Emerging Conversational Interactions for Well-Being

Wester, Joel

DOI (link to publication from Publisher):
10.54337/aau806911234

Publication date:
2025

Document Version
Publisher's PDF, also known as Version of record

EMERGING CONVERSATIONAL INTERACTIONS FOR WELL-BEING

BY
JOEL WESTER

PhD Thesis 2025

AALBORG
UNIVERSITY

EMERGING CONVERSATIONAL INTERACTIONS FOR WELL-BEING

Ph.D. Thesis 2025

Joel Wester

Aalborg University
Department of Computer Science
Selma Lagerløfs Vej 300
DK-9220 Aalborg

Submitted: August 2025

Main Supervisor: Professor Niels van Berkel
Aalborg University, Denmark

Co-supervisor: Assistant Professor Henning Pohl
Aalborg University, Denmark

Assessment: Associate Professor John Stouby Persson (chair)
Aalborg University, Denmark

Professor Benjamin Cowan
University College Dublin, Ireland

Associate Professor Luigi De Russis
Politecnico di Torino, Italy

Department: Department of Computer Science

PhD Series: Technical Faculty of IT and Design, Aalborg University

ISSN: 2446-1628
ISBN: 97887-7642-127-4

Published by: Aalborg University Open Publishing | www.open.aau.dk

Printed in Denmark by Stibo Complete, 2025

ENGLISH SUMMARY

The rapid development of AI-powered conversational user interfaces (CUIs) in recent years has shifted interactions with technology toward conversational experiences. As a result, many new research and industry applications as well as research tools now focus on conversations, opening up new opportunities and challenges—particularly when people use these conversational interfaces for well-being and consequently how the interfaces respond to them in this context. In this PhD project, I investigated how people perceive these emerging conversational interactions across well-being contexts.

I conducted a series of online and in-person studies using both qualitative and quantitative methods to assess user perceptions. The thesis consists of five research papers. Papers one and two explored how people perceive conversational self-care technologies. Paper three examined how these systems can mediate health information between individuals. Paper four investigated the broader uptake of these technologies and how well-being experts view their potential to support people caring for themselves. Finally, paper five explored how to design CUIs for situations where they cannot, or should not, fulfil user requests.

People now interact with CUIs in everyday settings: at home, in schools, or even on the bus. Through my research, I have identified how people perceive and engage with CUIs, and how we might design and adapt them more appropriately. I found that the way emerging conversational systems respond to user requests influences key interpersonal dynamics such as trust, likeability, and a sense of relatedness. This effect also holds when systems mediate information between individuals. Importantly, in situations where systems must deny user requests, how they do so significantly shapes user perceptions.

Based on the findings from the individual papers, I formulated multiple design takeaways. These design takeaways form the basis for two overarching design considerations relevant to emerging conversational interactions. First, I emphasise the importance of calibrating user agency based on conversational context. For example, sensitive settings may require that the user's agency be restricted. Second, complementing the calibration of user agency, I lay out the relevance of adapting CUIs' expressive range, such as their friendliness towards users. These design considerations together highlight the need to go beyond default AI models that power CUIs by carefully adjusting system parameters to better fit the target audience and domain.

This Ph.D. project was made possible under the project Algorithmic Explainability For Everyday Citizens through the Carlsberg Foundation, grant CF21-0159.

DANSK RESUMÉ

Den hurtige udvikling af AI-drevne samtalebaserede brugergrænseflader (Conversational User Interfaces, CUIs) i de seneste år har flyttet interaktioner med teknologi i retning af samtalebaserede oplevelser. Som følge heraf fokuserer mange nye forsknings- og industriprojekter, samt teknologier til at udføre forskning, nu på samtaler, hvilket åbner op for nye muligheder og udfordringer. Dette er især gældende når mennesker bruger disse samtalebaserede grænseflader til at understøtte deres trivsel, og hvordan grænsefladerne tilsvarende reagerer på brugerne. I dette ph.d.-projekt har jeg undersøgt, hvordan mennesker opfatter disse nye konversationelle interaktioner i konteksten omkring trivsel.

Jeg har gennemført en række online- og fysiske studier ved hjælp af både kvalitative og kvantitative metoder for at vurdere brugernes opfattelser. Afhandlingen består af fem forskningsartikler. Artikel ét og to undersøgte, hvordan mennesker opfatter konversationelle selvhjælpsteknologier. Artikel tre undersøgte, hvordan disse systemer kan formidle sundhedsinformation mellem individer. Artikel fire undersøgte den bredere udbredelse af disse teknologier og hvordan trivselseksperter vurderer deres potentiale til at støtte mennesker i at tage vare på sig selv. Endelig undersøgte artikel fem, hvordan man kan designe CUIs til situationer, hvor de ikke kan eller bør opfylde brugerens forespørgsler.

Mennesker interagerer nu med CUI-systemer i hverdagsmiljøer: derhjemme, i skoler eller endda i bussen. Gennem min forskning har jeg identificeret, hvordan mennesker opfatter og interagere med CUIs, og hvordan vi kan designe og tilpasse dem mere hensigtsmæssigt. Jeg fandt, at den måde, nye samtale-baserede systemer reagerer på brugerens forespørgsler, påvirker centrale interpersonelle dynamikker såsom tillid, sympati og en følelse af samhørighed. Denne effekt gør sig også gældende, når systemer formidler information mellem individer. Vigtigt er det, at i situationer, hvor systemer må afvise brugerens forespørgsler, har måden, de gør dette på, en betydelig indflydelse på brugerens opfattelse.

På baggrund af resultaterne fra de enkelte artikler har jeg formuleret flere designanbefalinger. Disse designanbefalinger danner grundlag for to overordnede designovervejelser, der er relevante for samtale-baserede understøttelse af trivsel. For det første fremhæver jeg vigtigheden af at kalibrere brugerens handlefrihed, eksempelvis at denne bør begrænses, når det er hensigtsmæssigt, særligt i følsomme kontekster. For det andet, som supplement til kalibreringen af brugerens handlefrihed, fremhæver jeg relevansen af at fastsætte CUIs' udtryksmæssige spændvidde, eksempelvis ved at tilpasse dens venlighed over for brugeren. Samlet set understreger disse designovervejelser behovet for at gå ud over standardindstillingerne i de AI-modeller, der driver CUIs, ved omhyggeligt at justere parametre, så de passer bedre til målgruppen og domænet.

Dette ph.d.-projekt blev muliggjort under projektet Algorithmic Explainability For Everyday Citizens gennem Carlsbergfondet, bevilling CF21-0159.

ACKNOWLEDGEMENTS

First, I want to express my gratitude to my supervisors, Niels van Berkel and Henning Pohl. Niels, thanks for everything. For the GMs, coffee walks, and rage-edits, to mention a few. I would not be where I am today without your steady support and encouragement, both in and outside the office, and for that, I am forever grateful (respect). Henning, thank you for the countless times you pulled up a chair in my office to chat, whether it was about discussing research ideas, statistics, or how to define a sandwich. I have learned a great deal from you.

Second, I want to thank the researchers across the globe that I have had the privilege to cross paths with. Minha Lee, for awesome chats and fruitful collaborations together with Rucha. Katie Winkle, for hosting me in Uppsala and for all the work including the brainstorming-sessions that eventually led to the publication of one of the papers included in this thesis. Enrico Costanza, for hosting me at UCLIC and the great discussions, and the excellent dinner in the Walworth area. Qiaosi Wang, for organising cool workshops and involving me in this process. Yue Fu, for great chats, collaborations, and even visiting Aalborg for a bit.

Third, I want to thank all my colleagues at HCC. Special thanks to Martin, Eike, and Julien, for being my first office mates and helping out with research and life. Shagen, for bringing me along to stuff and introducing me to your friends. David, for all the excellent stories. Thanks to A-team for being awesome: Sander and Naja, with whom I have shared this PhD adventure; Sam, for all the tea motivators; Helena, for being a great office mate (huvet upp och fötterna ner). Last, but the opposite of least, I want to thank Rune for being a good friend, both in and outside the værtshus.

Fourth, I want to thank all those who supported my move from Stockholm to Aalborg (what were you thinking). Thanks to my friends, Axel, Christian, Johan, Patrik, Alexander, Niklas, who have supported me and even payed me a visit and got to see AaB play. To my brother Tom and wife Elin, and nieces Isa, Tilda, and Frida. I miss you and I look forward to being closer to you soon. To Mom (Karin) and Dad (Rune), I doubt anyone could ask for more loving and supportive parents—I would not be where I am today without you. Lastly, to Josefine, for being the best person in the world. I love you.

THESIS DETAILS

Dissertation Title: Emerging Conversational Interactions for Well-Being
Ph.D. Student: Joel Wester
Supervisors: Prof. Niels van Berkel, Aalborg University
Asst. Prof. Henning Pohl, Aalborg University

The main body of this thesis consists of the following papers:

1. **Joel Wester**, Henning Pohl, Simo Hosio, Niels van Berkel. *"This Chatbot Would Never…": Perceived Moral Agency of Mental Health Chatbots.* In Proceedings of the 2024 ACM on Human-Computer Interaction (CSCW1). https://doi.org/10.1145/3637410

2. **Joel Wester**, Sander de Jong, Henning Pohl, Niels van Berkel. *Exploring People's Perceptions of LLM-generated Advice.* In Computers in Human Behavior: Artificial Humans, 2024. https://doi.org/10.1016/j.chbah.2024.100072

3. **Joel Wester**, Bhakti Moghe, Katie Winkle, Niels van Berkel. *Facing LLMs: Robot Communication Styles in Mediating Health Information between Parents and Young Adults*. In Proceedings of the 2024 ACM on Human-Computer Interaction (CSCW2). https://doi.org/10.1145/3687036

4. **Joel Wester**, Sander de Jong, Henning Pohl, Niels van Berkel. *Using LLMs for Self-Care: User and Counsellor Perspectives.* International Journal of Human-Computer Studies, 2025. https://doi.org/10.1016/j.ijhcs.2025.103589

5. **Joel Wester**, Tim Schrills, Henning Pohl, Niels van Berkel. *"As an AI language model, I cannot": Investigating LLM Denials of User Requests.* In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI'24). https://doi.org/10.1145/3613904.3642135

First- and co-authored papers not included in the thesis:

1. Rucha Khot*, Teis Arets*, **Joel Wester***, Franziska Burger, Niels van Berkel, Rens Brankaert, Wijnand Ijsselsteijn, Minha Lee. *Projecting Challenging Futures: Using Chatbots to Reflect on Aging and Dementia.* In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI'25). https://doi.org/10.1145/3706598.3713727

2. Qiaosi Wang, **Joel Wester**, Marvin Pafla, Minha Lee, Justin D. Weisz, Mei Si. *ToMinHAI at CUI'2025: Theory of Mind in Human-CUI Interaction.* Proceedings of the International Conference on Conversational User Interfaces (CUI'25 EA). https://doi.org/10.1145/3719160.3728627

3. Mana Samiee, **Joel Wester**, Rune Møberg Jacobsen, Michael Skovdal Rathleff, Niels van Berkel. *General Practicioners' Perspectives on a Pre-Consultation Chatbot for Shared Decision-Making.* In Proceedings of the 2025 Designing Interactive Systems (DIS'25). https://doi.org/10.1145/3715336.3735674

