BEHAVE AI: BEst Practices and Guidelines for Human-Centric Design and EvAluation of ProactiVE AI Agents

Matthias Kraus
University of Augsburg & LMU
Munich
Augsburg, Germany
matthias.kraus@uni-a.de

Sebastian Zepf
Mercedes-Benz AG
Böblingen, Germany
sebastian.zepf@mercedes-benz.com

Rebecca Westhäußer
Mercedes-Benz AG
Böblingen, Germany
rebecca.westhaeusser@mercedes-benz.com

Isabel Feustel
Ulm University
Ulm, Germany
isabel.feustel@uni-ulm.de

Nima Zargham
University of Bremen, Digital Media
Lab
Bremen, Germany
zargham@uni-bremen.de

Ilhan Aslan
Aalborg University
Aalborg, Denmark
ilas@cs.aau.dk

Justin Edwards
University of Oulu
Oulu, Finland
justin.edwards@oulu.fi

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

Dimosthenis Kontogiorgos
Massachusetts Institute of Technology
Cambridge, MA, USA
dkonto@mit.edu

Nicolas Wagner
Otto-Friedrich-University Bamberg
Bamberg, Germany
nicolas.wagner@uni-bamberg.de

Elisabeth André
University of Augsburg
Augsburg, Germany
elisabeth.andre@uni-a.de

Abstract

The last years have seen a significant rise in interest in highly autonomous and proactive agents fueled by the progress of AI. While there is various research on the design, implementation, and evaluation of proactive agents, there remains a critical gap in the methodologies used for both design and evaluation, which are largely informed by reactive system design principles. Our full-day multidisciplinary workshop brings together researchers and practitioners from the IUI community in academia and industry to understand the challenges of designing and evaluating proactive agents in a human-centric manner. We will reflect on existing evaluation methods, identify challenges in designing proactive systems, and discuss potential solutions, best practices, and human-centric guidelines to bridge these gaps. Ultimately, our goal is to map out key focus areas and research challenges, fostering strong interdisciplinary relationships within and across fields related to Artificial Intelligence (AI) and Human-Computer Interaction (HCI).

CCS Concepts

• **Human-centered computing** → **HCI design and evaluation methods**; **Interaction design**; *Natural language interfaces*.

Keywords

Proactive Agents, Conversational Agents, Voice Assistants, Proactive AI

1 Introduction

In recent years, there has been a surge of interest in the development of highly autonomous and proactive agents, driven by advancements in AI. The rapid emergence of large language models (LLMs) and generative AI (GenAI) has led to a new era of HCI, with tools like the GitHub Copilot and Google's Gemini exemplifying the early stages of proactive AI in commercial use. These systems, capable of anticipating user needs and autonomously initiating actions, present both exciting opportunities and significant challenges [3, 19, 27]. While the potential of proactive agents is vast – ranging from increasing perceived helpfulness to improving task efficiency and calibrating trust [13, 20, 24] – there remains a critical gap in established guidelines and best practices for their human-centric design and evaluation.

Current evaluation methods, primarily designed for reactive systems, fall short of addressing the unique demands and complexities of interactions with proactive AI. For example, tools such as the

System Usability Scale (SUS) [4] or the *User Experience Questionnaire* (UEQ) [23], often used to measure the perceived usability of proactive agents, have been validated only in the context of reactive systems. They do not account for the nuances of proactive interactions, which involve not just user responses but also autonomous decision-making and action execution by the AI. Similarly, the measurement of social implications, such as perceived human-computer trust, relies on scales that refer to reactive GUIs and may, therefore, be of limited applicability [17].

The design of proactive AI also poses unique challenges. Meurisch et al. [18] pointed out that expectations regarding proactive AI behavior highly depend on the application area and user characteristics. Users generally prefer to maintain control, expecting initially low levels of system proactivity that may increase as they become more familiar with the system, e.g., see Glass et al. [6]. Some initial principles for designing proactive behavior have been suggested by Yorke-Smith et al. [25], drawing from the mixed-initiative principles by Horvitz [8]. These principles emphasize that a proactive agent should be valuable, pertinent, competent, controllable, transparent, deferent, unobtrusive, anticipatory, and safe. Hammer et al. [7] presented a Bayesian Network approach for deciding on proactive system actions based on a user trust model, taking into account principles, such as level of control and transparency. Yet, the guidelines for the design of human-AI interaction, such as those by Amershi et al. [1], only lightly touch on proactive behavior, underscoring the need for more comprehensive and standardized approaches. In summary, a deficiency of standardization in terms of design and evaluation methods can be observed in this domain [3].

