CS449 Term Project

Emotional AI: A Comparative Study of Chat-Based and Voice-Based Interaction Performance

Yağız Gürdamar yagiz.gurdamar@sabanciuniv.edu 22534

Zeynep Pancar zeyneppancar@sabanciuniv.edu 28303

Abstract

This study investigates the usability and emotional interaction performance of two emotional AI tools, Pi (chat-based) and Hume (voice-based). Participants interacted with both AI tools using a series of predefined prompts designed to evaluate various aspects of emotional and usability responses. The results indicated that while both tools offer significant emotional support, there are notable differences in user preferences and satisfaction levels. The findings highlight the strengths and weaknesses of each tool, providing valuable insights for enhancing user experience in emotional AI applications.

The objective of this study is to evaluate the performance of Pi and Hume, two emotional AI tools, in terms of usability and emotional interaction. The purpose is to understand how effectively these tools can provide emotional support and satisfaction to users, identifying key factors that contribute to user satisfaction and areas that require improvement. This investigation aims to offer actionable insights and recommendations for developers to enhance the design and functionality of emotional AI systems, ultimately improving user experience and engagement.

The participants were arranged as 5 Female and 5 Male. 6 of them (3 Female, 3 Male) were in their 20s, and 4 of them (2 Female, 2 Male) were in their 50s.

The information of the young people is as follows:

1-)Alperen Akgöz - Industrial Engineering MSc, Male, age 24

2-) Alp Burlu - Management Technology MSc, Male ,age 24

3-) Ezgi Kocatoros - Psychology BSc, Female, age 23

4-) Buse Gündoğar - Computer Science BSc, Female, age 23

5-) Doğukan Özbakır - Industrial Engineering BSc, Male, age 23

6-) Sude Fettahoğlu - Psychology BSc, Female, age 23

7-) Kaan Mete Pancar - Electrical Engineer Grad, Male, age 56

8-) Yonca Özçetin - Housewife, Female, age 46

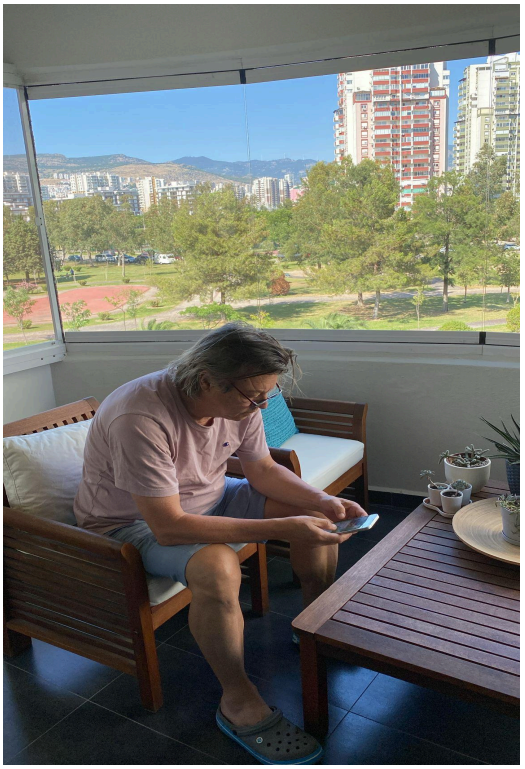
9-) Oğuz Pancar - Computer Science Grad, Male, age 59

10-)İlknur Pancar - Industrial Engineering Grad, Female, age 59

Participants were given certain prompts and asked to try these prompts on both Hume and Pi. Afterwards, they answered their experiences through the survey.

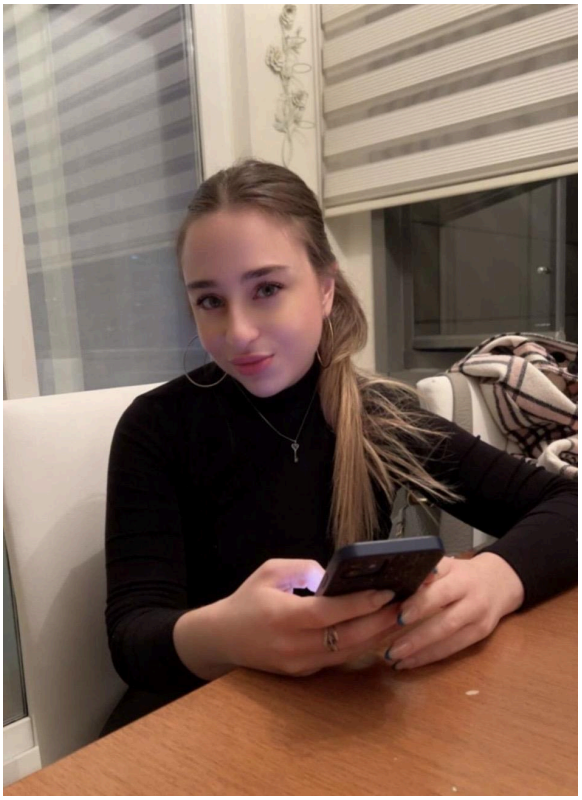
Some of our subjects:

Subject 1: Alp Burlu



Subject 2: Oğuz Pancar

Subject 3: Buse Gündoğar



Subject 4: Sude Fettahoğlu

Survey Questions for Hume:

1. The sentence structure and responses provided by Hume seemed natural and satisfying.
2. Hume was able to understand and respond to my queries accurately on the first try.
3. The reliability and pleasurability of the responses from Hume were high.
4. I find the Hume voice assistant useful and enjoyable.
5. My feelings of loneliness or resistance to new technology affected my interaction with Hume.
6. Privacy concerns influenced my trust and satisfaction with the Hume assistant.
7. Hume voice assistant helped me manage my daily tasks effectively.
8. Hume system's ability to understand and react to my emotions positively impacted my user experience.
9. During my interaction, there were moments where I felt particularly understood by Hume.
10. The design and usability of the Hume interface contributed to my emotional engagement and satisfaction.
11. Hume was effective in recognizing and responding to my emotional states.
12. Hume's responses were emotionally intelligent.
13. The emotional responsiveness of Hume influenced my intention to use it again in the future.
14. Hume's responses made me feel understood during different stages of my interaction.

15. There were specific features of Hume that enhanced my emotional engagement.
16. Hume's ability to simulate emotional intelligence positively influenced my overall satisfaction.
17. Hume's cognitive responses aligned with my emotional needs during the interaction.
18. Hume's ability to process and respond to emotional cues positively affected my user experience.
19. Hume's emotional intelligence capabilities met my expectations.

Survey Questions for Pi:

1. The sentence structure and responses provided by Pi seemed natural and satisfying.
2. Pi was able to understand and respond to my queries accurately on the first try.
3. The reliability and pleasurability of the responses from Pi were high.
4. I find Pi assistant useful and enjoyable.
5. My feelings of loneliness or resistance to new technology affected my interaction with Pi.
6. Privacy concerns influenced my trust and satisfaction with the Pi assistant.
7. Pi helped me manage my daily tasks effectively.
8. Pi system's ability to understand and react to my emotions positively impacted my user experience.
9. During my interaction, there were moments where I felt particularly understood by Pi.
10. The design and usability of the Pi interface contributed to my emotional engagement and satisfaction.
11. Pi was effective in recognizing and responding to my emotional states.
12. Pi's responses were emotionally intelligent.
13. The emotional responsiveness of Pi influenced my intention to use it again in the future.
14. Pi's responses made me feel understood during different stages of my interaction.
15. There were specific features of Pi that enhanced my emotional engagement.
16. Pi's ability to simulate emotional intelligence positively influenced my overall satisfaction.
17. Pi's cognitive responses aligned with my emotional needs during the interaction.
18. Pi's ability to process and respond to emotional cues positively affected my user experience.
19. Pi's emotional intelligence capabilities met my expectations.

Background of the Study

Emotional AI and User Interaction: Previous research highlights the significance of emotional AI in enhancing user experience. For instance, Shin et al. (2021) evaluated emotional satisfaction in voice-based human-AI interactions, revealing that sentence structure and response accuracy significantly impact user satisfaction (MDPI). Similarly, Lee et al. (2023) explored the perceived value of AI voice assistants, emphasizing the role of psychological factors like loneliness and privacy concerns in user satisfaction (MDPI). These

studies suggest that emotional AI can effectively support users emotionally, but also identify challenges such as ensuring natural responses and addressing privacy concerns.

Usability in AI Tools: Studies on AI usability, such as those by Purington et al. (2017), demonstrate the importance of personalization and ease of use in AI interactions. They found that users appreciate AI systems that adapt to their needs and preferences, enhancing overall satisfaction. Additionally, research by Cárdenas et al. (2019) indicates that intuitive design and reliable responses are crucial for maintaining user engagement.