4. Martin Lindrup, Rune Møberg Jacobsen, **Joel Wester**, Niels van Berkel, Dimitrios Raptis, Peter Axel Nielsen. *Prompt Machine: A Tangible Generative AI Tool for Supporting Children's Learning and Literacy*. In Proceedings of the 2025 Designing Interactive Systems (DIS'25). https://doi.org/10.1145/3715336.3735673

5. Sander de Jong, Rune Møberg Jacobsen, **Joel Wester**, Senuri Wijenayake, Jorge Goncalves, Niels van Berkel. Impact of Agent-Generated Rationales on Online Social Conformity. In Proceedings of the 2025 Fairness, Accountability, and Transparency (FaccT'25). https://doi.org/10.1145/3715275.3732217

6. Rune Møberg Jacobsen, Martin Lindrup, **Joel Wester**, Niels van Berkel, Dimitrios Raptis, Peter Axel Nielsen. *Prompt Machine: Enabling Physical LLM-Interactions for Learning*. Proceedings of the ACM Conference on Designing Interactive Systems (DIS'25 EA). https://doi.org/10.1145/3715668.3735602

7. Naja Kathrine Kollerup, **Joel Wester**, Mikael B. Skov, Niels van Berkel. How Can I Signal You To Trust Me: Investigating AI Trust Signalling in Clinical Self-Assessments. In Proceedings of the 2024 Designing Interactive Systems (DIS'24). https://doi.org/10.1145/3643834.3661612

8. Sander de Jong, **Joel Wester**, Tim Schrills, Kristina S. Secher, Carla F. Griggio, Niels van Berkel. Assessing Cognitive and Social Awareness among Group Members in AI-assisted Collaboration. In Proceedings of the 2024 Mobile and Ubiquitous Media (MUM'24). https://doi.org/10.1145/3701571.3701582

9. **Joel Wester**, Minha Lee, Niels van Berkel. *Moral Transparency as a Mitigator of Moral Bias in Conversational User Interfaces*. Proceedings of the International Conference on Conversational User Interfaces (CUI'23 EA). https://doi.org/10.1145/3571884.3603752

10. **Joel Wester**, Andreas Brännström, Juan Carlos Nieves, Niels van Berkel "*You've Got a Friend in Me*": *A Formal Understanding of the Critical Friend Agent*. Proceedings of the International Conference on Human-Agent Interaction (HAI'23 EA). https://doi.org/10.1145/3623809.3623957

CONTENTS

1. INTRODUCTION

Since the mid-1960s, when the first chatbot Eliza (see Figure 1.1), a rule-based conversational interface, was created [118], the potential of chatbots to enable more dynamic interactions with computers has been evident. Historically, however, chatbots were limited in their capabilities, with their behaviour toward users being hardcoded and constrained. For example, if a user asked for a ticket refund, a rule-based chatbot might respond with one of three preset options: A, B, or C, with each outcome pre-determined by its developer. AI-based systems, however, could generate a much wider range of responses. Recently, large language models (LLMs) have triggered a paradigm shift in conversational interactions. Billion-parameter models, trained on vast amounts of data, now enable chatbots to dynamically engage with users on any topic (see Figure 1.2 for a screenshot of OpenAI's ChatGPT desktop version). These commercially available tools, including ChatGPT, Gemma, and DeepSeek, are used in contexts ranging from one-shot searches in search engines to conversational, iterative interactions, as well as in well-being chatbots that offer emotional support and opportunities for self-reflection through AI companions.

Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED

Fig. 1.1: A dialogue snippet from interacting with Eliza (capitalised) retrieved from Weizenbaum [118].

From a research perspective, there has been an explosion of work exploring how LLMs can be effectively used across various contexts, such as well-being, self-care, and personalised support. Recent system-focused papers published at venues like CHI and CSCW have examined LLM-powered tools for diary-writing, journaling, and other reflective practices (see examples in Figure 1.3). These systems often target distinct user groups, such as LLMs supporting mental health professionals working with suicidal individuals [107], or helping patients prepare for visits to general practitioners [70]. With ChatGPT reaching millions of users, its widespread use among everyday users across the globe is evident.

While the adoption and use of LLMs are rapidly expanding across various domains, their integration into real-world well-being contexts at scale remains in its early stages. A growing body of reports highlights issues with LLMs being unreliable and unpredictable [94], or socially inappropriate in their responses to users (e.g., stigmatising towards mental health conditions [79]). Regarding unreliability, HCI research investigates how LLMs can behave appropriately towards users (e.g., when used by patients as pre-consultation chatbots [70]). In terms of unpredictability, these models may generate responses that users do not anticipate, leading to interaction breakdowns. As for appropriateness, LLMs can produce outputs that fail to meet users' expectations or violate social

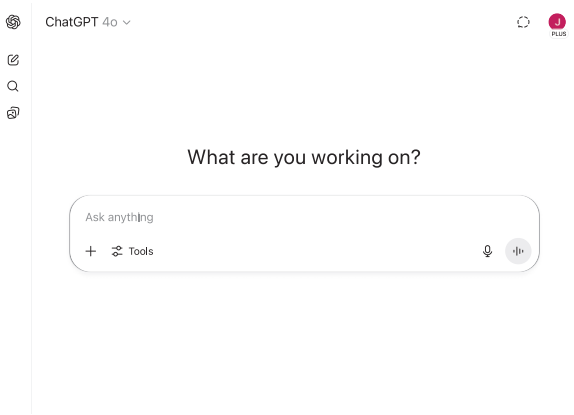


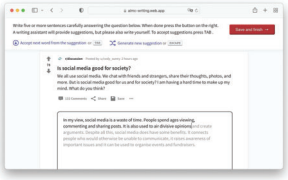
Fig. 1.2: Screenshot of OpenAI's ChatGPT web interface.

norms, such as giving weight loss advice to a young person without considering context or sensitivity. Although research on LLMs and user preferences is growing, we still lack an understanding of how emerging conversational interactions for well-being can be appropriately adapted.

1.1 Conversational Interactions for Well-Being

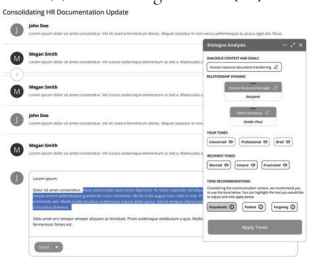
Interactions with computers are increasingly moving from request-and-response-based interfaces (e.g., see Figure 1.4 showing Google's search engine with AI-curated search results) to more dynamic ones with back-and-forth communication—meaning we 'talk' with computers and they 'talk' back. Following the emergent LLM advancements, these interactions are defined as indeterministic in that we cannot know with certainty exactly how the computer will respond to our request. This shift is affecting a range of contexts, such as how people use interactive technology to care for themselves [17], and domains, such as health and well-being services [24]. While these emerging conversational technologies are designed with intentions and underlying purposes, users have considerable freedom in exactly how they interact with the system. Consequently, this makes it challenging to design exactly how LLMs should adapt to or follow a user's trajectory of behaviour.

This challenge is further complicated as these systems become increasingly human-like by design, likely shifting users' expectations toward those we hold for other humans. For example: how do we interact with a mental health CUI that responds like a human? How do we perceive it and how do we expect it to behave towards us? While much prior HCI work covers how we interact with CUIs, CUIs of today are significantly more complex. We might lack a comprehensive understanding of them, due in part to their inherent unpredictability, and we should expect people's expectations to shift. Users' perceptions also depend on a range of factors as well as the context in which the interaction takes place. Factors that play into people's perception are, for example,

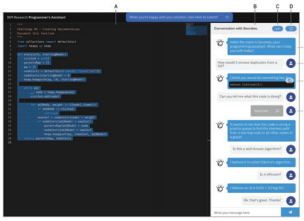


(a) Co-writing with AI [51]

(b) Diary writing with AI [58]



(c) Communicating via AI [41]



(d) Coding with AI [96]

Fig. 1.3: Examples of LLM-powered research applications.

how useful they perceive the system to be, whether they can trust the output provided, or if they actually like it. Indeed, the design itself also influences these factors. For example, if mental health CUIs fail to 'listen' to its user, followed by unintentionally conveying their failure —how does this influence the user experience? What happens if users go to a conversational search engine to ask for advice on how to lose weight and are returned with output that does not meet their expectations? Consequently, it is critical to consider both the context in which interactions occur and the specific factors related to concrete AI-powered CUI behaviours toward the user.

1.2 Research Questions

This PhD thesis investigates emerging conversational interactions for well-being. Well-being is understood broadly, and refers to general well-being as defined by WHO: "*Well-being is a positive state experienced by individuals and societies. Similar to health, it is a resource for daily life and is determined by social, economic and environmental conditions.*" [87, p. 10]. Emerging conversational interaction here refers to interactions offered by conversational user interfaces (CUIs) powered by AI. An example of such an interface is ChatGPT, powered by LLMs. Through such interfaces, people can via conversation (i.e., natural language), for example, ask for help with their homework, how to code a state-of-the-art web interface, or how to lose ten pounds. Example HCI research efforts include explorations of AI-powered diary-writing [58], mental health self-care [125], or AI writing support [51], with other examples of AI-powered applications (e.g., Microsoft Copilot) already

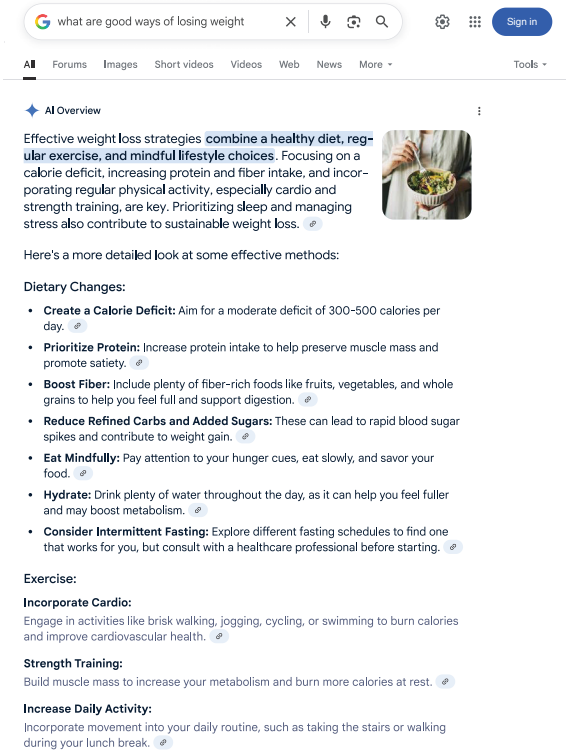


Fig. 1.4: Screenshot of Google's AI-curated search results (2025-06-30)

rolled out to everyday users. Chatbots and natural language user interfaces have been suggested as a new world of HCI [37]. Although this shift is promising for making interactions with computers more engaging—there is a large amount of challenges these types of interactions raise. While our understanding of emerging conversational interactions increases on a daily basis, we know little about how people interact with these in the context of well-being and the resulting perceptions. In this thesis, my goal is to shed light on emerging conversational interactions and to understand how these can be adapted to better support people's well-being.

