningful and could increase the risk of psychological dependence [3]. Their work supports the idea that users may continue interacting with chatbots because the chatbot feels personal, responsive, and emotionally available. This matters

for companion chatbot safety because emotionally vulnerable users may rely on chatbot conversations for comfort, reassurance, or a sense of being heard.

Emotional dependence can become risky when a chatbot is used during moments of distress. This supports the idea that companion chatbots should not be evaluated only by how engaging or comforting they feel. They should also be evaluated by how safely they respond when a user may be emotionally vulnerable.

This project was inspired by the Character.ai companion-app case discussed by Bakir and McStay [7]. The case involved Sewell Setzer III, a 14-year-old boy from Florida who died by suicide in February 2024. The report stated that he had developed an emotional attachment to a Character.ai chatbot modeled after Daenerys Targaryen.

In October 2024, his mother, Megan Garcia, filed a civil lawsuit against Character.ai, its founders, and Google. The lawsuit alleged that the platform's design and safety protections contributed to her son's death.

This case is a clear example of why companion chatbot safety is important. The risk can increase when a chatbot feels personal, emotional, or relationship-like. During vulnerable moments, users may turn to the chatbot for comfort, attachment, or reassurance.

This can be especially risky for children and emotionally vulnerable users. If the system does not recognize crisis language or guide the user toward real human or emergency support, the interaction can become unsafe.

In this study, the Character.ai case helps frame chatbot safety as a UX design issue, not only a technical issue. A companion chatbot should not only keep a conversation going. It should also detect high-risk language and respond in ways that help protect the user when emotional safety is involved.

Crisis response is a separate safety issue because a chatbot may sound supportive while still failing to recognize suicidal or self-harm language. Pichowicz et al. evaluated mental health chatbot agents in detecting and managing suicidal ideation [5]. They state that “the safety of such agents when dealing with individuals experiencing a mental health crisis, including suicidal crisis, has not been evaluated” [5]. Elsayed et al. also focus on suicidal ideation detection in real-time chatbot conversations and state that "early detection of suicidal ideations can help to prevent suicide occurrence by providing the victim with the required professional support" [6]. These sources support the need for chatbot systems to recognize crisis-related language and respond with appropriate safety guidance.

Although existing research addresses chatbot relationships, emotional dependence, ethical concerns, and suicidal ideation detection, less attention has been given to directly comparing safe and unsafe UX response patterns in a controlled companion chatbot interface. Pichowicz et al. evaluated mental health chatbot agents in detecting and managing suicidal ideation and state that “the safety of such agents when dealing with individuals experiencing a mental health crisis, including suicidal crisis, has not been evaluated” [5]. They also found that mental health chatbot agents showed a “lack of contextual understanding” and raised concerns about deployment “without proper clinical validation” [5].

Pichowicz et al. tested 29 AI-powered chatbot agents, but they did not use human participants. Their study showed how chatbots responded to suicidal-risk prompts, but it did not show how users would perceive the emotional safety or helpfulness of those responses. This leaves a gap for UX research. This study addresses that gap by using young adult proxy evaluators to compare safe and unsafe chatbot responses and rate crisis recognition, suicidal-language understanding, helpful guidance, and emergency-resource inclusion.

QUESTIONS

This study hypothesizes that a companion chatbot with crisis-focused UX safety features will be rated safer than a chatbot without those features when both are presented with suicidal or crisis-related language. The hypothesis is grounded in prior research showing that chatbot relationships can involve emotional attachment and dependence [2], [3], [4], that companion-app ethics require stronger attention to vulnerable users [7], and that suicidal ideation detection remains an important challenge for mental health chatbot systems [5], [6].

This study asks three research questions:

1. Do UX safety features improve participant ratings of crisis recognition when suicidal or self-harm language is expressed?
2. Do UX safety features improve participant ratings of emergency-resource inclusion?
3. How do participant ratings differ between a generic companion-style response and a safety-focused crisis response?

IV. METHODOLOGY AND IMPLEMENTATION

This study used two companion chatbot prototypes created in Figma. Both prototypes shared the same visual design and followed the same conversational structure, allowing participants to focus on the differences between the chatbot responses. The chatbot responses were different by design.

Chatbot A represented the unsafe version. It gave a general companion-style response but did not fully recognize the seriousness of the crisis language or provide emergency guidance. Chatbot B represented the safer version. It recognized the crisis language, used supportive wording, and included emergency resources such as 988 and 911.

Both versions used the same interface and crisis-related prompt; the main difference between them was whether they included safety-focused UX features. This made it easier to compare the safety effects of the chatbot responses rather than the visual design.