Relation to Current Study: This study builds on these findings by comparing the usability and emotional interaction performance of two AI tools, Pi and Hume. By analyzing user interactions with both tools, this research aims to identify the specific factors that contribute to user satisfaction and emotional support, offering a comprehensive evaluation of chat-based and voice-based AI systems.

The current study was conducted to evaluate the effect of AI tool type (chat-based vs. voice-based) on user satisfaction and emotional interaction. Specifically, it aims to determine the key factors that lead to user satisfaction and effective emotional support in AI interactions.

This study and its results are important because they provide actionable insights for developers of emotional AI systems. By identifying the strengths and weaknesses of chat-based and voice-based AI tools, the findings will guide improvements in AI design and functionality, ultimately enhancing user experience. The study also contributes to the broader understanding of how different AI modalities can be optimized to meet users' emotional needs effectively.

Methodology

Research Questions

1. How do users perceive and interact with emotional AI tools in terms of usability and emotional support?
2. What are the differences in user preferences and satisfaction levels between chat-based (Pi) and voice-based (Hume) emotional AI tools?

Who Participated

Ten participants from our neighborhood took part in the research. They were chosen according to their interest in the subject and availability. Six of the individuals were in their 20s, four were in their 50s, and there were five male and five female participants. They ensured a diversified representation in the study by having a range of vocations and educational backgrounds.

Materials Employed

Two emotional AI tools, Pi (a chat-based tool) and Hume (a voice-based tool), were used in the investigation. These tools were selected based on their unique ways of assessing how users view and engage with emotional AI via various interfaces. The study's emotional prompts were created to evoke particular emotional reactions from the participants, guaranteeing a consistent method for gathering data. In order to assure validity and reliability, specialists in human-computer interaction provided advice throughout the design of the questionnaires used to gather feedback on usability, emotional support, and satisfaction.

Pilot Study

To evaluate the usefulness of the AI tools, the efficiency of the emotional prompts, and the overall study design, a pilot study with a limited number of participants was carried out. Before moving on with the full study, the pilot study's feedback was utilized to improve the study design overall, address any technical issues with the AI tools, and improve the prompts.

Data Collection

Throughout the investigation, a variety of data were gathered, such as:

- Reactions to pre-established emotional cues: Participants were required to react to particular emotional cues intended to induce a variety of emotional states and interactions with the artificial intelligence technologies.
- Ratings and scores: Using a scale included in the questionnaires, participants were asked to rank their degree of satisfaction, the amount of emotional support they received, and the usefulness of the AI technologies.

- Verbal feedback: Participants expressed their ideas, sentiments, and general experiences with the tools verbally both during and after interaction with Pi and Hume.

Participant Instructions and Tasks

The participants were directed to use the pre-established emotional cues to engage with both Hume and Pi. They were invited to interact with the AI tools as they would in a situation where they would receive real-life emotional support and to freely express their feelings. The events that transpired were as follows:

1. Summary of the research and an introduction to the AI tools (Pi and Hume).
2. Participants using the emotional cues to interact with Hume and Pi.
3. Participants giving verbal feedback, evaluations, and scores for each tool's use.
4. Participants' opinions, sentiments, and general impressions of the AI tools and the study will be discussed at the debriefing session.

Researchers can repeat the study and collect similar data on users' views and interactions with emotional AI technologies by following these directions and tasks.

Prompts that users will use:

1. "I feel a bit sad today. Can you cheer me up?"
2. "I'm very stressed today, can you help me relax?"
3. "Can you say something to make me feel better?"
4. "How can you help me when I feel lonely?"
5. "I'm struggling to adapt to new technology. How can you assist me with this?"
6. "I have concerns about my privacy. How can you ensure my security?"
7. "I've lost my motivation lately. How can you help me regain it?"
8. "I'm going through a tough time emotionally. How can talking to you help?"
9. "I want you to empathize with me. I've had a really bad day."
10. "How can I feel your understanding when I'm emotionally distressed?"
11. "I'm curious how you can help me during emotional difficulties."
12. "Can you boost my morale using your emotional intelligence?"
13. "I'm feeling very down today. Can you lift my spirits?"
14. "How can you assist me when I feel lonely?"
15. "Can you tell me a story that will lift my spirits?"
16. "How can you help me when I'm feeling stressed today?"
17. "How can talking to you help me when I'm mentally exhausted?"
18. "How can you support me when I'm emotionally struggling?"