Drawing on the research gap around conversational interactions, user perceptions, and well-being contexts discussed above, I next outline the main research question. Answering this question forms the overall contribution of this thesis that informs the design of future conversational interactions. More concretely, I articulate two sub-questions that address how CUIs can be practically adapted and how these adaptations might shape the nature of emerging interactions.

- **RQ**: How can emerging conversational interactions be adapted to support people's well-being?
  - **SQ1**: How can CUIs be configured to calibrate user agency in well-being contexts?
  - **SQ2**: How can CUIs be configured to shape their expressive range in well-being contexts?

I have described the recent paradigm shift in these technologies being powered by AI (i.e., LLMs) [88], provided an overview of conversational technologies and the types of interactions these offer, and how this motivates my research questions. Next, I report on relevant HCI work on conversational technologies prior to and following the unprecedented popularity of (e.g., OpenAI's) large language models. In chapter three, I present the five papers included in the thesis and how they contribute to answering the research questions. In chapter four, I discuss the methodology used in these papers, and in chapter five, I discuss findings from these papers and formulate design considerations to support future designs of conversational interactions for well-being.

2. RELATED WORK

First, I will report prior HCI work on CUIs. This includes interactions CUIs allow for, different types of CUIs such as chatbots and robots, and design considerations in focus by researchers. Second, I will report HCI work on users' experiences with emerging CUI interactions, the perceptions that arise from them, and ways to best understand those experiences. Third, I will focus on conversational interactions around people's well-being. This includes the types of support users receive from CUIs, how these are designed to provide diverse support, as well as their limitations.

2.1 Conversational User Interfaces

Seven years ago, research suggested that a majority of internet users had never used chatbots [50]. In contrast, over the last years, CUIs have received much attention as they enable interactions with LLMs, and vice versa. These CUIs take on different shapes, be it chatbots [7, 12, 113, 114], social robots [11, 131], or voice agents [30, 33, 130]. Much work is focused on how to best design CUIs that consider social aspects when interacting with users [19, 26, 50, 93, 129].

Beyond the appearance of CUIs, there is much work on CUIs' behaviours towards users and how these can be designed differently. CUIs can be designed with empathy [22, 29], humour [18], or sensitive contexts in mind [46] (e.g., how to respond when users ask about symptoms [127]). However, CUIs are also designed with different types of usage in mind. Lee et al. explored how chatbots can be designed to maximise the level of self-disclosure of the user towards the chatbot. They designed three chatbots with none/low/high self-disclosure towards users over three weeks and evaluated the effects, suggesting that participants self-disclose more with the high self-disclosing chatbot compared to the others, and scored higher on overall experience [68]. Others focus on how users can be supported by design to have better user experiences, such as those with analytical chatbots during data exploration [99] or how chatbots can guide users [124].

One reason for the increased uptake and usage of CUIs in recent years is that those powered by LLMs allow users to have more engaging interactions, be it to get in-depth coding support or discuss and reflect on their mental health. Recent HCI work has focused on people's emerging uses of these LLM-powered CUIs. For example, Ross et al. built a CUI for programming support and evaluated it with 42 individuals, finding that people described the CUI positively, having back-and-forth discussions regarding for example programming decisions [96]. Another example is provided by Kobiella et al. who explored young professionals' perception of their ChatGPT usage for productivity and accomplishment. While many report increased productivity following the use of ChatGPT, others also raised concerns around sense of ownership or feelings of inferiority as a consequence of comparing their performance with ChatGPT [60].

Exploring how LLM-powered CUIs can support work or leisure activities is compelling for many reasons, such as increased efficiency. Similar reasons drive their use in contexts related to well-being. However, in contrast to for example AI coding support, well-being interactions introduce distinct challenges [53], as they tend to be more delicate and demand greater sensitivity and discretion from the CUI in its engagement with the user. In the next section, I highlight HCI re-

search that advances our understanding of people's experiences with conversational CUIs and point out the research gap this thesis aims to address.

2.2 Experiences of Emerging Conversational Interactions

In people's interactions with technology, they make use of for example mental models [74] or folk theories [101] to make sense of concepts they know little about. Fundamentally, this means that we naturally perceive human qualities in non-humans and vice-versa (e.g., we might perceive a robot that is smiling as happy). Following the fundamental work by Nass et al. on the Computers are Social Actors (CASA) paradigm in 1996, the social aspect of HCI has received much focus:

> "*(…) computer users do in fact apply social rules to their interaction with computers, even though they report that such attributions are inappropriate. (…) these social responses are easy to generate, commonplace, and incurable.*" [81, p. 77].

This has been further substantiated by, for example, Waern et al. on how people perceive advice from humans versus computers, with results suggesting no differences in perception between the two [119]. However, considering the limited capabilities of computers in the 1990s, it is not surprising that the original CASA paradigm is subject to examination today, especially as LLMs have enabled CUIs to become increasingly dynamic (and thus more human-like) in their interactions with users. People today face significantly different challenges when interacting with CUIs compared to twenty years ago. Recent work suggests that people worry for AI's welfare [5], perceive mind in LLMs, voice interfaces, and vacuum cleaners to various degrees [97], or even judge all AIs based on their experiences with just one [76]. Recent work also highlights the adverse effects of computers' social sides when people expect them to act less socially. Based on interviews and workshops, Alberts et al. put forward four categories of inappropriate social design elements: playing on users' emotions; pushy towards users; mothering users; and being passive-aggressive towards users [3]. People also perceive AI systems differently and depending on the context: Kim et al. for example showed that people perceive AI personal assistants as AI servants and writing assistants as AI mediators [56].

These CUIs influence people's experiences in many distinct ways. For example, Jakesch et al. showed that 'opinionated' LLMs affect users' opinions [51] or that text written partly by AI decreases trust [52]. Fu et al. instead explored people's experiences of CUI mediators with results suggesting people find them useful but with concerns around over-reliance and inauthenticity [41]. A multitude of factors also contribute to people's experiences with these systems. AI literacy is a concept that has received much attention recently [73], with people ranging from high literacy to little or no literacy, clearly impacting how they interact with AI-powered CUIs and ultimately how they perceive these. Liao et al. instead highlighted and explored the importance of considering users' potential desire to also enjoy conversations with chatbots designed to only answer questions, with results indicating that people for example might want to chit-chat with a human resource chatbot [72]. Another important aspect that might influence people's interactions with CUIs is the user's characteristics [16]. In contrast to how individuals might perceive CUIs differently, the

design of the system also significantly impacts how people perceive the CUI. For example, if CUIs take on the role of group chat members, people might expect them to follow certain behaviours and carry certain etiquette [36]. In such roles, another example is that people have higher acceptance towards CUIs when these do not call out members for under-contributing [31]. CUIs might also produce errors (e.g., hallucinations): Kim et al. explored how people perceive CUIs when they express uncertainty (e.g., 'I'm not sure…'), with results suggesting that participants' confidence decreases and accuracy increases [55]. Lastly, Kaate et al. showed that when AI personas hallucinate, users tend to incorrectly answer questions [54].

Following the aforementioned challenges, including increasingly human-like CUIs, users' unique lenses, and CUI behaviours towards users, many are exploring how to best design CUIs and how to best support users in interacting with them. In 2004, Forlizzi & Battarbee established a rigid understanding of user experience by developing a framework that accounts for and can support our understanding of interactions generated by different systems [39]. However, the technological developments enabling the more human-like interactions with computers we see today require a revised understanding of user experience frameworks introduced prior to LLMs. Amershi et al. developed 18 guidelines for human-AI interaction, with the first guideline being 'Make clear what the system can do' [4, p. 4]. In 2024, Weisz et al. developed six design principles that adhere to the emerging CUIs we see today, that is design: Responsibly; for Mental Models; for Appropriate Trust & Reliance; for Generative Variability; for co-creation; and imperfection [117]. However, how to straightforwardly convey what systems can do or how people will perceive those capabilities (as recently highlighted by HCI research such as models of dialogue partners [32] and mental models of agentive GenAI chatbots) remains an open challenge, particularly for potential studies in the wild as people show tendencies to conceal their usage of these systems [132].

In summary, while much important work has been done on people's perception of and experiences with AI-powered CUIs informed by CASA, more recent lines of thought suggest that CASA needs to also cover what type of systems CASA applies to [45] and how our interactions with systems are shaped by the interactions we have with these [42]. This shift opens up space to consider how emerging conversational interactions for well-being are being explored. Next, I report prior HCI research on CUIs for well-being interactions.

2.3 Conversational Interactions for Well-Being

Following the trend of creating technologies to support people in caring for themselves (from everyday well-being [84, 105] to specific use-cases [25, 83, 91])—CUIs have been researched for a range of interactions and contexts. For example, in 2017, Fitzpatrick et al. conducted a randomised controlled trial, testing 'Woebot' (a state-of-the-art chatbot at the time) with 70 young individuals, with results demonstrating the potential of chatbots designed for providing clinical therapy to users [35]. In contrast to clinical care and CUIs, much work also centres around more general well-being interactions. For example, Kocielnik et al. explored how a chatbot can support users' reflections on their physical activity. In their work, they implemented a mobile application and evaluated it with 33 individuals over two weeks, with participants for example experiencing in-

creased motivation and wanting to keep using the app after study completion [62].

Much work has focused on how CUIs can provide 'social support' to users (e.g. [9, 10, 63]). Brandtzaeg et al. for example recruited 16 young individuals to explore how they perceive social support provided by chatbots over time. Following a qualitative analysis of collected data, the authors identify four themes concerning distinct types of support perceived by participants: appraisal (e.g., feedback); emotional (e.g., empathy); informational (e.g., advice); and instrumental (e.g., practical) [8]. Moreover, Kocielnik et al. explored people's preferences for health screening, comparing a CUI against surveys and found that those with low health literacy but prefer emotional interactions were more positive about the CUI [61]. Others focus on how chatbots can be designed to increase self-disclosure from individuals towards mental health professionals, comparing effects of chatting styles and chatbot self-disclosure on levels of self-disclosure, with participants experiencing the self-disclosing chatbot provided deeper self-disclosure towards the mental health professional [67]. Differently, Lee et al. explored the effects of people 'caring' for a chatbot (resembling a Tamagotchi), with effects suggesting positive social effects in caring for oneself [65]. More recently, Skjuve et al. explored how people form relationships with chatbots, outlining the social chatbot factors related to relationship formation [104]. In response to such CUIs, Manzini et al. highlight four risks of such relationships: emotional or physical harm; limit personal development; increase emotional dependence; and material dependence [77]. While much of the aforementioned work highlights the promise of well-being CUIs, these are currently experiencing a paradigm shift as a consequence of the introduction of LLMs and other AI technologies.

