# Crafting Human-AI Collaborative Analysis for User Experience Evaluation

Emily Kuang
Rochester Institute of Technology
Rochester, New York, USA
ek8093@rit.edu

## ABSTRACT

AI has been increasingly adopted in user experience (UX) analysis, in which UX evaluators review test recordings to identify usability problems. However, most AI-infused systems apply fully automatic approaches, leading to distrust from UX evaluators. In my dissertation work, we consider AI as assisting, not replacing human judgment. Through an international survey, we investigated the current practices and challenges of UX evaluators and identified an opportunity for AI assistance. We then studied nuanced cooperative work between UX evaluators and AI, by employing either non-interactive visualizations or interactive conversational assistants (CAs). The next steps include building upon our findings about the reactive Q&A dynamic with CAs, by exploring how a proactive approach or a combination of visualizations and CAs may better support UX evaluators. This research will identify interactions and representations that give rise to productive and trusting collaborations with AI.

## CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; **Natural language interfaces**.

## KEYWORDS

User experience; Usability testing; Human-AI collaboration, Conversational agents

**ACM Reference Format:**
Emily Kuang. 2023. Crafting Human-AI Collaborative Analysis for User Experience Evaluation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3544549.3577042

## 1 CONTEXT AND MOTIVATION

User experience (UX) is central to the adoption of technology because usefulness and usability define the efficacy and capacity of the human-system relationship [19, 35, 42]. In order to achieve smooth user experiences, UX evaluators conduct usability testing to detect and address usability problems. However, analyzing usability test recordings is challenging and time-consuming [5, 7, 10, 36]. UX evaluators have limited time and resources, which could lead to missed information or misinterpreted problems if only one person completes analysis [10, 17, 26, 36]. Despite the value of working with others to improve reliability and completeness, few evaluators employ collaboration in practice [10, 11, 26]. In the international survey of 279 UX evaluators that I conducted in 2021, only 37% reported collaboration when analyzing the same recordings [26] and in other cases, it was found that matching teams or pairs was costly in terms of time, resources, and effort [7]. Thus, prior work shows that although individual analysis can be problematic, effective collaboration is often hindered by limited resources.

To alleviate the scarcity of human-human collaboration, AI assistance is considered an effective tool that could boost the efficiency of UX evaluators and reliability of results [6]. Some commercial analytical platforms already contain features derived from AI and machine learning (ML) (e.g., emotion detection [44], sentiment analysis [43]). Researchers have also incorporated ML and AI into the UX field [16, 22, 37, 38, 47, 49], primarily using fully automated methods that track user interaction events like keystrokes or mouse movements. However, they lead to incomplete results because some usability problems are very difficult to capture, or cannot be captured automatically at all [14]. Furthermore, fully automated systems lack transparency and are viewed with distrust: UX evaluators face challenges in understanding AI's capabilities and are skeptical of AI-generated results without explanations of the underlying algorithm [41, 46]. Thus, instead of full automation, we consider AI as an assistant, to *augment* manual analysis rather than replace skilled knowledge and reasoning. We posit that an interactive human-technology partnership may help UX evaluators to discover usability problems that may be overlooked by solely manual or automatic methods. In sum, my dissertation research involves the following steps towards crafting human-AI collaborative usability analysis:

- We investigate the current practices of UX evaluators and identify opportunities for AI assistance [26].
- We design and evaluate a visual analytics tool that displays AI-driven features extracted from usability test recordings to UX evaluators [41].
- We then explore how interactive conversational assistants (CAs) may enhance human-AI collaborative analysis by understanding the range of questions UX evaluators would ask an AI assistant [*under review*].
- We propose a user study that investigates nuanced proactive interactions between UX evaluators and conversational AI assistants.
- Finally, we propose a summative study to validate our approach. This study will examine how UX evaluators utilize our tool that provides a combination of visualizations and CAs to improve analytic performance.