Therefore, this full-day multidisciplinary workshop aims to bring together members of the Intelligent User Interfaces (IUI) community to critically examine the challenges of human-centric design and evaluation of proactive AI agents. By uniting researchers from industry and a wide array of academic disciplines, we seek to explore best practices and guidelines for designing and evaluating proactive AI and autonomous agents. Through collaborative discussions and knowledge sharing, our goal is to foster the development of a cohesive, diverse community dedicated to advancing the field of proactive HCI and its human-centric evaluation.

2 Previous History

In May 2024, the Shonan Seminar, a Dagstuhl-style seminar held in Japan and managed by the National Institute of Informatics (NII) Japan, was dedicated to the topic "Intelligent Interaction with Autonomous Assistants in the Wild." Among the manifold research challenges identified, the lack of standardized methods and metrics for evaluating autonomous and proactive agents has emerged as one of the most relevant gaps that need to be addressed to advance the research field of interaction with autonomous agents. As a result, a subgroup consisting of both organizers and participants of the Shonan Seminar was formed, and they would like to elaborate on this topic in more depth in a dedicated workshop. While there have been related events in the domain of proactive systems, such as tutorials on proactive dialogue design at ACM's WSDM and SIGIR [15, 16], to the best of our knowledge, there have been no prior workshops that have focused on human-centric design and evaluation of proactive and autonomous agents.

3 Organizers and Committee

The key organizers are listed as authors of this proposal. Workshop organizers include researchers from both academia and industry who have a strong background, expertise, and interest in research and business in AI Assistants covering various application domains such as automotive [28, 29], health and well-being [2, 22], smart home [12, 21, 26], robotics [9, 10, 14], and learning [5, 11]. We believe that the exchange driven by this interdisciplinary team can help advance the field of interaction with proactive, autonomous agents. The organizers are as follows:

- **Matthias Kraus** is currently an interim professor at LMU Munich and Post-Doc at the Chair of Human-Centered Artificial Intelligence at Augsburg University. He previously co-organized a workshop at ACM CHI 2022.
- **Sebastian Zepf** is an HCI Researcher at Mercedes-Benz AG, working on designing and developing AI assistants that anticipate user needs and proactively initiate actions for and interactions with users whenever suitable.
- **Rebecca Westhäußer** is a PhD student at Mercedes-Benz AG, focusing on research related to the interaction with proactive agents.
- **Isabel Feustel** is a PhD student in the Dialogue Systems group at the University of Ulm.
- **Nima Zargham** is a postdoctoral researcher in the Digital Media Lab at the University of Bremen. His research focuses on human-centered approaches for designing speech-based systems that elicit desirable user experiences. He previously organized CUI-related workshops at notable conferences such as ACM/IEEE HRI 2023, ACM CUI 2023-24, and ACM CHI 2024.
- **Ilhan Aslan** is an associate professor and member of the Human-Centered Computing group at Aalborg University, Denmark. He previously co-organized workshops at CHI, nordiCHI, mobile HCI, and TEI.
- **Justin Edwards** is a postdoctoral researcher in the Learning and Educational Technology (LET) research lab at the University of Oulu. He previously co-organized the CUI@-workshop series with workshops at CHI 2019, IUI 2020, and Automotive UI 2020 as well as the EduCUI workshop at CUI 2024.
- **Sven Mayer** (https://sven-mayer.com) is an assistant professor of computer science at LMU Munich, Germany. His research sits at the intersection between Human-Computer Interaction and Artificial Intelligence, where he focuses on the next generation of computing systems. He has served as a program committee member at numerous conferences, e.g., ACM CHI, and in various organizing committees, e.g., as General Chair for the International Conference on Hybrid Human-Artificial Intelligence (HHAI'23).
- **Dimosthenis Kontogiorgos** is a PostDoctoral Researcher at the Massachusetts Institute of Technology, and his research explores how to develop explainable multimodal AI technologies for collaborative interactive robots. He has previously co-organized the Robo-Identity series of workshops at HRI 2021-2024, as well as workshops for Interspeech 2017, SigDial 2019, ICMI 2020, and CUI 2023. He has also been

involved in the organising committee of ICMI 2024 and in AC editorial roles for ROMAN 2024, ACII 2024, CHI 2025, RA-L, and THRI.

- **Nicolas Wagner** is a researcher in the Natural Language Generation and Dialogue Systems group at the University of Bamberg.
- **Elisabeth André** is a full professor of Computer Science and Founding Chair of Human-Centered Artificial Intelligence at Augsburg University, Germany. She has served as a General and Program Co-Chair of top-tier ACM SIGCHI conferences, such as ACM IUI in 2003 and 2011 and ACM ICMI in 2016 and 2023.