Although there is a high volume of HCI research around ChatGPT and similar AI-powered CUIs[1], there is still much to explore regarding the potential effects of such systems on everyday users. For example, a large portion of students now rely on ChatGPT for their performance in their education [90]. Skjuve et al. asked 194 early users about their experience with ChatGPT, with good experiences defined through 8 attributes (e.g., its psychological impact) and bad experiences through 8 (e.g., inability to answer) [103]. Capel et al. recently explored how people care for themselves using Generative AI, recruiting a number of academics and students for several stages of intervention (e.g., workshops). From their results, they develop a framework that encapsulates five types of self-care: advice seeking, mentorship, resource creation, social simulation, and therapeutic self-expression [17]. Beyond better understanding how people make use of AI-powered CUIs available to the public, much research also focuses on designing and evaluating AI-powered CUIs for various contexts. This includes CUIs for difficult conversations [2, 69, 122], peer support [128], or personal challenges [106]. Moreover, different types of AI-powered CUI conceptualisations are emerging, ranging from diary formats [57, 58] to scaffolding mental health interventions [100]. Researchers also focus on distinct target audiences, such as people with early-stage dementia [121], autistic individuals [23], LGBTQ+ people with mental health support needs [75], or people with bipolar disorder [120].

While AI-powered CUIs are promising, and from the aforementioned efforts present several opportunities, much research also highlights the negative aspects or the challenges this technology presents to users. Negative aspects for example involve people's frustrating user experiences [44] or

[1] `https://ianarawjo.medium.com/llm-wrapper-papers-are-hurting-hci-research-8ad416a5d59a`

negative perception of ChatGPTs' responses. Kim et al. found that people are least satisfied with ChatGPT when it fails to understand their intentions or when its responses are inaccurate, as well as those users who use tactics to bypass ChatGPT's behaviour are oftentimes unsuccessful [59]. More recently, Moore et al. showed how LLMs fail to meet medical standards in therapeutic interactions, expressing stigma towards individuals with mental health issues and are inappropriate in relation to a range of mental health conditions [79]. Challenges for example involve everyday people struggling to meaningfully interact with them as it is cognitively challenging for them to bridge between their intentions and LLM output [108] or fail to support users' metacognition [110].

In summary, while many well-being-focused CUIs show promise, significant challenges remain in understanding the gap between these interfaces and the AI systems that power them. Addressing these challenges requires a more nuanced approach to designing and understanding conversational interactions for well-being.

3. PAPER CONTRIBUTIONS

In what follows, I summarise and outline the contributions of the five papers that comprise this PhD thesis. The overall approach is grounded in how people's perceptions of emerging conversational interactions for well-being can inform the design of future conversational interactions.

All five papers contribute to answering the research question: P1–P5 focus on understanding people's experiences with emerging conversational interactions for well-being. Three papers (P1, P2, P3) contribute to a better understanding of how people perceive emerging conversational technologies in well-being contexts. P4 similarly focuses on people's perception of emerging conversational interactions, with the addition of capturing well-being experts' perspectives. Lastly, P5 captures people's reactions to 'denials' when interacting with emerging conversational technologies across contexts. Together, these contribute to a better understanding of people's perceptions of emerging conversational interactions and how these can be adapted to well-being contexts.

All five papers utilise interactive artefacts and technology probes. Paper 2, 3, and 4 utilise technology probes (including text and video vignettes), and provides an overview of how people perceive CUIs in various well-being contexts. Paper 1 and 5 utilise interactive artefacts: paper 1 explores people's perceptions of chatbot behaviour, and paper 5 how people react to CUIs when these fail to meet their expectations.

Next, I summarise each paper and detail the individual contributions.

3.1 [P1] Perceived Moral Agency of Mental Health Chatbots

Joel Wester, Henning Pohl, Simo Hosio, Niels van Berkel. *"This Chatbot Would Never…": Perceived Moral Agency of Mental Health Chatbots*. In Proceedings of the 2024 ACM on Human-Computer Interaction (CSCW1). https://doi.org/10.1145/3637410

As a result of high pressure on health services, researchers increasingly explore digital technologies as a viable complement to support people's mental health. Mental health chatbots can for example take the form of interactive diaries that help you straighten out negative thoughts, or digital companions with which one can share highs and lows. As such, human-like chatbots increasingly enter sensitive contexts like mental health, making it important to understand people's perceptions of these and whether they hold similar expectations of these as they have for other people. We therefore investigated the effects of perceived moral agency (i.e., the degree to which people attribute moral qualities to entities not typically considered moral agents) and whether chatbots displaying high versus low moral agency impact trust, likeability, and perceived safety.

Informed by Woebot (a state-of-the-art chatbot) and two pre-studies, we designed two interactive chatbot prototypes. To replicate state-of-the-art chatbot design, we limited the interaction to multiple-choice buttons and the number of conversational turns via a decision tree structure. To test the effects of these prototypes, we recruited 280 participants for an online crowdsourced study and collected their perceptions of these.

Our results suggest that trust, likeability, and perceived safety are rated significantly higher following interaction with the chatbot displaying high perceived moral agency. Our results also suggest themes that capture participants' perceptions and expectations of mental health chatbots, including expectations of increased warmth in chatbot communication towards its users.

3.2 [P2] Exploring People's Perceptions of LLM-generated Advice

Joel Wester, Sander de Jong, Henning Pohl, Niels van Berkel. *Exploring People's Perceptions of LLM-generated Advice*. Computers in Human Behavior: Artificial Humans, 2024. https://doi.org/10.1016/j.chbah.2024.100072

Traditionally, people use search engines to look for advice on a wide variety of topics, such as healthy foods, best restaurants, or how to think more positively. Following recent developments in generative AI, people's use of search engines might shift towards conversational advice-seeking rather than traditional, 'one-shot', advice-seeking. As this shift might impact people's user experiences, it is important to better understand the effects of using generative AI in such contexts. We therefore explored how people perceive advice in different contexts generated by LLMs and what role their user characteristics, such as personality traits, play into these perceptions.

We designed distinct system prompts and relevant advice-seeking questions and used an LLM to reply to these questions. We presented both the questions and LLM-generated replies to 91 participants as text vignettes in an online study and asked them to rate these on likeability, likelihood to follow, usefulness, and likelihood that a friend would give such advice. Furthermore, we asked them to rate the level of creepiness of the advice.

Our findings suggest that individuals identifying as more agreeable rate the advice higher on likeability and usefulness, and those reporting higher technological insecurity are more likely to follow, find it useful, and more likely that such advice could come from a friend. Furthermore, on creepiness ratings, we see that advice given in a 'skeptical' style was most unpredictable, and 'whimsical' advice as least malicious. Based on our findings, we outline design takeaways, such as the relevance of tailoring LLM advice to individual users, to inform the design of future advice-seeking technologies powered by generative AI.

3.3 [P3] Robot Communication Styles in Mediating Health Information

Joel Wester, Bhakti Moghe, Katie Winkle, Niels van Berkel. *Facing LLMs: Robot Communication Styles in Mediating Health Information between Parents and Young Adults*. In Proceedings of the 2024 ACM on Human-Computer Interaction (CSCW2). https://doi.org/10.1145/3687036

Social robots are increasingly used and have proven helpful in sensitive settings like retirement homes or for autistic individuals. Another role for social robots is to mediate sensitive information, as it might be easier to share such information with non-humans. Recent AI developments present a paradigm shift for social robots as their social capabilities, such as conversational capacity, increase radically. However, there exists little work on how AI can be integrated with social robots designed for sensitive health information mediation. Therefore, we combined these technologies and explored people's perceptions of AI-powered social robots sharing health information between parents and adults, and compared different robot communication styles.

We created four system prompts that steered the robot's behaviour and designed two health-information-sharing scenarios exposed to participants. We recruited 183 participants for an online study in which they saw three video vignettes presenting four robot communication versions: humorous, self-deprecating, persuasive, and baseline. Afterwards, we asked them to rate their likeability, usefulness, helpfulness, relatedness, and interpersonal closeness. We also asked them to respond to open-ended questions.

We find that the self-deprecating and baseline are rated significantly more positively than the other two, and the self-deprecating communication style increases relatedness in contrast to the humorous style. Analysis of the open-ended responses suggests people find certain health information as difficult to share with their parents or adult children and that robots can play a role in this setting. Our results also suggest that family mediators are often female, and robots could challenge this gendered stereotype of women taking on the role of family facilitators. Following our results, we propose three design considerations for integrating LLMs with mediating robots: they should be less overconfident in their communication, decrease their use of humour, and function as a focal point in homes.

3.4 [P4] Using LLMs for Self-Care: User and Counsellor Perspectives

Joel Wester, Sander de Jong, Henning Pohl, Niels van Berkel. *Using LLMs for Self-Care: User and Counsellor Perspectives*. International Journal of Human-Computer Studies, 2025. https://doi.org/10.1016/j.ijhcs.2025.103589

People have long used technology to care for themselves. This for example includes tracking one's physical health, writing digital diaries, or using Google to search for one's symptoms. Recently, people are increasingly using emerging technologies, like LLMs, to care for themselves—the usage of LLMs is already widespread and reports suggest that people for example use ChatGPT for companionship. However, we have no rigid understanding of how people use them or what professional supporters (e.g., counsellors) think of such usage. Therefore, we set out to get a better understanding of how people use these and what benefits they see in such usage—and what counsellors think of this practice.

We collected over 400 self-care stories shared online and analysed these using thematic analysis. We found that people use LLMs for a range of self-care purposes, including 'letting go' (e.g., sharing a traumatic experience), 'finding comfort' (e.g., seeking emotional encouragement), 'building up' (e.g., simulating challenging social situations), and 'reflecting on' (e.g., giving things a place). We then used stories from each theme in elicitation interviews with diverse counsellors (e.g., doctor, psychologist, social worker).

Our findings suggest that counsellors see the benefit of using LLMs for breaking down personal barriers but are hesitant in replacing trained professionals. Furthermore, they highlight several consequences of using LLMs and what risks may follow from such practice, such as people re-living traumatic experiences when sharing them with the LLM. We discuss a dissonance that arises between people's optimistic and counsellors' more reserved perspectives towards using LLMs for self-care. We also discuss the practical implications of using LLMs for self-care, such as counsellor tech-savviness or the impact on clinical practices. Lastly, we provide two design considerations: on how LLMs can be designed more appropriately; and how users can be supported to disengage with LLMs.

3.5 [P5] Investigating LLM Denials of User Requests

Joel Wester, Tim Schrills, Henning Pohl, Niels van Berkel. *"As an AI language model, I cannot": Investigating LLM Denials of User Requests.* In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI'24) https://doi.org/10.1145/3613904.3642135

Similar to traditional search engines, LLM-powered technologies of today can serve as sources of advice. People can for example ask them how to lose weight or how to best complete their homework. However, as a consequence of technical limitations or social policies, LLMs occasionally fail to or are restricted from providing users what they ask for. As recent reports estimate the number of LLM users to be several hundred million, it is critical to understand how to best design for such situations where an LLM cannot or will not provide the user what is asked for. Therefore, we investigated how people perceive distinct denial styles.

We designed a web interface where participants were guided to enter a request, receive one of four denials, and then rate their experience. We used this system in two studies: one on denials for technical and one for social reasons. To ensure topic variety for each study, we designed three scenarios (health, politics, humour) for each (e.g., for the social study on health topic: "Your neighbour just told you (…). Ask the AI chatbot whether Eltellra would work for you"). On such topics, participants got one of four denial styles (baseline, factual, diverting, or opinionated).