*CHI EA '23, April 23–28, 2023, Hamburg, Germany*

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3577042

# 2 BACKGROUND AND RELATED WORK

## 2.1 Current UX Analysis Practices and the Need for Collaboration

Usability testing is a frequently employed method for detecting usability problems [7]. UX evaluators assess recorded user sessions by observing user actions and writing notes simultaneously [5]. However, analyzing recordings using these manual approaches is challenging and time-consuming because evaluators have limited time and resources, which could lead to missed information or misinterpreted problems [8, 10, 17, 36]. Thus, collaboration is crucial since it allows UX evaluators to balance reliability and efficiency by dividing the workload and consolidating their results [7, 17, 18]. Collaborations also alleviate the *"evaluator effect,"* the condition in which different evaluators identify different sets of UX problems even when analyzing the same test session [17, 20], and therefore ensures comprehensive coverage.

Despite the value of working with others to improve reliability and completeness, few evaluators employ collaboration in practice [10]. A survey of 197 UX practitioners found that more than half (56%) of the participants analyzed data and wrote informal reports *alone* [7]. Similarly, our CHI22 survey showed that only 37% reported collaboration when analyzing the same recordings [26]. These findings show that although individual analysis can be problematic, effective collaboration is often hindered by limited resources. Thus, we identify an opportunity for AI assistance as an effective tool to augment the efficiency of UX evaluators and reliability of results. Advances in natural language processing and machine learning (ML) enable automatic cues detection from acoustic, textual, and visual channels available in recorded sessions [6, 14]. Whereas UX evaluators may miss banal usability problems, especially under time pressure, leveraging AI to capture low level and commonplace issues could sustain high accuracy and efficiency [6]. We consider different approaches in the next section.

## 2.2 Incorporating AI into Usability Analysis

The potential benefits of integrating AI in usability analysis gave rise to an increasing trend of using AI to detect UX problems [14, 16, 22, 37, 38, 47, 49]. User interaction events were utilized to create machine learning (ML) classifiers to detect usability issues of websites [14, 38], mobile applications [22], and virtual reality (VR) applications [16]. However, only two-thirds of usability problems were detected by automated algorithms when compared to manual testing [14]. These results indicate that although automated methods can find meaningful problems, they cannot replace human reasoning required for completeness. Furthermore, these automatic methods were primarily based on users' interaction logs [22], which are only applicable for specific types of products, such as digital interfaces. Due to the limitations of automated methods, there is growing interest in human-AI collaboration where human decision making is supplemented with AI assistance [27]. Recent work developed tools where UX evaluators can utilize visualizations of ML-driven features to inform their identification of usability problems [8]. My dissertation research builds upon the tenet of AI as an assistant to *augment* manual analysis instead of automatic detection without humans-in-the-loop.

## 2.3 Human-AI Collaboration via Interactive Conversational Assistants

AI's increasing ability to better detect patterned usability issues in rich visual and audio data streams enables assistive agents to work with humans toward increased productivity and problem solving [33, 45]. Conversation is a key mode of human-computer interaction [30]. Conversational assistants are increasing in both professional and personal use, where 70% of white-collar workers are expected to interact with text chatbots on a daily basis in 2022 [13]. Since text and speech are the two main ways to interact with conversational assistants, prior work has compared the two modalities and demonstrated solid differences in user behavior between them [23, 24, 31, 34]. Recent research also demonstrated a need for AI to match relevant social norms and to show contextually relevant information [1, 46], and for conversational assistants to exhibit emotionally appropriate responses [12, 48]. Furthermore, providing explanations for predictions has been shown to alleviate users' uncertainty and enhance their trust in AI systems [29]. Although prior research investigated perceptions of conversational assistants in collaborative games [2, 3] and productivity applications [15], the use of a conversational agent for UX analysis has been unexplored. We hypothesize that an interactive assistant—in the form of a conversational agent—better supports analysis on rich behavioral data while also enhancing trust and engagement to guide UX evaluators in discovering problems otherwise overlooked. Through the reciprocal dialogue between humans and AI, a deeper cooperation is established, bringing huge benefits analogous to when people with different capabilities collaborate to achieve a goal [4, 21]. In sum, my dissertation research explores data rich capabilities alongside interactive engagement with conversational assistants, toward leveraging AI capability beyond data logs and toward collaborative and contextually relevant usability analysis and information sharing.