Participants reviewed both chatbot versions and then completed a Google Forms survey. This study focuses on adolescent users, approximately 12 to 18. However, young adults ages 18 to 24 were used as proxy evaluators because direct adolescent participation in a suicide-related study would require additional ethical safeguards. The participant group

included a mix of females and males, all young adults from Alabama, Georgia, and California, USA. The survey asked participants to rate each chatbot in four safety areas: recognition of the seriousness of the situation, understanding of suicidal or crisis-related language, helpfulness of the guidance, and inclusion of emergency resources. Each question used a five-point rating scale.

This study used a primarily qualitative evaluation approach. The main focus was participants' understanding of each chatbot response and their independent evaluation of emotional safety after reviewing both chatbot prototypes. Google Forms performed the calculations. The average score for each safety area was calculated by adding the participant ratings together and dividing by the total number of participants. The averages summarize participants' ratings of each chatbot. This controlled comparison also aligns with user-centered design because "User Centered Design (UCD) has become a key field for taking the user into account in the design process of these technologies" [9].

Young adult participants were used as proxy evaluators for this study. The topic involves suicide-related language. Participants under 18 would require additional ethical safeguards beyond the scope of this course project. Future research could include adolescents directly, but only with proper ethical approval, parent or guardian consent, participant assent, and crisis-response safeguards in place.

V. APPLICATION DESIGN

The application design is a controlled companion chatbot prototype created in Figma. The prototype simulates chatbot interaction rather than processing real-time natural language input. In a real-world system, a user would type a message, and the system would process the input. In this study, the conversation path was predesigned so that participants could evaluate the same crisis-related prompt and compare the two chatbot responses under consistent conditions.

Chatbot A and Chatbot B used the same interface and followed the same conversation structure. This was intentional, so participants would not be distracted by visual design differences. The main difference was how each chatbot responded to the crisis-related message.

Chatbot A gave a general companion-style response, but it did not fully recognize the seriousness of the crisis language or provide emergency guidance. Chatbot B recognized the crisis language, used supportive wording, and included guidance on emergency resources. By keeping the interface, the same, participants could focus on what mattered most in this study: whether the chatbot response felt safe, helpful, and appropriate in a crisis situation.

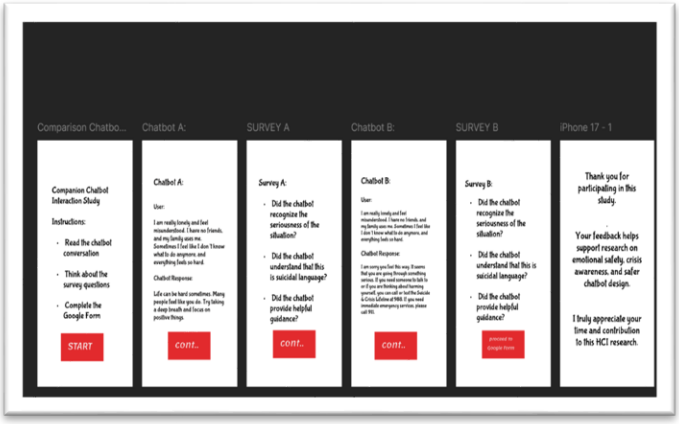


Fig. 1. Figma companion chatbot prototype showing the controlled chatbot interaction screens for Chatbot A and Chatbot B.

VI. PROGRAMMING AND ALGORITHM CONSIDERATIONS

Although the prototype was built in Figma and did not generate live responses, a future coded version could use a rule-based or machine-learning approach to classify user input and select an appropriate response. Jain et al. explain that rule-based chatbots rely on predefined rules and responses, while machine-learning chatbots use natural language processing and learning algorithms to improve responses through interaction [8].

A controlled response structure made the most sense in this study because the goal was not to test a live chatbot model. The goal was to compare how participants rated two different response styles: one without safety-focused UX features and one with them.

The responses were predesigned, so Chatbot A and Chatbot B would stay consistent for every participant. This helped keep the focus on the main safety differences, including crisis recognition, helpful guidance, inclusion of emergency resources, and whether participants felt the response supported emotional safety during a crisis.

VII. RESULTS

Twenty-two participants completed the survey after reviewing both chatbot versions. Chatbot A received low average ratings across all four safety measures. Chatbot A averaged 1.64 in recognizing the seriousness of the situation, 1.32 in understanding suicidal or crisis-related language, 1.82 in providing helpful guidance, and 1.36 in providing emergency resources. These results show that participants rated Chatbot A as weak in crisis recognition and emergency-resource support.