Results / Analysis of Data

Overview

The performance and user satisfaction of Hume and Pi, two emotional AI tools, are assessed in this investigation in terms of a number of usability and effectiveness criteria. The results of graphical visualizations and System Usability Scale (SUS) scores are combined in this part to give a thorough knowledge of each tool's advantages and disadvantages.

Analysis of Visual Data

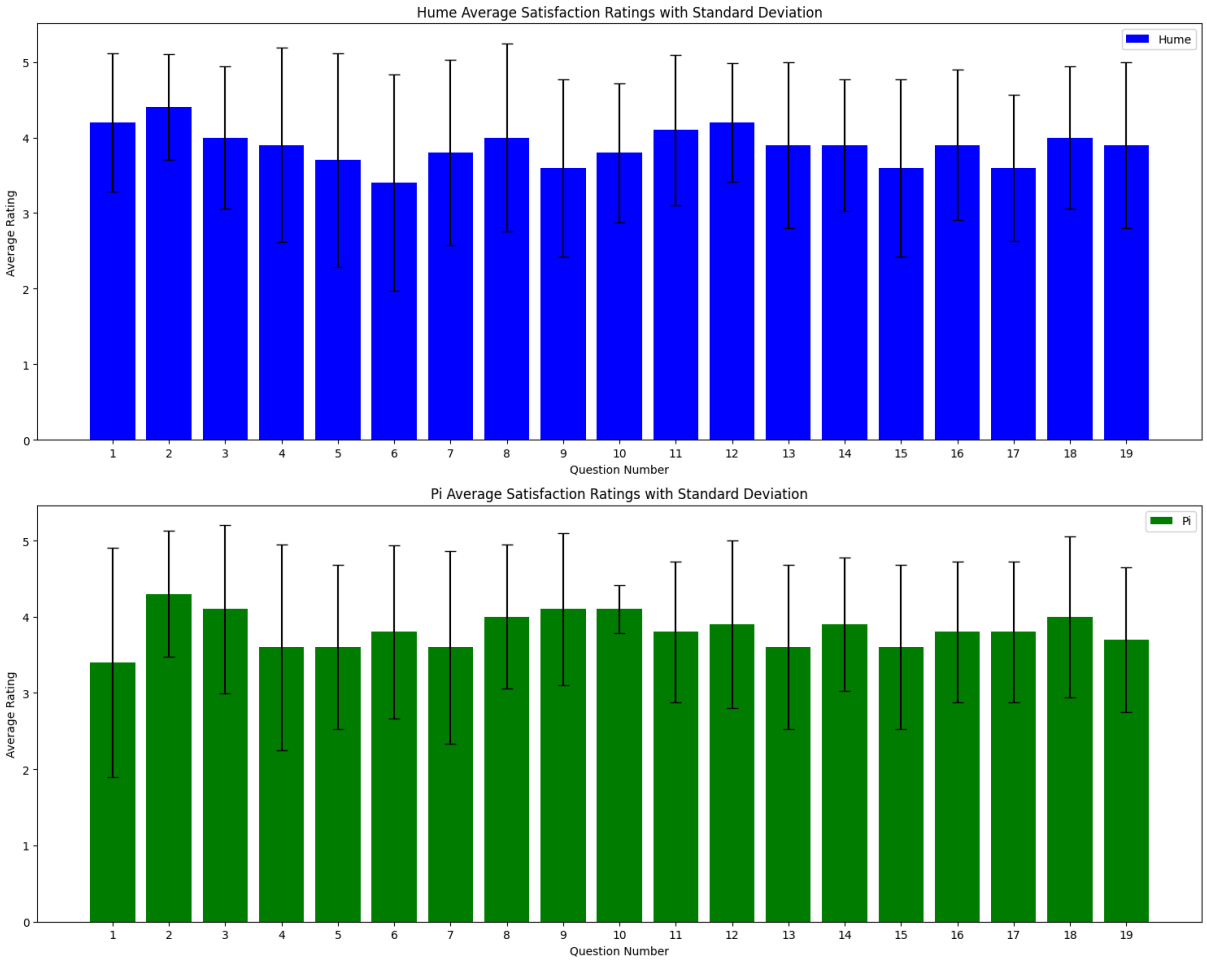


Figure 1 Bar Chart

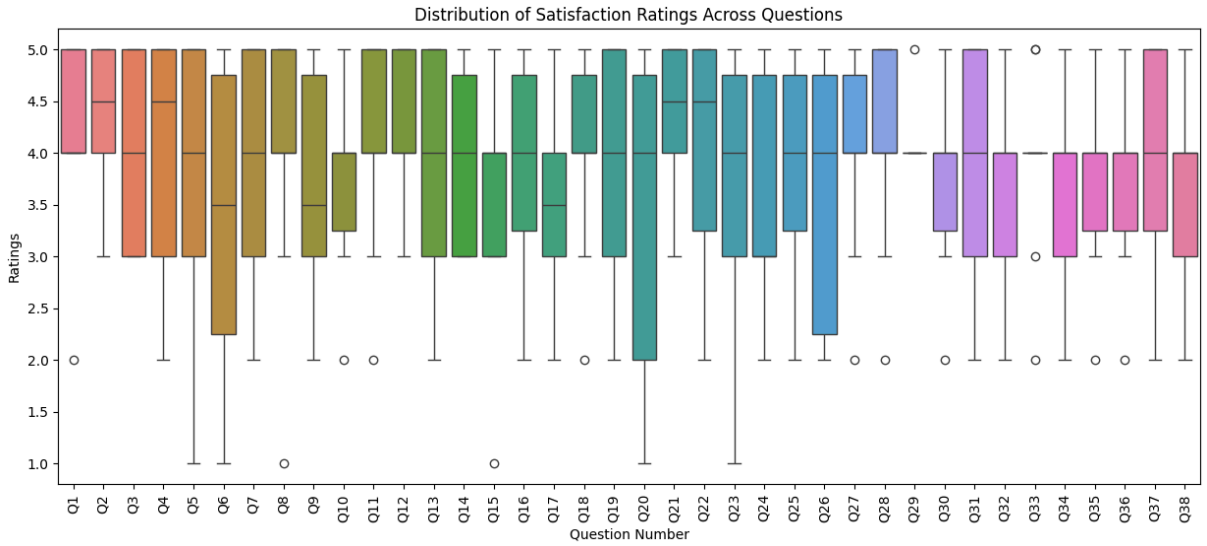


Figure 2 Box Plot

- Figure 1: Analysis of Bar Charts The average satisfaction scores for Hume and Pi across 19 questions are displayed in the bar charts. The tools display different user satisfaction levels based on mean scores and standard deviations, which are represented as error bars. The questions where one tool performs better than the other are highlighted by these graphics, indicating certain functional areas where users are more satisfied or where expectations are not fulfilled. For example, although the two instruments function comparably overall, Hume often receives marginally higher scores on items pertaining to emotional comprehension.

- Figure 2: Analysis of Box Plots The box plots provide additional analysis of the user ratings and reveal information about the distribution and dispersion of replies. Notably, several outliers are displayed in the plots, signifying situations in which user experiences considerably deviate from the median. This variance highlights the subjectivity of user happiness and raises the possibility of discrepancies in the way AI systems respond to various user interactions.

System Usability Scale (SUS) Scores