# 3 RESEARCH QUESTIONS

Despite AI being increasingly adopted in various areas of work, there is a need for better approaches to defining the role of AI as an assistant to UX evaluators. My dissertation explores the following research questions:

- **RQ1: What are the current practices, challenges, and desired improvements for collaborative data analysis of UX evaluators?** [26]
- **RQ2: How does providing visualizations of AI-driven features support evaluators in analyzing usability test sessions?** [41]
- **RQ3: What types of questions will UX evaluators ask a conversational AI assistant during analysis? How do their behavior differ between text and voice assistants?** [*under review*]
- **RQ4: How do UX evaluators interact with a proactive AI assistant that provides suggestions? When should these suggestions be displayed?** [*proposed*]
- **RQ5: How does providing a combination visualizations and proactive conversational assistants support UX evaluators?** [*proposed*]

# 4 RESEARCH METHODS AND FINDINGS TO DATE

## 4.1 Survey on Current Practices and Challenges (RQ1)

Our prior work investigated current practices and challenges in conducting usability testing analysis through an online survey study with 279 UX professionals from six continents with different levels of UX experience, which was published at CHI 22 [26]. We asked about resources and collaborations leveraged during analysis, as well as desired features of new tools to better support their practices. Most respondents encountered challenges related to lack of time for analysis (66%) and difficulty merging analysis from multiple evaluators (70%). Our results showed that UX evaluators needed an integrated platform for session review and annotation, and that they needed support to relieve them from manual and time-consuming aspects of data analysis. When asked about their willingness to use tools that involved AI—due to the increasing trend of incorporating AI into the UX field [47, 49]—to support their analysis, 73% of the respondents were open to AI assistance to determine if a usability problem had occurred. This finding indicates that AI assistance has the opportunity to be widely adopted by UX evaluators.

## 4.2 Collaborative AI-Assisted UX Analysis Tool (RQ2)

In response to the interest in AI and the need for an integrated platform from our survey results [26], we built a collaborative visual analytics tool, *CoUX*, which seamlessly supports usability problem identification, annotation, and discussion in an integrated environment [41]. To ease the discovery of usability problems, CoUX visualizes a set of problem-indicators based on acoustic, textual, and visual features as shown in Fig 1. These features were automatically extracted from the video and audio of a think-aloud session using machine learning techniques. The design of CoUX was informed by a formative study with two UX experts and on insights derived from the literature. We conducted a user study with six pairs of UX evaluators on collaborative usability recording analysis tasks. The results indicated that CoUX was useful and effective in facilitating problem identification and collaborative teamwork. Then, we drew insights from participant survey and interview responses on how different ML-driven features were used to support independent analysis. For example, the acoustic, textual, and visual features were used by participants as hints, anticipation of problems, and anchors to revisit in the second-pass analysis. We found that participants trusted the suggested sentiment analysis and UX keywords because they could intuitively draw connections to usability problems. For relatively new features (e.g., speech rate, scrolling speed, and scene breaks), they did not understand the underlying algorithm of how these features were derived and were less reliant on these during analysis. Since the Feature Panel was non-interactive, participants did not have a way to express their hesitancy towards certain ML-driven features or ask for an explanation of the underlying algorithm. Based on this investigation and observed users' behavior, we consider that an interactive assistant—in the form of a conversational agent—may provide timely explanations for uncertain ML-driven suggestions.