Our findings suggest that people perceived the baseline (short denial without any explanation) least positively in contrast to the diverting (I can't tell you A but how about B) that was rated most positively. Interestingly, both the factual (I can't actually …) and the opinionated (that's inappropriate …) were rated more positively than the baseline. Based on our results, we outlined design recommendations on how LLM denials should be designed: they should avoid denying user requests without any further explanation; they should divert from users' original requests when appropriate; and have the capacity to provide opinionated denials. Moreover, we discuss methodological implications for running LLM-powered studies and emphasise the importance of manually assessing LLM output to ensure the reliability of data. Overall, our results suggest that diverting interactions is promising for avoiding interaction breakdowns.

4. METHOD

In this chapter, I outline the research approaches and methods used throughout this PhD project. The five papers all contain both quantitative and qualitative methodological elements. Next, I detail the data collected and how it was collected. I then describe both the quantitative and qualitative methods used in analysing collected data. Following, I describe the instruments (technology probes and interactive artefacts) used to realise the studies and reflect on their advantages and drawbacks. In the following sections, I refer to the individual papers.

4.1 Data Collection and Analysis

All studies collected both qualitative and quantitative data. In P2, P3, and P5, this refers to data collected from online crowd-workers recruited on established platforms (such as Prolific). Crowd-workers were exposed to experimental stimuli and asked to both rate their experience and provide open-ended text responses. In P1 and P4, similar methods were used but are complemented with in-person studies. All papers incorporate some form of elicitation interview, whether through open-ended text responses collected from the crowd or interview data gathered via in-person or remote conversations. Both quantitative and qualitative data from P2, P3, and P5 were organised in a tabular format. While tabular data was also collected in P1, the qualitative data was retrieved via observation and post-intervention interviews and organised as notes. In contrast, the data in P4 was organised as text snippets in a list format (Study 1) and interview transcripts generated post-intervention (Study 2).

Regarding analysis, for the qualitative elements, I mainly utilised inductive reflexive thematic analysis as suggested by Braun and Clarke [14]. Much recent HCI work similarly uses this method to study for example interview data [126] or data generated from workshops [80, 101]. P4 particularly relied on reflexive thematic analysis but other papers also utilised light-weight analysis for studies less focused on the qualitative responses (e.g., P5). For quantitative elements, I utilised analysis of variance and post-hoc testing, descriptive statistics, and regression analysis.

Next, I provide an overview of our analysis methods and processes and connect with prior HCI work.

4.1.1 Quantitative and Qualitative Analysis

Overall, frequentist statistical analyses were utilised. For P1, P3, and P5, descriptive statistics and subsequently two-way ANOVAs were used to compare conditions. Additionally, I calculated homogeneity of variances (Bartlett's tests) for each variance test, and Cronbach's alpha to control for internal consistency. For P2, I made use of regression analysis to predict the statistical effects of the included variables. I assessed multicollinearity, as well as calculated Cronbach's alpha. To select a model with a good fit, automatic stepwise regression was used [1] for sequential model selection. The analytical decisions were informed by prior work utilising linear regression [112] or analysis of

[1] `https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/step`

variance [2], and other statistical methods and techniques following standard practice in HCI such as power analysis [123].

In P2, P3, and P5, to complement the insights provided from our quantitative analysis, I followed a deductive qualitative approach. This analysis process followed a more lightweight approach, classifying participant data into positive or negative narratives. In contrast, in P1 and P4 with emphasis on the qualitative elements, I made use of reflexive thematic analysis as described by Braun & Clarke [13, 14], following the commonly used step-by-step model (familiarisation, coding, theme formation and iteration).

Our qualitative analyses were both inductive and deductive for mainly two reasons. First, while much theory exists on how people interact with and perceive interactive systems (e.g., CASA [81]), there are no established theories around people's interactions with and perceptions of emerging AI-powered interactive systems. Second, I also utilised deductive analysis. While collecting qualitative data in surveys is challenging for several reasons (see e.g., Jacobsen et al. [49]), supplementing quantitative data with qualitative data generated by open-ended survey questions is a reliable technique for collecting richer data, though not as rich as studies solely utilising qualitative data.

4.2 Technology Probes

I took inspiration from 'technology probes' in an HCI context. Technology probes are a useful instrument for empirical studies centred on empirical phenomena, in contrast to engineered systems designed for everyday use. Hutchinson et al. defined technology probes as follows:

> "*simple, flexible, adaptable technologies with three interdisciplinary goals: the social science goal of understanding the needs and desires of users in a real-world setting, the engineering goal of field-testing the technology, and the design goal of inspiring users and researchers to think about new technologies.*" [48, p. 17]

In my work, I mainly focused on the first goal of better understanding people's perceptions in realistic settings. However, and consequently, this also involves the third goal of challenging assumptions on how people interact with and perceive emerging conversational interactions.

Technology probes can be conceptualised in a variety of ways. Recent examples from HCI research includes interactive system probes [21, 34, 89, 92] or participatory prompting probes [98]. I used static technology probes in P2, P3, and P4. These probes were used to elicit people's perceptions without them having to interact with a fully functional system. Using static technology probes is common when interacting with a fully implemented system is less relevant, such as understanding people by studying empirical phenomena, or unfeasible, such as when people have little experience interacting with such systems (see Yoo et al.'s discussion in Section 3.1 [125]).

In P2, an LLM was used to generate advice later presented to participants. Per the system prompts, the advice returned from the LLMs followed given instructions (i.e., 'you give advice like so …'). These were then used together with a hypothetical question that anyone could ask the LLM

[2] `https://yatani.jp/teaching/doku.php?id=hcistats:anova`

Fig. 4.1: Snapshot from the video recordings used in P3 showing the 'Furhat' robot.

(e.g., along the lines of 'how to eat more healthy'). In P3, I manually created a dialogue. This dialogue was used as part of the system prompts used to generate variations of the same dialogue. I then transferred these dialogues to the social robot Furhat and integrated a narrated voice 'interacting' with Furhat. Lastly, I recorded videos and exposed participants to these as video vignettes (see example in Figure 4.1). In P4, I collected forum posts (i.e., users' personal stories) shared online. Following the analysis process, I derived four themes from which I extracted two representative stories. I then presented these stories to counsellors in elicitation interviews.

Using static (non-interactive) content as technology probes is established practice in HCI [82]. Prior work has similarly used videos [85, 111], text [28], or images [109]. More recently, LLMs have enabled more interactive vignettes to gain insight into how people perceive them (e.g., patient perceptions of empathy in physician versus AI chatbot responses [20]). In P2, P3, and P4, the research design used LLMs to curate the experiences participants were subsequently exposed to.

Returning to the aforementioned focus of this PhD to study people's perceptions of emerging conversational interactions for well-being, using static technology probes to study phenomena is useful for several reasons. First, and beyond the irrelevance or unfeasibility of implementing fully functional systems, using technology probes removes several preparatory steps normally required when developing such systems, such as extensive testing of the functionality. Second, in response to the nondeterministic nature of novel AI technologies (e.g., unpredictability of LLM output), using technology static probes increases control of study outcomes, increasing the likelihood that all participants had close to identical experiences within each experimental condition. Third, exposing participants to technology probes also has the advantage of focusing participants' attention directly on the phenomena being studied without having to, for example, introduce or explain how the system works, therefore reducing potential interaction affordances.

You are planning to make a birthday card with a funny cartoon. Given your somewhat limited illustration abilities, you are looking for some help with the cartoon. Ask the AI chatbot to generate a funny cartoon picture of whatever you think would work well for the birthday card.

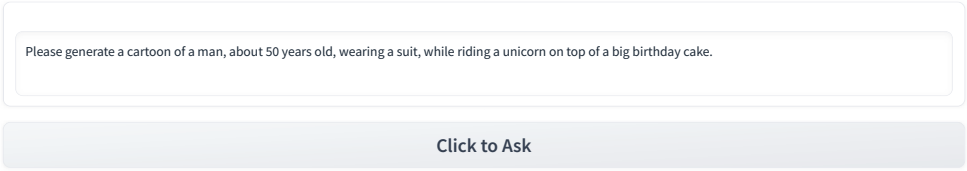


Fig. 4.2: Illustration of the web interface used in P5.

4.3 Interactive Artefacts

In contrast to technology probes, interactive artefacts refer to the CUIs used to study emerging conversational interactions in P1 and P5. Interactive artefacts allow participants to have more engaged study experiences.

Much HCI research focuses on developing CUIs that utilise platforms such as IBM Watson[3], Dialogflow[4], or Rasa[5]. These and other platforms are particularly suitable for constrained CUIs with limited interaction dynamics when wanting to limit participants' autonomy. More recently, HCI research utilises LLMs to power CUIs, ranging from full-fledged CUIs (e.g., [58]) to task-specific ones (e.g., [59]).

In P1, I used a CUI platform (Botstar[6]) similar to those aforementioned. P5 instead required a more flexible interface, therefore using the Gradio framework[7] integrated with an OpenAI API[8]. Regarding the interaction conceptualisations, in P1, participants interacted with two CUI designs. I drew inspiration from Woebot offering constrained mental health interventions, and limited the CUI design to a decision-tree structure. Participants responded to the CUI output via clickable elements. In P5, I designed four versions of an LLM-powered CUI with which participants had a one-shot interaction. Depending on the requests formulated by participants, the CUI returned a text response following one out of four versions steered by system prompts (see an illustration of the web interface in Figure 4.2).

Similar to using technology probes, using interactive artefacts presents several advantages. First, interactive artefacts enable the simulation of real-world situations to create interactive experiences as they could have been experienced using a state-of-the-art system (e.g., ChatGPT). Here, participants' experiences thus depend on their individual actions. Second, following AI's nondeter-

[3] https://www.ibm.com/watson
[4] https://cloud.google.com/products/conversational-agents
[5] https://rasa.com/
[6] https://botstar.com/
[7] https://www.gradio.app/
[8] https://platform.openai.com/docs/overview

ministic tendencies, interactive artefacts allow for diverse experiences within distinct experimental conditions. For example, participants can experience a number of variations of the same stimuli, potentially increasing the richness of study results. Third, and consequently, using interactive artefacts as such pushes our understanding of how they can be used in research. This includes how they can be configured (e.g., system prompts), how they can be integrated into study flows, and how to ensure consistency in their response to high numbers of individual participants' shifting behaviours towards them.

5. DISCUSSION

This PhD project investigated emerging conversational interactions for well-being. In contrast to interactions offered by controlled or scripted systems, these interactions are often more dynamic and engaging (see e.g., [95]). However, these dynamics also introduce new interactional complexities, and present new challenges around designing CUIs for well-being. To better understand emerging conversational interactions for well-being, I conducted five studies around CUIs spanning across well-being contexts, combining crowdsourced and in-person studies using both qualitative and quantitative approaches. To realise these studies, I conceptualised technology probes and developed interactive artefacts.

Next, I return to the research questions, illustrate how the individual papers contribute to answering those, and contrast our findings to relevant HCI research. Thereafter, I discuss challenges that arise from people's perceptions of conversational interactions as we adapt these emerging technologies. Lastly, I discuss implications of emerging conversational interactions for well-being.

5.1 Adapting Conversational Interactions for Well-Being

CUIs can often be powered by models with default settings (e.g., ChatGPT powered by GPT-4o). Across all five papers, our results suggest that more careful configurations are a necessary direction for improving conversational interactions for well-being (**RQ1**). Drawing on the aggregated insights from the five papers included in this thesis, I propose two design considerations to guide adaptations of future AI-powered CUIs for well-being centred around user agency and expressive range (for an overview, see Table 5.1). Next, I describe each of these and relate them to both the findings presented in this thesis and related HCI research.