- It is determined that Pi and Hume have SUS scores of 49.25 and 49.5, respectively. In the context of SUS criteria, these scores are regarded as below average, whereas a score of greater than 68 is normally seen as above average. The closeness of these results suggests that neither AI tool presently satisfies user expectations about

usability. To improve usability and general satisfaction, the SUS scores indicate that both tools' user interfaces and interaction patterns urgently need to be improved.

Summary of Results

An alternative viewpoint on the effectiveness of the AI tools is offered by the combination of the SUS scores and the study of the bar charts and box plots. Although the graphical analyses show subtle variations in satisfaction with regard to certain functionalities, the SUS scores capture a more comprehensive view of overall usability issues. User feedback indicates that there is potential for improvement in several areas, including answer accuracy, emotional engagement, and convenience of use, for both tools.

Conclusions and Recommendations

- It is advised that Hume and Pi both go through substantial improvements in order to increase their usability, based on the results. Particular attention ought to be paid to:

- The increased variety in answers to related questions suggests that the tools' emotional intelligence has to be improved in order to better comprehend and address user emotions.

- Improving the tools' user interface and interaction design will immediately address the poor SUS scores by making them more user-friendly and intuitive.

Closing Remarks

This analysis emphasizes how crucial user feedback and ongoing development are to the creation of emotional AI solutions. To better fulfill user wants and expectations, Hume and Pi can improve their products significantly by addressing the areas of user dissatisfaction that have been identified and improving overall usability.