## 4.3 Conversational Assistants for UX Evaluation (RQ3)

To address the limitations of non-interactive visualizations, we investigated how interactive AI assistants augment UX evaluator analytic productivity and trust. To build AI assistants that could respond to a full range of questions from UX evaluators, we must first understand what that full range might be. Thus, we conducted a design probe in which evaluators used an AI assistant with two modalities (voice and text) to ask any questions that they considered to be relevant to their analysis. Using the AI assistant as a probe, we investigated if the Q&A dynamic and modality of interaction improved evaluator efficiency and trust during analysis. As it is still challenging to leverage state-of-the-art AI algorithms to accurately detect usability problems and provide natural language responses [9], we adopted a Wizard of Oz approach to simulate the AI assistant so that we could focus on answering our research questions. Wizard of Oz has been commonly used to circumvent technical limitations in prior research (e.g., [31, 32, 40]), and offers an ability to explore the fidelity of the experience for participants, regardless of technical capability. Through a user study with 20 evaluators, we found that they mostly asked questions about user actions and mental model, as well as for suggestions from the AI assistant. We also found that participants felt text assistants were more efficient than voice assistants. Finally, we highlighted design considerations for improving future conversational AI assistants for UX analysis. The findings from this study are currently under submission for review.

# 5 EXPECTED NEXT STEPS

## 5.1 Proactive Conversational Assistants for UX Evaluation (RQ4)

The Q&A dynamic used in the design probe study for RQ3 was *reactive* (i.e., the CA only responded when prompted). In contrast, *proactive* dialogue initiated by the CA has been shown to provide adequate and timely assistance [25] and offer better support to older adults than passive approaches [39]. Although proactive interactions carry the risk of interruption, prior work has encouraged future research on how to improve the reception of such interactions by reducing their interruption costs and increasing their value [28]. Thus, we were motivated to explore whether a proactive approach may better support UX evaluators. For example, the AI assistant may monitor actions that UX evaluators take in the tool. If the evaluator pauses for an extended period of time or misses a usability problem, the AI assistant may initiate a conversation to ask if the evaluator would like to see which features were detected. We foresee that this active and ongoing monitoring allows the AI assistant to proactively make suggestions to aid UX evaluators. We will explore if UX evaluators would be willing to accept proactive suggestions and if so, how these suggestions should be presented. Again using the Wizard of Oz approach, we will conduct a user study with UX evaluators and collect interactions recordings to analyze the participants' usage of the AI assistant, including actions taken after each suggestion (e.g., ignored the suggestion, entered a

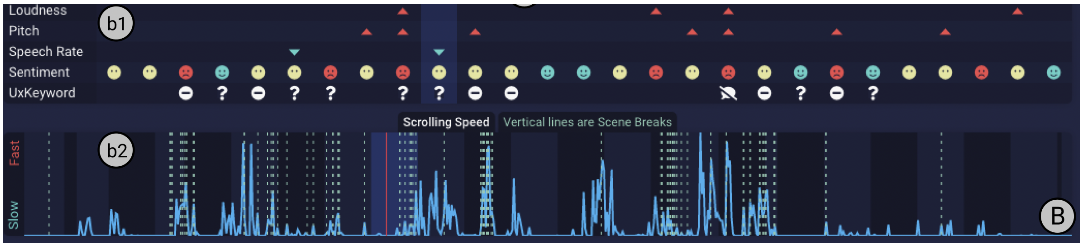


**Figure 1: (B) The CoUX Feature Panel which contains: (b1) various features that indicate usability problems; (b2) scrolling speed graph.**

usability problem), and surveys and interviews to glean feedback about the experience. Findings from this proposed work would indicate the desirability of proactive agents, and the timing of preferred proactive notifications.