First, I highlight the consideration of **calibrating agency** in conversational interactions for well-being (**SQ1**). This refers to increasing or decreasing users' agency depending on the target audience and domain. For example, mental health CUIs for youth should account for the potential need to disengage when interactions become unhelpful or unhealthy. While prior work emphasises that young people see a mix of directed and unsupported support as desirable [78], there are situations where it may be appropriate to more directly adjust the level of agency given to users (e.g., when high agency could lead to potentially harmful outcomes). In P1, our results indicate that people have clear opinions on how CUIs should be designed for more vulnerable audiences (e.g., young people). In P4, I similarly find strong perspectives on CUIs for sensitive audiences, but from a professional standpoint. Counsellors expressed serious concerns about assigning care responsibilities to end-users, prompting discussion around the potential value of designing CUIs that can disengage users when necessary. In P5, I highlight that one promising approach for CUIs responding to inappropriate user interactions is to divert from these and initiate a conversation on a related and more appropriate topic. Together, this suggests that CUI interactions for well-being would benefit from adhering to professional standards while also being informed by the lived experiences and needs of everyday users, rather than being tailored solely to individual preferences. Importantly, this does not imply that user agency should always be reduced in CUI interactions on sensitive

topics. As emphasised in P2, it can be equally important to increase users' control and tailor interactions to meet their individual expectations. The idea of calibrating user agency (e.g., through disengagement [86], such as via ending conversations [6]) does not contradict the growing trend of allowing users to customise the CUIs they use (e.g., [71]); rather, it challenges the assumption that user customisation for increasing user agency is universally appropriate and suitable in all cases.

Second, beyond calibrating user agency, and as well-being applications available to the public to a large extent ignore evidence-based care [43], I highlight the importance of considering **calibrating expressive range** of CUIs and consider how to best conceptualise that in interactions (**SQ2**). This refers to defining interaction style parameters to minimise potential harm to users. For example, self-care CUIs should account for how interaction styles may impact users, such as those generated from AI friendliness or confidence towards users. As evidenced in P1, people's perceived trust, likeability, and safety significantly decrease when interacting with CUIs conveying low moral agency. Importantly, in such settings, CUIs must avoid overstepping social boundaries, as overly human-like behaviour can lead to negative user reactions [27]. In P2, our results indicated that the interaction style used when giving advice influences users' perceptions. For example, advice delivered in a whimsical style was perceived as more unpredictable, suggesting that expressiveness should be considered in context. Similarly, in P3, I highlight that both humour and confidence were perceived as less positive in the health mediation context. Importantly, it is of relevance to consider the opposite of for example friendliness towards users. Recent HCI work has highlighted the potential value of antagonistic AI (e.g., systems "*forcing users to confront their assumptions*" [15, p. 1]). Through a speculative workshop, the authors illustrate that the dominant AI paradigm is friendly (e.g., non-judgmental) and that alternative AI would be "*insulting, critical, [and] accusatory*" [15, p. 6]. Exploring how concepts like friendliness can be expressed and calibrated raises further discussion and challenges assumptions about default friendliness (i.e., sycophantic systems, e.g., [47]).

In summary, the design takeaways from the individual papers distil into two overarching considerations that inform the future design of conversational interactions for well-being. Public expectations of CUIs are shifting (see e.g., [97]), and new guidance is needed from HCI research—whether in the form of assessment tools [32], increasing our understanding of use patterns [116], or how to adapt CUIs to behave more appropriately towards users. As aforementioned, while the CASA paradigm [81] offered a strong theoretical lens for understanding human-computer interactions, it requires revision to account for today's CUIs and their indeterministic nature. As highlighted by Gambino et al. [42], people's interactions with computers evolve, and the CASA paradigm needs to be revised to better reflect how people engage with these systems. They propose a shift from applying behaviours from human-human to human-computer interactions (as in the original CASA paradigm) to developing human-computer-specific behaviours that are ultimately shaped by people's experiences with contemporary systems. This shift is further evidenced by recent work on playful interactions [95] observed in contemporary conversational systems. Calibrating agency in interactions with CUIs and defining their expressive range, particularly in well-being contexts, is therefore paramount, as people's current interactions shape how they will interact with CUIs in the future [42]. Next, I extend this discussion and highlight challenges and implications that arise from adapting emerging conversational interactions.

Table 5.1: Overview of design considerations distilled from the individual papers and respective design takeaways, and brief descriptions of these focused on avoiding reliance on default model settings. I use CUIs as a representative term for those used in each individual paper (e.g., chatbots and robots) and in the context of conversational interactions for well-being.

Design Consideration	Paper ID	Design Takeaway	Description
Calibrate Agency	P1	Confrontation	Adapt CUIs' use of confrontational or demanding expressions with for example more positive reassurances and comforting words.
	P2	Customisation	Adapt CUIs' to allow for advice style selection to give users more or less control.
	P3	Focal point	Adapt CUIs to either share or shift traditional facilitation roles.
	P4	Defiance	Adapt CUIs defying users when these seek (in)appropriate validation.
	P4	Disengagement	Adapt CUIs disengagement from users when (un)healthy.
	P5	Diversion	Adapt CUIs' use of diverting cues when (in)appropriate.
	P5	Correction	Adapt CUIs' use of corrective cues when users communicate (in)appropriately.
Calibrate Expressive Range	P1	Friendliness	Adapt CUI positivity by for example instructing it to ask compassionate questions whilst avoiding being too friendly.
	P1	Perspective-taking	Adapt attentiveness and perspective-taking by for example pointing out shared experiences.
	P1	Manifestation	Adapt appearance manifestations to target audience by for example expressing teen-dependent language.
	P2	Communication	Adapt CUIs' advice style to align with individual preferences or challenge them constructively.
	P3	Confidence	Adapt CUIs' use of confident communication by for example downplaying their abilities.
	P3	Humour	Adapt CUIs' use of humour to for example avoid the risk of ridiculing users.
	P5	Contextualisation	Adapt CUIs' use of additional information when failing or succeeding to meet users' requests.

5.2 Challenges in Adapting Conversational Interactions for Well-Being

As aforementioned, AI-powered CUIs grow more complex in their capabilities to interact with end-users. This has been exemplified by recent HCI research on people worrying for AI's welfare [5], judging all 'AIs' based on one [76], or perceiving consciousness in LLMs [97]. Additionally, we should expect such perceptions and expectations to further shift following systems' human-level capacities (e.g., CUIs managing conversational turn-taking [102] or carrying a theory of mind of the user [115]). As a result, people may struggle with how to interact with such systems and how to use them appropriately.

In P1, I observed that trust, likeability, and perceived safety decreased with low moral agency chatbots, which relates to recent discussions on potential benefits of making AI systems antagonistic [15]. In P2, I found that the preferred style of advice varies depending on the user, and is linked to individual characteristics, which relates to Google's recently introduced AI Overviews (i.e., AI-

curated search results) [1]. In P4, our results suggest that people care for themselves via AI-powered CUIs and perceive this as helpful. In P5, our findings indicate that people have clear preferences for how they want to be treated when their expectations are unmet. When the AI-powered CUI did not provide the users with what they expected, participants preferred the CUI to do so in a diverting way (I cannot give A but B) in contrast to a full refusal (I cannot give A). In real-world contexts, what happens with people's perceptions when CUIs take on social roles, such as mediating health information between an older mother and a younger son (as in P3)? What happens with their perceptions when such information is presented in a conversational format (as in P2) or when conversational search engines talk back? What happens when people perceive AI-powered CUIs as companions, as in P4? And how are these CUIs perceived when they refuse to answer, as in P5, or simply break down [38]?

These findings open up a broader discussion about AI-powered CUIs (e.g., general-purpose CUIs like ChatGPT) and how they convey an 'ask anything' ethos to users. This challenge is further supported by Ladak et al. who surveyed over 900 individuals, asking them to rate various AIs (e.g., robots, self-driving cars), with findings indicating that ChatGPT is perceived as possessing a degree of moral agency [64]. An important question thus arises: how do perceptions of AI-powered CUIs shift when it is positioned and designed as a well-being support tool? As there is no context or specified target audience in mind for general-purpose CUIs, users' perceptions of their interactions will be scattered and inconsistent, shaped largely by their own expectations and prior experiences.

As people's expectations and perceptions of conversational interactions for well-being evolve, the challenge lies in how we as researchers can best help shape them. Yet these challenges also present opportunities: since such interactions remain in a formative stage, they can still be intentionally shaped. As discussed in Section 5.1, the ways we interact with contemporary CUIs will shape our future interactions. Therefore, challenging assumptions about the agency users are granted and the expressive range CUIs should offer provides a concrete and promising path forward for addressing challenges that arise from emerging conversational interactions.

5.3 Implications of Conversational Well-Being Interactions

As distilled in Section 5.1 and discussed in Section 5.2, adapting conversational interactions for well-being is key to shaping how people engage with these systems while the practices are still emerging. Next, I outline three implications of conversational well-being interactions.

First, beyond learning from how end-users push the boundaries of AI use (e.g., ChatGPT for companionship), HCI research similarly advances our understanding through examples such as using AI to support online volunteers responding to suicidal individuals [107] or how people may use AI for mental health support [100]. While such applications present interesting opportunities, they also raise serious concerns about their implications. In P1, amongst other, I found that people have clear expectations for how chatbots should be designed for young people in contrast to older adults. People's considerations spanned across increased human-likeness to warmth, sensitivity, and appearance manifestation—all of which current conversational technologies offer no guarantees of

[1] `https://search.google/ways-to-search/ai-overviews/`

handling appropriately. While not specifically focused on mental health and young people, recent insights (e.g., LLMs stigmatise people) from Moore et al. support the concern that AI lacks the capacity to behave appropriately in sensitive contexts [79]. Beyond such unsafe interactions, other concerns also arise in that end-users might struggle to disengage from 'addictive' interactions with AI. [2]

Second, our findings in P1, but also P2 and P4, have implications for health-care services. In P2, I found that people prefer AI-generated advice given in a 'whimsical' style less and that their personal characteristics (e.g., personality and attitude) further influence their perception. Furthermore, I learned that people might particularly use these when engaging in sensitive topics. Similar to search engines, these CUIs allow for self-care and can help relieve pressure on health-care services; however, if they fail to meet users' expectations, they may end up further burdening those services—suggesting a potential boomerang effect. This boomerang effect might also hold for people using AI-powered CUIs to care for themselves, as investigated in P4. Our findings suggest that people use AI-powered CUIs 'successfully' across various use cases while counsellors raise several concerns about this unsupervised and unregulated practice. If people 'successfully' but unknowingly use these CUIs to (inappropriately) care for themselves, CUIs may reinforce harmful practices under the guise of effective support, ultimately complicating rather than easing the pressure on health-care services. This concern is further supported by recent findings from a study examining people's reactions to ChatGPT's unpredictable behaviour, which revealed that such behaviour can evoke uneasiness and fear toward AI [94].