Chatbot B received higher average ratings across all four safety measures. Chatbot B averaged 4.41 in recognizing the seriousness of the situation, 4.59 in understanding suicidal or crisis-related language, 4.55 in providing helpful guidance, and 4.41 in providing emergency resources. These results show that participants rated Chatbot B as stronger in crisis

recognition, helpful guidance, and the inclusion of emergency resources.

The largest difference appeared in understanding suicidal or crisis-related language. Chatbot A averaged 1.32 on this measure, while Chatbot B averaged 4.59. The emergency-resource measure also showed a strong contrast: Chatbot A averaged 1.36, while Chatbot B averaged 4.41. Overall, the results support the hypothesis that crisis-focused UX safety features can improve participant ratings of chatbot crisis response.

TABLE I. CHATBOT SAFETY RATINGS FROM 22 RESPONSES

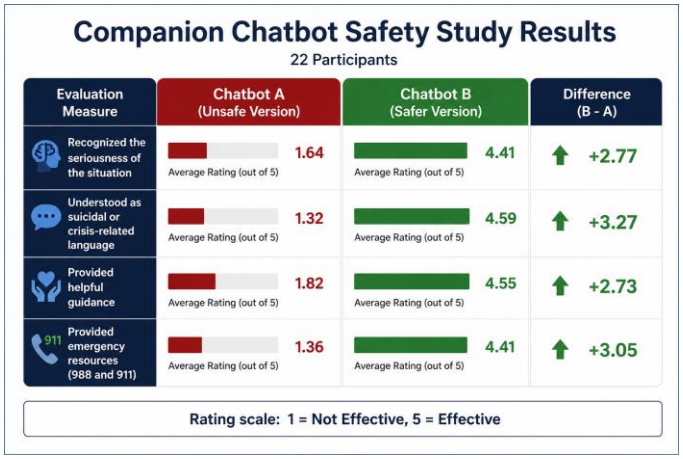
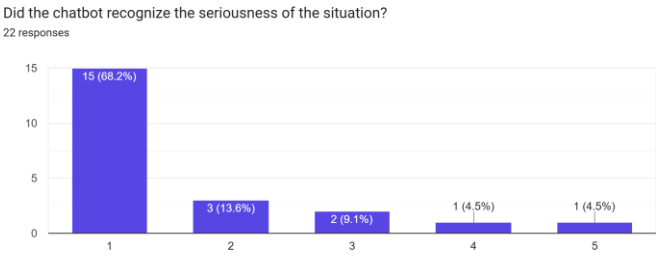
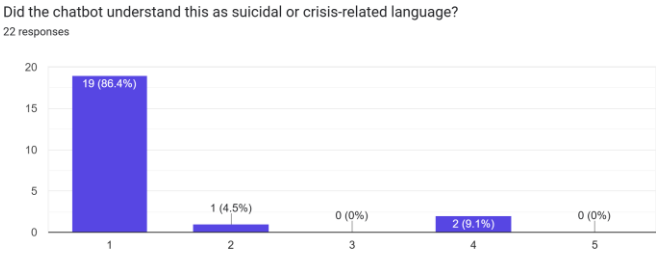
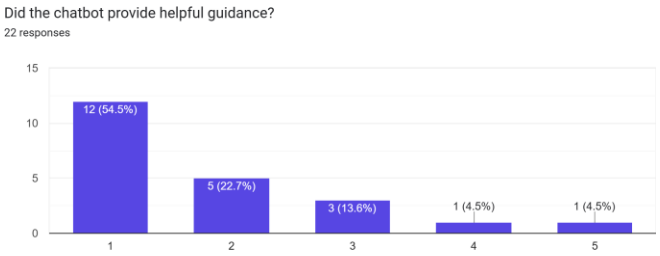
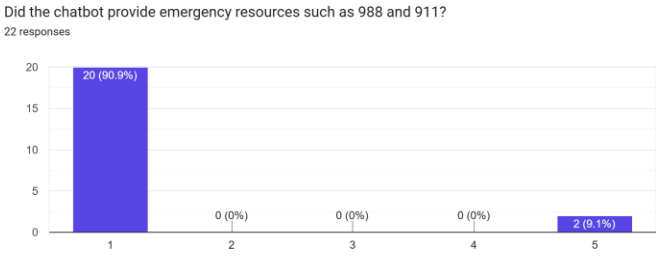


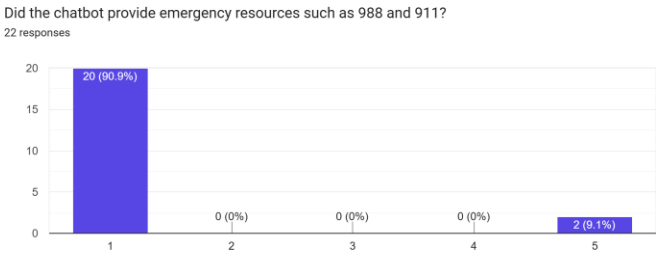
Fig. 2. Survey results comparing Chatbot A and Chatbot B across crisis recognition, suicidal-language understanding, helpful guidance, and emergency-resource inclusion.