### 5.2 Summative Study with AI Represented as both Visualizations and Conversational Assistants (RQ5)

Based on the results from prior studies (RQ2 and RQ3), we found that visualizations and interactive agents offered unique advantages. For example, objective information such as how many clicks the user made could be answered with little controversy, where as other information like redesign recommendations were typically subjective and required more judgment. Compared to subjective responses from the AI assistant, participants were more trusting of factual and objective information, which were also mundane and easier to miss. Participants in the design probe (RQ3) felt it would be helpful to see quantitative statistics common to all recordings as a summary, which would reduce the need to repeat the same questions for each recording. On the other hand, the conversational assistant offered many benefits, such as making some participants feel as if they were collaborating with a colleague with the chat window persisting as a record of their analysis. Thus, the advantages of both methods could be combined by providing UX evaluators with an overview of objective information in the form of a dashboard while having the opportunity to ask higher level subjective questions using the conversational interface. Informed by our findings on interactions with conversational AI assistants (RQ3, RQ4), we will design an integrated web-based tool that harnesses non-interactive visualizations for detectable information (RQ2) and conversational AI assistants for subjective suggestions. We will conduct a summative user study validating the effectiveness of our new tool for usability evaluation, which will involve asking participants to analyze a series of usability test recordings using the new tool.

Another open question in this step may be to compare interactive conversational interfaces against non-interactive visualizations as a baseline. Although we posit that the CA may provide more autonomy and timely information since visualizations always display information regardless of whether the evaluator needs it or not, we did not conduct a direct comparison of the two designs. There is ample room to change this project plan based on feedback and I think that this work will benefit tremendously from the suggestions of peers as well as experienced HCI researchers.

## 6 ANTICIPATED CONTRIBUTIONS

The rapid pace at which complex features and new products are being developed translates into an escalated need for more efficient UX testing and analysis. My dissertation research aims to elucidate factors contributing to human-AI collaborative work, particularly in the domain of usability analysis. This work will demonstrate the viability of human-AI collaboration, and the necessary technical and interactive elements that enhance such collaboration. Human-AI collaboration is the future of work and will proliferate, therefore, determining factors that lead to smooth integration and high performance is critical. We anticipate project outcomes in two key areas: (1) contributions to knowledge about how to design interactions with AI assistants for effective human-AI collaboration in the UX domain; (2) practical guidelines and web-based tools (involving text and voice assistants) to support UX evaluators in conducting analysis. These guidelines are also applicable to other domains of work where AI provides suggestions, since our findings will identify interactions and representations that give rise to productive and trusting collaborations with AI.

## 7 DISSERTATION STATUS AND LONG TERM GOALS

I am currently a third-year Ph.D. student in Computing and Information Sciences at Rochester Institute of Technology, advised by Dr. Kristen Shinohara. I am co-advised by Dr. Mingming Fan from the Hong Kong University of Science and Technology. I have completed all required coursework for the program (60 total semester credit hours). I successfully passed the Research Potential Assessment (RPA) at the end of my first year, which is intended to determine early in a student's academic life if they have the potential to successfully obtain a Ph.D. from the program. I aim to complete the proposed research projects and dissertation defense before Spring 2025. I have completed internships at industrial research labs during Summer 2021 and Summer 2022. Upon graduation, I intend to seek research positions in industry where I can continue investigating factors that lead to effective human-AI collaboration and advocating for the importance of UX research.

## ACKNOWLEDGMENTS

I am grateful to my advisors Professor Kristen Shinohara and Professor Mingming Fan, as well as my research collaborators for their mentorship and support. I also thank the respondents in my survey and participants in the user studies for taking the time to share their experiences and provide valuable feedback.