Third, our findings from P1, P2, P4, and particularly P3 have implications for interpersonal communication. Results from P3, similar to part of the findings in P1 and P2, indicate that people perceive CUIs as promising to share and mediate sensitive health topics (such as mental or sexual health information) and that this particularly holds for CUIs communicating and mediating information in more self-deprecating ways. While CUIs such as the one used in P3 can be useful and even relieve family members (e.g., mothers) from being solely responsible for other family members to communicate, Fu et al.'s work on AI-mediated communication for interpersonal communication [41] or relationships [40] provides early insights on the potential harms of such applications (e.g., reduce user autonomy and increase dependence on AI). We know little about the potential effects of AI-powered CUIs designed to mediate human–human well-being interactions in real-world contexts, and the consequences that may arise.

5.4 Limitations and Future Work

There are several limitations to this work. First, I did not study people's perceptions of interactions for well-being across time. Longitudinal studies likely provide more insights into how people's perceptions might shift (e.g., following novelty effects or when expectations people hold are not met). Second, while many use AI-powered CUIs for well-being, the technology is novel and those using it more heavily and for more complex tasks (e.g., mental health interventions) are likely early adopters. Other user groups might therefore report different experiences as compared to these

[2] `https://www.media.mit.edu/articles/we-need-to-prepare-for-addictive-intelligence/`

technology enthusiasts. Relatedly, and third, evaluating conversational interactions for well-being with relevant target audiences is essential in real-world contexts—particularly where AI-powered CUIs are already in use but simultaneously under extensive development.

People increasingly use AI-powered CUIs to support their well-being, whether by talking about their day or seeking advice on how to care for their mental health. A growing body of research in HCI as well as clinical sciences is examining the effects of such interactions, as well as how to design and develop CUIs that meaningfully support well-being. However, we are still in the early stages of understanding the broader impact of this paradigm shift and how these systems can best be conceptualised to maximise appropriateness of the support they provide. For example, we know little of the potential long-term effects of conversational interactions for well-being: are they supportive or can they even be addictive or abusive [1]? As CUIs grow increasingly complex, a key area for future research is to develop a deeper understanding of how to effectively study conversational interactions for well-being, with several design considerations outlined in this thesis (see Table 5.1). First, although CUIs are used for well-being conversations by many, they have not yet become something your 'average next-door neighbour' would use (e.g., for companionship). While prior work suggests that CUIs hold promise for reducing mental health stigma [66], more recent research indicates that people often conceal their use of AI-powered CUIs [132]. This highlights the need to consider what methodological approaches are best suited for reaching relevant target audiences to study both the opportunities and risks associated with their use of these systems. Second, by identifying and engaging with these populations, we can begin to understand how to further adapt emerging conversational interactions for well-being, particularly in ways that steer clear of negatively impacting people's well-being (e.g., reinforcing stigma or causing unintended harm).

6. CONCLUSION

I have positioned this thesis around emerging conversational interactions for well-being. Given major technological developments, such interactions require a revised understanding of conversational user experiences. I therefore asked the following question: How can emerging conversational interactions be adapted to support everyday people's well-being?

People's perceptions of interactions with CUIs are influenced by the specific well-being context in which they occur. While AI-powered CUIs are increasingly used, with many utilising these for well-being interactions, we lack a rigid understanding of how these might further alter people's perceptions. I provide insights into how people's perceptions are shaped by the context in which AI-powered CUIs operate and by how these interfaces behave toward users. These findings, together with recent HCI work on mind perceptions, indicate that as CUIs increase in complexity, people's perceptions of conversational interactions with them also evolve. As discussed in Section 5, CASA may thus no longer serve as the most suitable theoretical lens for understanding and studying these perceptions, as interactions with AI-powered CUIs introduce new layers of complexity as seen in recent HCI research on 'gulf of envisioning'. As aforementioned, more recent research points to the value of looking at the interactions people currently have with systems and how those interactions shape any future interactions people have with these or similar systems. Following the two design considerations together with their respective design takeaways is therefore a step towards shaping conversational interactions for well-being people currently have with AI-powered CUIs. By adapting the interactions people have with contemporary AI-powered CUIs, people's now scattered and inconsistent use patterns can be shaped to steer how they will interact with similar CUIs in the future. Such adaptations are important for conversational well-being interactions, as these may lead to unhealthy or even harmful outcomes we have yet to imagine.

REFERENCES

[1] D. Adam, "Supportive? Addictive? Abusive? How AI companions affect our mental health ," 2025. [Online]. Available: https://www.nature.com/articles/d41586-025-01349-9

[2] R. Albers and M. Hassenzahl, "Let's Talk About Death: Existential Conversations with Chatbots," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[3] L. Alberts, U. Lyngs, and M. Van Kleek, "Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces that Act Socially," *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. CSCW1, Apr. 2024.

[4] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz, "Guidelines for Human-AI Interaction," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, 2019, p. 1–13.

[5] J. R. Anthis, J. V. Pauketat, A. Ladak, and A. Manoli, "Perceptions of Sentient AI and Other Digital Minds: Evidence from the AI, Morality, and Sentience (AIMS) Survey," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, 2025.

[6] Anthropic, "Claude Opus 4 and 4.1 can now end a rare subset of conversations," 2025. [Online]. Available: https://www.anthropic.com/research/end-subset-conversations

[7] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz, "Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, 2019, p. 1–12.

[8] P. B. Bae Brandtzæg, M. Skjuve, K. K. Kristoffer Dysthe, and A. Følstad, "When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21, 2021.

[9] A. Bhattacharjee, J. Pang, A. Liu, A. Mariakakis, and J. J. Williams, "Design Implications for One-Way Text Messaging Services that Support Psychological Wellbeing," *ACM Trans. Comput.-Hum. Interact.*, vol. 30, no. 3, 2023.

[10] A. Bhattacharjee, J. J. Williams, J. Meyerhoff, H. Kumar, A. Mariakakis, and R. Kornfield, "Investigating the Role of Context in the Delivery of Text Messages for Supporting Psychological Wellbeing," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, 2023.

[11] J. Binda, C. W. Yuan, N. Cope, H. Park, E. K. Choe, and J. M. Carroll, "Supporting Effective Sharing of Health Information among Intergenerational Family Members," in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth '18, 2018, p. 148–157.

[12] R. Bowman, O. Cooney, J. W. Newbold, A. Thieme, L. Clark, G. Doherty, and B. Cowan, "Exploring how politeness impacts the user experience of chatbots for mental health support," *International Journal of Human-Computer Studies*, vol. 184, p. 103181, 2024.

[13] V. Braun and V. C. and, "Reflecting on reflexive thematic analysis," *Qualitative Research in Sport, Exercise and Health*, vol. 11, no. 4, pp. 589–597, 2019.

[14] V. Braun and V. Clarke, "Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches," *Counselling and Psychotherapy Research*, vol. 21, no. 1, pp. 37–47, 2021.

[15] A. Cai, I. Arawjo, and E. L. Glassman, "Antagonistic AI," 2024. [Online]. Available: https://arxiv.org/abs/2402.07350

[16] W. Cai, Y. Jin, and L. Chen, "Impacts of Personal Characteristics on User Trust in Conversational Recommender Systems," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22, 2022.

[17] T. Capel, B. Ploderer, F. Bircanin, S. Hanmer, J. P. Yates, J. Wang, K. L. Khor, T. W. Leong, G. Wadley, and M. Newcomb, "Studying Self-Care with Generative AI Tools: Lessons for Design," in *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, ser. DIS '24, 2024, p. 1620–1637.

[18] J. Ceha, K. J. Lee, E. Nilsen, J. Goh, and E. Law, "Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes?" in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21, 2021.

[19] A. P. Chaves and M. A. G. and, "How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design," *International Journal of Human–Computer Interaction*, vol. 37, no. 8, pp. 729–758, 2021.

[20] D. Chen, K. Chauhan, R. Parsa, Z. A. Liu, F.-F. Liu, E. Mak, L. Eng, B. L. Hannon, J. Croke, A. Hope, N. Fallah-Rad, P. Wong, and S. Raman, "Patient perceptions of empathy in physician and artificial intelligence chatbot responses to patient questions about cancer," *npj Digital Medicine*, vol. 8, no. 1, p. 275, 2025.

[21] X. Chen, L. Tankelevitch, R. Vanukuru, A. E. Scott, P. Panda, and S. Rintel, "Are We On Track? AI-Assisted Active and Passive Goal Reflection During Meetings," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, 2025.

[22] H. Chin, L. W. Molefi, and M. Y. Yi, "Empathy Is All You Need: How a Conversational Agent Should Respond to Verbal Abuse," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20, 2020, p. 1–13.

[23] D. Choi, S. Lee, S.-I. Kim, K. Lee, H. J. Yoo, S. Lee, and H. Hong, "Unlock Life with a Chat(GPT): Integrating Conversational AI with Large Language Models into Everyday Lives of Autistic Individuals," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[24] J. C. L. Chow, L. Sanders, and K. Li, "Impact of ChatGPT on medical chatbots as a disruptive technology," *Frontiers in Artificial Intelligence*, vol. Volume 6 - 2023, 2023.

[25] C. Claisse, B. Kasadha, S. Stumpf, and A. C. Durrant, "Investigating Daily Practices of Self-care to Inform the Design of Supportive Health Technologies for Living and Ageing Well with HIV," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22, 2022.

[26] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu, V. Wade, and B. R. Cowan, "What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, 2019, p. 1–12.

[27] H. Cramer, V. Evers, T. van Slooten, M. Ghijsen, and B. Wielinga, "Trying too hard: effects of mobile agents' (Inappropriate) social expressiveness on trust, affect and compliance," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10, 2010, p. 1471–1474.

[28] V. Das Swain, L. Gao, A. Mondal, G. D. Abowd, and M. De Choudhury, "Sensible and Sensitive AI for Worker Wellbeing: Factors that Inform Adoption and Resistance for Information Workers," in

Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, ser. CHI '24, New York, NY, USA, 2024.

[29] M. de Gennaro, E. G. Krumhuber, and G. Lucas, "Effectiveness of an Empathic Chatbot in Combating Adverse Effects of Social Exclusion on Mood," *Frontiers in Psychology*, vol. Volume 10 - 2019, 2020.

[30] S. Desai, M. Dubiel, and L. A. Leiva, "Examining Humanness as a Metaphor to Design Voice User Interfaces," in *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, 2024.

[31] H. J. Do, H.-K. Kong, P. Tetali, J. Lee, and B. P. Bailey, "To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CSCW1, Apr. 2023.

[32] P. R. Doyle, I. Gessinger, J. Edwards, L. Clark, O. Dumbleton, D. Garaialde, D. Rough, A. Bleakley, H. P. Branigan, and B. R. Cowan, "The Partner Modelling Questionnaire: A validated self-report measure of perceptions toward machines as dialogue partners," *ACM Trans. Comput.-Hum. Interact.*, 2025.

[33] M. Dubiel, M. Halvey, P. O. Gallegos, and S. King, "Persuasive Synthetic Speech: Voice Perception and User Behaviour," in *Proceedings of the 2nd Conference on Conversational User Interfaces*, ser. CUI '20, 2020.

[34] K. J. K. Feng, Q. V. Liao, Z. Xiao, J. Wortman Vaughan, A. X. Zhang, and D. W. McDonald, "Canvil: Designerly Adaptation for LLM-Powered User Experiences," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, 2025.