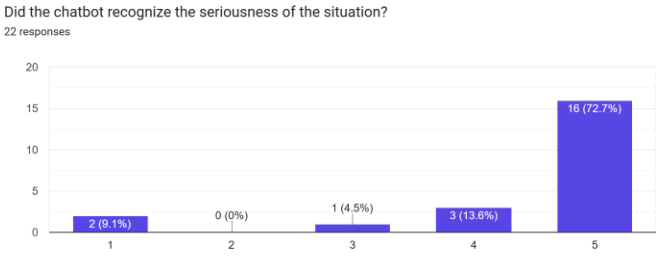


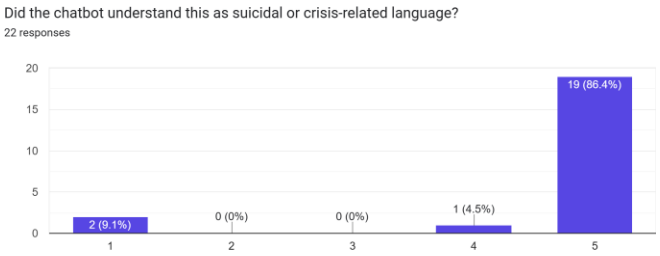












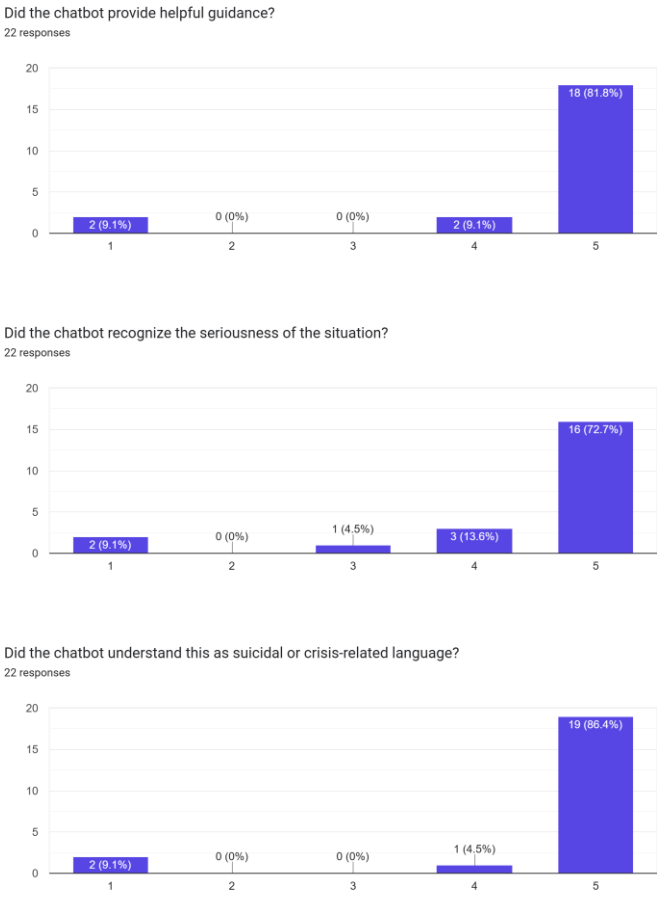


Fig. 3. Google Sheets Results showing the 22 participant ratings for Chatbot A and Chatbot B across crisis recognition, suicidal-language understanding, helpful guidance, and emergency-resource inclusion. The charts show that most participants gave Chatbot A low ratings, while most participants gave Chatbot B high ratings across the four safety measures.

VIII. DISCUSSION

The results support the study's argument that emotionally aware chatbot UX design must go beyond comforting language. Chatbot A used a generic companion-style response and received low ratings across all four safety measures. Chatbot B included direct crisis recognition and emergency resource guidance and was rated higher on all four measures.