## REFERENCES


[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '19)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3290605.3300233

[2] Zahra Ashktorab, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2021. Effects of Communication Directionality and AI Agent Differences in Human-AI Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '21)*. Association for Computing Machinery, 1–15. https://doi.org/10.1145/3411764.3445256

[3] Zahra Ashktorab, Q. Vera Liao, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2020. Human-AI Collaboration in a Cooperative Game Setting: Measuring Social Perception and Outcomes. 4 (2020), 96:1–96:20. Issue CSCW2. https://doi.org/10.1145/3415167

[4] Pietro Battistoni, Marco Romano, Monica Sebillo, and Giuliana Vitiello. 2021. A Change in Perspective About Artificial Intelligence Interactive Systems Design: Human Centric, Yes, But Not Limited to. In *HCI International 2021 - Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence* (Cham) *(Lecture Notes in Computer Science)*, Constantine Stephanidis, Masaaki Kurosu, Jessie Y. C. Chen, Gino Fragomeni, Norbert Streitz, Shin'ichi Konomi, Helmut Degen, and Stavroula Ntoa (Eds.). Springer International Publishing, 381–390. https://doi.org/10.1007/978-3-030-90963-5_29

[5] Parmit K. Chilana, Jacob O. Wobbrock, and Andrew J. Ko. 2010. Understanding Usability Practices in Complex Domains. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 2337–2346. https://doi.org/10.1145/1753326.1753678

[6] Mingming Fan, Yue Li, and Khai N. Truong. 2020. Automatic Detection of Usability Problem Encounters in Think-Aloud Sessions. *ACM Transactions on Interactive Intelligent Systems* 10, 2 (June 2020), 1–24. https://doi.org/10.1145/3385732

[7] Mingming Fan, Serina Shi, and Khai N Truong. 2020. Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey. *Journal of Usability Studies* 15, 2 (2020), 85–102.

[8] Mingming Fan, Ke Wu, Jian Zhao, Yue Li, Winter Wei, and Khai N. Truong. 2020. VisTA: Integrating Machine Intelligence with Visualization to Support the Investigation of Think-Aloud Sessions. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (Jan. 2020), 343–352. https://doi.org/10.1109/TVCG.2019.2934797

[9] Mingming Fan, Xianyou Yang, TszTung Yu, Q. Vera Liao, and Jian Zhao. 2022. Human-AI Collaboration for UX Evaluation: Effects of Explanation and Synchronization. 6 (2022), 96:1–96:32. Issue CSCW1. https://doi.org/10.1145/3512943

[10] Asbjørn Følstad, Effie Lai-Chong Law, and Kasper Hornb. 2012. Analysis in Practical Usability Evaluation: A Survey Study. In *Proceedings of the 30th SIGCHI Conference on Human Factors in Computing Systems - CHI '12*. ACM Press, Austin, Texas, 2127–2136. https://doi.org/10.1145/2207676.2208365

[11] Asbjørn Følstad, Effie Lai-Chong Law, and Kasper Hornbæk. 2010. Analysis in Usability Evaluations: An Exploratory Study. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10)*. Association for Computing Machinery, New York, NY, USA, 647–650. https://doi.org/10.1145/1868914.1868995

[12] Asma Ghandeharioun, Daniel McDuff, Mary Czerwinski, and Kael Rowan. 2019. EMMA: An Emotion-Aware Wellbeing Chatbot. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 1–7. https://doi.org/10.1109/ACII.2019.8925455 ISSN: 2156-8111.

[13] Laurence Goasduff. 2019. *Chatbots Will Appeal To Modern Workers*. https://www.gartner.com/smarterwithgartner/chatbots-will-appeal-to-modern-workers

[14] Julián Grigera, Alejandra Garrido, José Matías Rivero, and Gustavo Rossi. 2017. Automatic detection of usability smells in web applications. *International Journal of Human-Computer Studies* 97 (2017), 129–148.

[15] Ted Grover, Kael Rowan, Jina Suh, Daniel McDuff, and Mary Czerwinski. 2020. Design and evaluation of intelligent agent prototypes for assistance with focus and productivity at work. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (New York, NY, USA) *(IUI '20)*. Association for Computing Machinery, 390–400. https://doi.org/10.1145/3377325.3377507

[16] Patrick Harms. 2019. Automated usability evaluation of virtual reality applications. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 3 (2019), 1–36.