4 Participants

Our workshop aims to bring together 15-20 world-leading researchers and industry representatives from a broad range of communities related to human-machine interaction, multimodal interaction, conversational AI, and human-centered AI. We aim to highlight the importance of a multidisciplinary approach to solving issues related to the design and evaluation of proactive autonomous agents, broadening the reach of proactive agent research at IUI and within HCI in general to other research communities (e.g., speech technology, linguistics, dialogue research, cognitive sciences). Further, we seek to build a collaborative, diverse, and cross-disciplinary conversational interaction community strongly connected to IUI.

5 Workshop Format

We follow a mini-conference model inviting submissions of short papers of max. 8 pages in length (excluding references) following the CEUR format. The papers have been independently reviewed by two reviewers. We have invited participants who wish to join the workshop without submitting a paper to submit a short abstract or position statement. We believe that this will help stimulate the diversity of participants by attracting researchers from industry with relevant experience in the field as well as researchers from related fields who also deal with human-centered evaluation. The workshop's interesting themes are evaluation metrics, questionnaires, study design, and existing guidelines. The plan is to have a half-day workshop, following the structure outlined:

- Introduction (15 Minutes)
- Keynote by Benjamin R Cowan (University College Dublin) (30 Minutes)
- Paper Session (60 minutes)
- Roundtable Discussion (60 minutes)
- Breakout: Streamlining Promising Existing Methods and Challenges (60 minutes)
- Summary and Debriefing (30 minutes)

6 Planned Outcomes

Following the workshop, our aims are to:

(1) Propose key research directions and identify potential collaborations to drive the standardization of proactive system evaluation and the establishment design guidelines.
(2) Organize further workshops on topics relevant to the establishment of guidelines for the human-centric evaluation

of proactive AI agents. This will include targeting INTER-SPEECH, SIGDIAL, CogSci, AutoUI, CHI, CUI, and IUI conferences, as well as identifying further collaborative opportunities across disciplines.
(3) Connect and foster the community of research and development between IUI researchers and industry partners.

Acknowledgments

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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233
[2] Ilhan Aslan, Dominik Neu, Daniela Neupert, Stefan Grafberger, Nico Weise, Pascal Pfeil, and Maximilian Kuschewski. 2023. How to Compliment a Human-Designing Affective and Well-being Promoting Conversational Things. *Interaction Design and Architecture (s) Journal* 58 (2023), 157–184. https://doi.org/10.55612/s-5002-058-007
[3] Caterina Bérubé, Marcia Nißen, Rasita Vinay, Alexa Geiger, Tobias Budig, Aashish Bhandari, Catherine Rachel Pe Benito, Nathan Ibarcena, Olivia Pistolese, Pan Li, et al. 2024. Proactive behavior in voice assistants: A systematic review and conceptual model. *Computers in Human Behavior Reports* (2024), 100411.
[4] John Brooke. 2013. SUS: a retrospective. *Journal of usability studies* 8, 2 (2013).
[5] Justin Edwards, Andy Nguyen, Marta Sobocinski, Joni Lämsä, Adelson de Araujo, Belle Dang, Ridwan Whitehead, Anni-Sofia Roberts, Matti Kaarlela, and Sanna Jarvela. 2024. MAI - A Proactive Speech Agent for Metacognitive Mediation in Collaborative Learning. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 46, 5 pages. https://doi.org/10.1145/3640794.3665585
[6] Alyssa Glass, Deborah L McGuinness, and Michael Wolverton. 2008. Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces*. 227–236.
[7] Stephan Hammer, Michael Wißner, and Elisabeth André. 2015. Trust-based decision-making for smart and adaptive environments. *User Model. User Adapt. Interact.* 25, 3 (2015), 267–293. https://doi.org/10.1007/S11257-015-9160-8
[8] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 159–166. https://doi.org/10.1145/302979.303030
[9] Dimosthenis Kontogiorgos, Andre Pereira, Boran Sahindal, Sanne van Waveren, and Joakim Gustafson. 2020. Behavioural Responses to Robot Conversational Failures. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 53–62. https://doi.org/10.1145/3319502.3374782
[10] Dimosthenis Kontogiorgos, Sanne van Waveren, Olle Wallberg, Andre Pereira, Iolanda Leite, and Joakim Gustafson. 2020. Embodiment Effects in Interactions with Failing Robots. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376372
[11] Matthias Kraus, Diana Betancourt, and Wolfgang Minker. 2023. Does it affect you? Social and learning implications of using cognitive-affective state recognition for proactive human-robot tutoring. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 928–935.
[12] Matthias Kraus, Nicolas Wagner, Wolfgang Minker, Ankita Agrawal, Artur Schmidt, Pranav Krishna Prasad, and Wolfgang Ertel. 2022. KURT: A household assistance robot capable of proactive dialogue. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 855–859.
[13] Matthias Kraus, Nicolas Wagner, Ron Riekenbrauck, and Wolfgang Minker. 2023. Improving Proactive Dialog Agents Using Socially-Aware Reinforcement Learning. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (Limassol, Cyprus) *(UMAP '23)*. Association for Computing Machinery, New York, NY, USA, 146–155. https://doi.org/10.1145/3565472.3595611
[14] Jan Leusmann, Chao Wang, and Sven Mayer. 2024. Comparing Rule-based and LLM-based Methods to Enable Active Robot Assistant Conversations. In *Proceedings of the CUI@CHI 2024: Building Trust in CUIs –From Design to Deployment*