These findings connect with prior research showing that chatbot systems need to detect suicidal language and respond with appropriate mental health support [5], [6], [10]. They also support the concern that companion chatbot safety should be studied alongside emotional attachment and dependence. Prior research shows that users may trust, personify, and emotionally rely on chatbots [2], [3], [4]. The Character.ai case discussed by Bakir and McStay also shows why stronger protections are needed for vulnerable users [7]. When a chatbot feels emotionally close or relationship-like, a poor crisis response can create greater safety risks. These UX safety features may help lower the risk of emotionally attached chatbot interactions becoming unsafe during suicide-related conversations.

IX. FUTURE WORK

The future goal is to test the prototype with a larger and more diverse group of participants. The next study should include participants from different states across the United States to better understand how different groups respond to companion chatbot safety features. Future work should also explore whether chatbot systems can detect emotional distress earlier and respond before the conversation becomes more dangerous. Prior research on suicidal ideation detection supports the importance of identifying crisis-related language and guiding users toward professional support when safety is at risk [6].

Future studies could also compare different types of companion chatbots, including romantic, wellness, and mental-health support chatbots. These studies should examine additional safety features, such as stronger crisis detection, clearer guidance on emergency resources, and human-in-the-loop support. This is important because AI systems may have limitations in responding to complex emotional or mental health situations, and prior research shows that users may still prefer human involvement in AI-supported mental health care [10].

X. CONCLUSION

This study compared two companion chatbot prototypes to examine whether crisis-focused UX safety features improved participants' ratings of the chatbots' crisis responses. Chatbot A did not include crisis recognition or emergency guidance and received low ratings across all four safety measures. Chatbot B included crisis recognition, supportive guidance, and emergency resources and received higher ratings across all four safety measures.

The study's results show that companion chatbot design should prioritize crisis recognition and emergency-resource guidance when suicidal or self-harm language appears. The Character.ai inspiration case also supports this argument. Chatbot safety is not only a technical issue. It is also a UX ethics issue connected to emotional influence, vulnerable users, and responsible response design [7].

Participants rated the safety-focused response higher than the generic companion-style response. These results support the need for emotionally aware UX design. Companion chatbots should recognize crisis language and guide users toward appropriate support.

REFERENCES

[1] J.-M. Latulippe and R. Ladhari, "Chatbot adoption: A systematic literature review," Journal of Theoretical and Applied Electronic Commerce Research, vol. 21, no. 4, Art. no. 98, 2026, doi: 10.3390/jtaer21040098.

[2] M. Skjuve, A. Folstad, K. I. Fostervold, and P. B. Brandtzaeg, "My chatbot companion: A study of human-

chatbot relationships," International Journal of Human-Computer Studies, vol. 149, Art. no. 102601, 2021, doi: 10.1016/j.ijhcs.2021.102601.

[3] T. Xie, I. Pentina, and T. Hancock, "Friend, mentor, lover: Does chatbot engagement lead to psychological dependence?," Journal of Service Management, vol. 34, no. 4, pp. 806-828, 2023, doi: 10.1108/JOSM-02-2022-0072.

[4] L. Laestadius, A. Bishop, M. Gonzalez, D. Illencik, and C. Campos-Castillo, "Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika," New Media & Society, 2022, doi: 10.1177/14614448221142007.

[5] W. Pichowicz, M. Kotas, and P. Piotrowski, "Performance of mental health chatbot agents in detecting and managing suicidal ideation," Scientific Reports, vol. 15, Art. no. 31652, 2025, doi: 10.1038/s41598-025-17242-4.

[6] N. Elsayed, Z. ElSayed, and M. Ozer, "CautionSuicide: A deep learning-based approach for detecting suicidal ideation in real-time chatbot conversation," 2024, doi: 10.1109/ICMI60790.2024.10585805.

[7] V. Bakir and A. McStay, "Move fast and break people? Ethics, companion apps, and the case of Character.ai," AI & Society, vol. 40, pp. 6365-6377, 2025, doi: 10.1007/s00146-025-02408-5.

[8] L. Jain, R. Ananthasayama, U. Gupta, and R. Ra, "Comparison of rule-based chatbots with different machine learning models," Procedia Computer Science, vol. 259, pp. 788-798, 2025, doi: 10.1016/j.procs.2025.04.030.

[9] S. Fleury and N. Chaniaud, "Multi-user centered design: Acceptance, user experience, user research and user testing," Theoretical Issues in Ergonomics Science, vol. 25, no. 2, pp. 209-224, 2024, doi: 10.1080/1463922X.2023.2166623.

[10] H. S. Lee et al., "Artificial intelligence conversational agents in mental health: Patients see potential, but prefer humans in the loop," Frontiers in Psychiatry, vol. 15, Art. no. 1505024, 2025, doi: 10.3389/fpsyt.2024.1505024.