[17] Morten Hertzum and Niels Ebbe Jacobsen. 2001. The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction* 15, 1 (2001), 183–204. https://doi.org/10.1207/S15327590IJHC1501_14

[18] Morten Hertzum, Rolf Molich, and Niels Ebbe Jacobsen. 2013. What You Get Is What You See: Revisiting the Evaluator Effect in Usability Tests. *Behaviour & Information Technology* 33, 2 (April 2013), 144–162. https://doi.org/10.1080/0144929X.2013.783114

[19] ISO. 2020. ISO 9241-110:2020 Ergonomics of Human-System Interaction. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/52/75258.html.

[20] Niels Jacobsen, Niels E, Morten Hertzum, Morten, Bonnie John, and Bonnie E. 1998. The Evaluator Effect in Usability Tests. https://doi.org/10.1145/286498.286737

[21] Sophie F. Jentzsch, Sviatlana Höhn, and Nico Hochgeschwender. 2019. Conversational Interfaces for Explainable AI: A Human-Centred Approach. In *Explainable, Transparent Autonomous Agents and Multi-Agent Systems* (Cham) *(Lecture Notes in Computer Science)*, Davide Calvaresi, Amro Najjar, Michael Schumacher, and Kary Främling (Eds.). Springer International Publishing, 77–92. https://doi.org/10.1007/978-3-030-30391-4_5

[22] JongWook Jeong, NeungHoe Kim, and Hoh Peter In. 2020. Detecting usability problems in mobile applications on the basis of dissimilarity in user behavior. *International Journal of Human-Computer Studies* 139 (2020), 102364.

[23] Jie Kang, Kyle Condiff, Shuo Chang, Joseph A. Konstan, Loren Terveen, and F. Maxwell Harper. 2017. Understanding How People Use Natural Language to Ask for Recommendations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (New York, NY, USA) *(RecSys '17)*. Association for Computing Machinery, 229–237. https://doi.org/10.1145/3109859.3109873

[24] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for Workplace Reflection: A Chat and Voice-Based Conversational Agent. In *Proceedings of the 2018 Designing Interactive Systems Conference* (New York, NY, USA) *(DIS '18)*. Association for Computing Machinery, 881–894. https://doi.org/10.1145/3196709.3196784

[25] Veronika Krauß, Alexander Boden, Leif Oppermann, and René Reiners. 2021. Current Practices, Challenges, and Design Implications for Collaborative AR/VR Application Development. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 454. Association for Computing Machinery, 1–15. https://doi.org/10.1145/3411764.3445335

[26] Emily Kuang, Xiaofu Jin, and Mingming Fan. 2022. "Merging Results Is No Easy Task": An International Survey Study of Collaborative Data Analysis Practices Among UX Practitioners. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. To Appear.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3491102.3517647

[27] Vivian Lai, Chacha Chen, Q. Vera Liao, Alison Smith-Renner, and Chenhao Tan. 2021. Towards a Science of Human-AI Decision Making: A Survey of Empirical Studies. arXiv:2112.11471 [cs] http://arxiv.org/abs/2112.11471

[28] Q. Vera Liao, Matthew Davis, Werner Geyer, Michael Muller, and N. Sadat Shami. 2016. What Can You Do? Studying Social-Agent Orientation and Agent Proactive Interactions with an Agent for Employees. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (New York, NY, USA) *(DIS '16)*. Association for Computing Machinery, 264–275. https://doi.org/10.1145/2901790.2901842

[29] Bingjie Liu. 2021. In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–AI Interaction. (2021). Issue zmab013. https://doi.org/10.1093/jcmc/zmab013

[30] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '16)*. Association for Computing Machinery, 5286–5297. https://doi.org/10.1145/2858036.2858288

[31] Michal Luria, Guy Hoffman, and Oren Zuckerman. 2017. Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '17)*. Association for Computing Machinery, 580–628. https://doi.org/10.1145/3025453.3025786