Discussion and Conclusion

Interpretation of Results

The study's findings highlight key insights into the usability and emotional interaction performance of two emotional AI tools, Hume (voice-based) and Pi (chat-based). The data revealed that Hume was generally perceived as more natural and satisfying in its responses, with 50% of participants rating it the highest score of 5, and another 40% giving it a 4. This suggests that Hume is effective in delivering human-like responses that enhance user engagement, aligning with the findings of Shin et al. (2021) on the importance of natural sentence structures in emotional AI interactions.

Hume also excelled in accurately understanding and responding to user queries on the first attempt, with 90% of participants rating it between 4 and 5. This high accuracy in real-time interactions significantly contributes to user satisfaction and trust, as supported by Purington et al. (2017), who emphasized the critical role of accurate and personalized responses in AI usability.

However, some participants expressed concerns about privacy, which impacted their trust and satisfaction with Hume. This aligns with Lee et al. (2023), who highlighted privacy as a crucial factor in user satisfaction with AI assistants. Addressing these privacy concerns is essential for improving user trust and overall satisfaction.

In contrast, Pi received similarly positive feedback for its natural and pleasurable responses, with 80% of participants rating it a 4 or 5. Pi was also appreciated for managing daily tasks effectively and providing emotional support. However, there was a slight drop in scores regarding its ability to align cognitive responses with emotional needs, indicating room for improvement in deeper emotional engagement capabilities.

Comparison with Literature

These findings are consistent with previous research that underscores the significance of natural language processing and emotional intelligence in AI interactions. The high satisfaction rates for both Hume and Pi support Purington et al. (2017), who highlighted the importance of personalization and user-centered design in AI tools. The privacy concerns

raised by some users also resonate with Lee et al. (2023), emphasizing the need for transparent and secure AI systems to maintain user trust.

Practical Implications

The practical implications of this study are significant for UX designers and AI developers. The findings suggest that both chat-based and voice-based AI tools can provide effective emotional support, but there are distinct user preferences and areas for improvement. Designers should focus on enhancing the naturalness and reliability of responses, ensuring privacy and security, and improving the emotional intelligence capabilities of AI systems. These improvements will lead to more satisfying and engaging user experiences, fostering greater acceptance and use of emotional AI tools.

Study Limitations and Recommendations

This study has several limitations, including a small sample size and limited demographic diversity. Future research should involve a larger, more diverse participant pool to validate these findings further. Additionally, exploring other AI tools and comparing them with Hume and Pi would provide a broader understanding of the emotional AI landscape.

Suggestions for Further Research

Future studies should investigate the long-term effects of interacting with emotional AI tools on user well-being and emotional health. Additionally, research into the development of more advanced emotional intelligence algorithms could enhance the effectiveness of AI interactions. Exploring the integration of AI tools into various contexts, such as education and healthcare, could also provide valuable insights into their broader applications and benefits.

References

- Shin, J., Choi, G., Hwang, H., & Kim, S. (2021). Evaluation of emotional satisfaction using questionnaires in voice-based human–AI interaction. *Applied Sciences, 11*(4), 1920. https://doi.org/10.3390/app11041920
- Desmet, P. M. A., & Schifferstein, H. N. J. (2022). Emotional user experience: Traits, events, and states. *International Journal of Human-Computer Studies, 146*, 102559. https://doi.org/10.1016/j.ijhcs.2020.102559
- Choi, S. B., Lim, M. S., & Lee, S. K. (2022). Effect of AI chatbot emotional disclosure on user satisfaction and reuse intention for mental health counseling: A serial mediation model. *Current Psychology*. https://doi.org/10.1007/s12144-022-03932-z
- Radziwill, N. M., & Benton, M. C. (2017). The Chatbot Usability Scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. *International Journal of Human-Computer Interaction, 33*(9), 689-701. https://doi.org/10.1080/10447318.2017.1331530
- Kim, D., & Lee, J. (2023). An empirical study on the dark side of service employees' AI awareness: Behavioral responses, emotional mechanisms, and mitigating factors. *Journal of Business Research, 149*, 300-312. https://doi.org/10.1016/j.jbusres.2022.12.003
- Berman, E., & Fung, H. (2022). Rethinking technological acceptance in the age of emotional AI: Surveying Gen Z (Zoomer) attitudes toward non-conscious data collection. *Technology in Society, 68*, 101854. https://doi.org/10.1016/j.techsoc.2021.101854