(2024-05-11). CUI, 1–5. https://sven-mayer.com/wp-content/uploads/2024/06/leusmann2024comparing.pdf

[15] Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive Conversational Agents in the Post-ChatGPT World. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 3452–3455. https://doi.org/10.1145/3539618.3594250

[16] Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive Conversational Agents in the Post-ChatGPT World. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 3452–3455. https://doi.org/10.1145/3539618.3594250

[17] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, Vol. 53. Citeseer, 6–8.

[18] Christian Meurisch, Cristina A. Mihale-Wilson, Adrian Hawlitschek, Florian Giger, Florian Müller, Oliver Hinz, and Max Mühlhäuser. 2020. Exploring User Expectations of Proactive AI Systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 146 (dec 2020), 22 pages. https://doi.org/10.1145/3432193

[19] Carl Oechsner, Sven Mayer, and Andreas Butz. 2022. Challenges and Opportunities of Cooperative Robots as Cooking Appliances. In *Proceedings of the 2022 Workshop on Engaging with Automation* (2022-04-30) *(AutomationXP'22)*. https://ceur-ws.org/Vol-3154/paper14.pdf

[20] Zhenhui Peng, Yunhwan Kwon, Jiaan Lu, Ziming Wu, and Xiaojuan Ma. 2019. Design and Evaluation of Service Robot's Proactivity in Decision-Making Support Process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300328

[21] Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. May I interrupt? Diverging opinions on proactive smart speakers. In *Proceedings of the 3rd conference on conversational user interfaces*. 1–10.

[22] Matthias Schmidmaier, Cedrik Harrich, and Sven Mayer. 2024. Increasing Large Language Models Context Awareness through Nonverbal Cues. In *Proceedings of Mensch und Computer 2024 –Workshop* (2024-09-01) *(MuC Workshoṕ24)*. GI, Bonn, Germany. https://doi.org/10.18420/muc2024-mci-ws09-190

[23] Martin Schrepp, Andreas Hinderks, and Joörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40.

[24] Maximiliane Windl, Jan Leusmann, Albrecht Schmidt, Sebastian S. Feger, and Sven Mayer. 2024. Privacy Communication Patterns for Domestic Robots. In *Tenth Symposium on Usable Privacy and Security (SOUPS 2024)* (2024-08-11). USENIX Association, Philadelphia, PA, USA. https://www.usenix.org/conference/soups2024/presentation/windl

[25] Neil Yorke-Smith, Shahin Saadati, Karen L Myers, and David N Morley. 2012. The design of a proactive personal agent for task management. *International Journal on Artificial Intelligence Tools* 21, 01 (2012), 1250004.

[26] Nima Zargham, Dmitry Alexandrovsky, Jan Erich, Nina Wenig, and Rainer Malaka. 2022. "I Want It That Way": Exploring Users' Customization and Personalization Preferences for Home Assistants. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 270, 8 pages. https://doi.org/10.1145/3491101.3519843

[27] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. Understanding circumstances for desirable proactive behaviour of voice assistants: The proactivity dilemma. In *Proceedings of the 4th conference on conversational user interfaces*. 1–14.

[28] Sebastian Zepf, Neska El Haouij, Jinmo Lee, Asma Ghandeharioun, Javier Hernandez, and Rosalind W. Picard. 2020. Studying Personalized Just-in-time Auditory Breathing Guides and Potential Safety Implications during Simulated Driving. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 275–283. https://doi.org/10.1145/3340631.3394854

[29] Sebastian Zepf, Arijit Gupta, Jan-Peter Krämer, and Wolfgang Minker. 2020. EmpathicSDS: Investigating Lexical and Acoustic Mimicry to Improve Perceived Empathy in Speech Dialogue Systems. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 2, 9 pages. https://doi.org/10.1145/3405755.3406125