[32] Indrani Medhi Thies, Nandita Menon, Sneha Magapu, Manisha Subramony, and Jacki O'Neill. 2017. How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians. In *Human-Computer Interaction - INTERACT 2017* (Cham) *(Lecture Notes in Computer Science)*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer International Publishing, 441–459. https://doi.org/10.1007/978-3-319-67744-6_28

[33] Geoff Musick, Thomas A. O'Neill, Beau G. Schelble, Nathan J. McNeese, and Jonn B. Henke. 2021. What Happens When Humans Believe Their Teammate is an AI? An Investigation into Humans Teaming with Autonomy. 122 (2021), 106852. https://doi.org/10.1016/j.chb.2021.106852

[34] Christine M. Neuwirth, Ravinder Chandhok, David Charney, Patricia Wojahn, and Loel Kim. 1994. Distributed collaborative writing: a comparison of spoken and written modalities for reviewing and revising documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '94)*. Association for Computing Machinery, 51–57. https://doi.org/10.1145/191666.191693

[35] Jakob Nielsen. 2012. Usability 101: Introduction to Usability. https://www.nngroup.com/articles/usability-101-introduction-to-usability/.

[36] Mie Nørgaard and Kasper Hornbæk. 2006. What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing. In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)*. Association for Computing Machinery, New York, NY, USA, 209–218. https://doi.org/10.1145/1142405.1142439

[37] Asil Oztekin, Dursun Delen, Ali Turkyilmaz, and Selim Zaim. 2013. A Machine Learning-Based Usability Evaluation Method for eLearning Systems. *Decision Support Systems* 56 (Dec. 2013), 63–73. https://doi.org/10.1016/j.dss.2013.05.003

[38] Fabio Paternò, Antonio Giovanni Schiavone, and Antonio Conti. 2017. Customizable automatic detection of bad usability smells in mobile accessed web applications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–11.

[39] Lazlo Ring, Barbara Barry, Kathleen Totzke, and Timothy Bickmore. 2013. Addressing Loneliness and Isolation in Older Adults: Proactive Affective Agents Provide Better Support. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 61–66. https://doi.org/10.1109/ACII.2013.17 ISSN: 2156-8111.

[40] Ameneh Shamekhi, Q. Vera Liao, Dakuo Wang, Rachel K. E. Bellamy, and Thomas Erickson. 2018. Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '18)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3173574.3173965

[41] Ehsan Jahangirzadeh Soure, Emily Kuang, Mingming Fan, and Jian Zhao. 2021. CoUX: Collaborative Visual Analysis of Think-Aloud Usability Test Videos for Digital Interfaces. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–11. https://doi.org/10.1109/TVCG.2021.3114822

[42] Usability.gov. 2014. *User Experience Basics*. https://www.usability.gov/what-and-why/user-experience.html Publisher: Department of Health and Human Services.

[43] UserTesting. 2021. UserTesting: The Human Insight Platform. https://www.usertesting.com/.

[44] UXTesting. 2022. *UXTesting: The Best and Most Intelligent User Experience Solution for Enterprises*. https://www.uxtesting.io/user-experience-program

[45] H James Wilson and Paul R Daugherty. 2018. Collaborative Intelligence: Humans and AI Are Joining Forces. (2018), 11.

[46] Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 585–596. https://doi.org/10.1145/3196709.3196730

[47] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301

[48] Yang Yang, Xiaojuan Ma, and Pascale Fung. 2017. Perceived Emotional Intelligence in Virtual Agents. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA) *(CHI EA '17)*. Association for Computing Machinery, 2255–2262. https://doi.org/10.1145/3027063.3053163

[49] John Zimmerman, Changhoon Oh, Nur Yildirim, Alex Kass, Teresa Tung, and Jodi Forlizzi. 2020. UX Designers Pushing AI in the Enterprise: A Case for Adaptive UIs. *Interactions* 28, 1 (Dec. 2020), 72–77. https://doi.org/10.1145/3436954