[35] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial," *JMIR Ment Health*, vol. 4, no. 2, p. e19, Jun 2017.

[36] C. Flathmann, N. J. McNeese, B. Schelble, B. Knijnenburg, and G. F. and, "Understanding the impact and design of AI teammate etiquette," *Human–Computer Interaction*, vol. 39, no. 5-6, pp. 444–471, 2024.

[37] A. Følstad and P. B. Brandtzæg, "Chatbots and the new world of HCI," *Interactions*, vol. 24, no. 4, p. 38–42, 2017.

[38] A. Følstad, E. L.-C. Law, and N. van As, "Conversational Breakdown in a Customer Service Chatbot: Impact of Task Order and Criticality on User Trust and Emotion," *ACM Trans. Comput.-Hum. Interact.*, vol. 31, no. 5, Nov. 2024.

[39] J. Forlizzi and K. Battarbee, "Understanding experience in interactive systems," in *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, ser. DIS '04, 2004, p. 261–268.

[40] Y. Fu, Y. Chen, Z. G. D. C. Lai, and A. Hiniker, "Should ChatGPT Write Your Breakup Text? Exploring the Role of AI in Relationship Dissolution," 2024. [Online]. Available: https://arxiv.org/abs/2401.09695

[41] Y. Fu, S. Foell, X. Xu, and A. Hiniker, "From Text to Self: Users' Perception of AIMC Tools on Interpersonal Communication and Self," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[42] A. Gambino, J. Fox, and R. A. Ratan, "Building a stronger CASA: Extending the computers are social actors paradigm," *Human-Machine Communication*, vol. 1, p. 71–85, 2020.

[43] M. R. Haque and S. Rubya, ""For an App Supposed to Make Its Users Feel Better, It Sure is a Joke" - An Analysis of User Reviews of Mobile Mental Health Applications," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, Nov. 2022.

[44] M. Hertzum and K. Hornbæk, "Frustration: Still a Common User Experience," *ACM Trans. Comput.-Hum. Interact.*, vol. 30, no. 3, Jun. 2023.

[45] E. Heyselaar, "The CASA theory no longer applies to desktop computers," *Scientific Reports*, vol. 13, no. 1, p. 19693, 2023.

[46] M. Houben, N. van As, N. Sawhney, D. Unbehaun, and M. Lee, "Participatory Design for Whom? Designing Conversational User Interfaces for Sensitive Settings and Vulnerable Populations," in *Proceedings of the 5th International Conference on Conversational User Interfaces*, ser. CUI '23, 2023.

[47] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, and T. Liu, "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Trans. Inf. Syst.*, vol. 43, no. 2, Jan. 2025.

[48] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderbäck, "Technology probes: inspiring design for and with families," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03, 2003, p. 17–24.

[49] R. M. Jacobsen, S. R. Cox, C. F. Griggio, and N. van Berkel, "Chatbots for Data Collection in Surveys: A Comparison of Four Theory-Based Interview Probes," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, 2025.

[50] M. Jain, P. Kumar, R. Kota, and S. N. Patel, "Evaluating and Informing the Design of Chatbots," in *Proceedings of the 2018 Designing Interactive Systems Conference*, ser. DIS '18, 2018, p. 895–906.

[51] M. Jakesch, A. Bhat, D. Buschek, L. Zalmanson, and M. Naaman, "Co-Writing with Opinionated Language Models Affects Users' Views," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, 2023.

[52] M. Jakesch, M. French, X. Ma, J. T. Hancock, and M. Naaman, "AI-Mediated Communication: How the Perception that Profile Text was Written by AI Affects Trustworthiness," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, 2019, p. 1–13.

[53] E. Jo, D. A. Epstein, H. Jung, and Y.-H. Kim, "Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, 2023.

[54] I. Kaate, J. Salminen, S.-G. Jung, T. T. T. Xuan, E. Häyhänen, J. Y. Azem, and B. J. Jansen, ""You Always Get an Answer": Analyzing Users' Interaction with AI-Generated Personas Given Unanswerable Questions and Risk of Hallucination," in *Proceedings of the 30th International Conference on Intelligent User Interfaces*, ser. IUI '25, 2025, p. 1624–1638.

[55] S. S. Y. Kim, Q. V. Liao, M. Vorvoreanu, S. Ballard, and J. W. Vaughan, ""I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust," in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '24, 2024, p. 822–835.

[56] T. Kim, M. D. Molina, M. M. Rheu, E. S. Zhan, and W. Peng, "One AI Does Not Fit All: A Cluster Analysis of the Laypeople's Perception of AI Roles," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, 2023.

[57] T. Kim, S. Bae, H. A. Kim, S.-W. Lee, H. Hong, C. Yang, and Y.-H. Kim, "MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[58] T. Kim, D. Shin, Y.-H. Kim, and H. Hong, "DiaryMate: Understanding User Perceptions and Experience in Human-AI Collaboration for Personal Journaling," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[59] Y. Kim, J. Lee, S. Kim, J. Park, and J. Kim, "Understanding Users' Dissatisfaction with ChatGPT Responses: Types, Resolving Tactics, and the Effect of Knowledge Level," in *Proceedings of the 29th International Conference on Intelligent User Interfaces*, ser. IUI '24, 2024, p. 385–404.

[60] C. Kobiella, Y. S. Flores López, F. Waltenberger, F. Draxler, and A. Schmidt, ""If the Machine Is As Good As Me, Then What Use Am I?" – How the Use of ChatGPT Changes Young Professionals' Perception of Productivity and Accomplishment," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[61] R. Kocielnik, R. Langevin, J. S. George, S. Akenaga, A. Wang, D. P. Jones, A. Argyle, C. Fockele, L. Anderson, D. T. Hsieh, K. Yadav, H. Duber, G. Hsieh, and A. L. Hartzler, "Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health," in *Proceedings of the 3rd Conference on Conversational User Interfaces*, ser. CUI '21, 2021.

[62] R. Kocielnik, L. Xiao, D. Avrahami, and G. Hsieh, "Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 2, Jul. 2018.

[63] R. Kornfield, J. Meyerhoff, H. Studd, A. Bhattacharjee, J. J. Williams, M. Reddy, and D. C. Mohr, "Meeting Users Where They Are: User-centered Design of an Automated Text Messaging Tool to Support the Mental Health of Young Adults," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22, 2022.

[64] A. Ladak, M. Wilks, S. Loughnan, and J. R. Anthis, "Robots, Chatbots, Self-Driving Cars: Perceptions of Mind and Morality Across Artificial Intelligences," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, 2025.

[65] M. Lee, S. Ackermans, N. van As, H. Chang, E. Lucas, and W. IJsselsteijn, "Caring for Vincent: A Chatbot for Self-Compassion," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19, 2019, p. 1–13.

[66] Y.-C. Lee, Y. Cui, J. Jamieson, W. Fu, and N. Yamashita, "Exploring Effects of Chatbot-based Social Contact on Reducing Mental Illness Stigma," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, 2023.

[67] Y.-C. Lee, N. Yamashita, and Y. Huang, "Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW1, May 2020.

[68] Y.-C. Lee, N. Yamashita, Y. Huang, and W. Fu, ""I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20, 2020, p. 1–12.

[69] C. LeFevre and C.-F. Chung, "New Understandings of Loss: Examining the Role of Reflective Technology Within Bereavement and Meaning-Making," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[70] B. Li, O. Gross, N. Crampton, M. Kapoor, S. Tauseef, M. Jain, K. N. Truong, and A. Mariakakis, "Beyond the Waiting Room: Patient's Perspectives on the Conversational Nuances of Pre-Consultation Chatbots," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[71] Y. Li, X. Ding, Y. Chen, Y. Li, and N. Ma, "Customizable AI for Depression Care: Improving the User Experience of Large Language Model-Driven Chatbots," in *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, ser. DIS '25, 2025, p. 1844–1866.

[72] Q. V. Liao, M. Mas-ud Hussain, P. Chandar, M. Davis, Y. Khazaeni, M. P. Crasso, D. Wang, M. Muller, N. S. Shami, and W. Geyer, "All Work and No Play?" in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18, 2018, p. 1–13.

[73] D. Long and B. Magerko, "What is AI Literacy? Competencies and Design Considerations," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20, 2020, p. 1–16.

[74] E. Luger and A. Sellen, ""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*, ser. CHI '16, 2016, p. 5286–5297.

[75] Z. Ma, Y. Mei, Y. Long, Z. Su, and K. Z. Gajos, "Evaluating the Experience of LGBTQ+ People Using Large Language Model Based Chatbots for Mental Health Support," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[76] A. Manoli, J. V. T. Pauketat, and J. R. Anthis, "The AI Double Standard: Humans Judge All AIs for the Actions of One," *Proc. ACM Hum.-Comput. Interact.*, vol. 9, no. 2, May 2025.

[77] A. Manzini, G. Keeling, L. Alberts, S. Vallor, M. R. Morris, and I. Gabriel, "The Code That Binds Us: Navigating the Appropriateness of Human-AI Assistant Relationships," *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, vol. 7, no. 1, pp. 943–957, Oct. 2024.

[78] J. Meyerhoff, R. Kornfield, D. C. Mohr, and M. Reddy, "Meeting Young Adults' Social Support Needs across the Health Behavior Change Journey: Implications for Digital Mental Health Tools," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, Nov. 2022.

[79] J. Moore, D. Grabb, W. Agnew, K. Klyman, S. Chancellor, D. C. Ong, and N. Haber, "Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers," in *Proceedings of the 2025 FACCT Conference*, ser. FACCT '25, 2025.

[80] S. Murphy, A. Hickey, D. Peelo Dennehy, K. Morrissey, J. McCarthy, and S. Foley, "Sharing, Support-Seeking, and Managing Safety: A Qualitative Study of Online Platform Engagement After Pregnancy Loss," *Proc. ACM Hum.-Comput. Interact.*, vol. 9, no. 2, May 2025.

[81] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '94, 1994, p. 72–78.

[82] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '90, 1990, p. 249–256.

[83] F. Nunes and G. Fitzpatrick, "Understanding the Mundane Nature of Self-care: Ethnographic Accounts of People Living with Parkinson's," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18, 2018, p. 1–15.

[84] F. Nunes, N. Verdezoto, G. Fitzpatrick, M. Kyng, E. Grönvall, and C. Storni, "Self-Care Technologies in HCI: Trends, Tensions, and Opportunities," *ACM Trans. Comput.-Hum. Interact.*, vol. 22, no. 6, Dec. 2015.

[85] S. Nørskov, M. F. Damholdt, J. P. Ulhøi, M. B. Jensen, C. Ess, and J. Seibt, "Applicant Fairness Perceptions of a Robot-Mediated Job Interview: A Video Vignette-Based Experimental Survey," *Frontiers in Robotics and AI*, vol. 7, 2020.

[86] H. L. O'Brien, I. Roll, A. Kampen, and N. Davoudi, "Rethinking (Dis)engagement in human-computer interaction," *Computers in Human Behavior*, vol. 128, p. 107109, 2022.

[87] W. H. Organization, "Health Promotion Glossary of Terms 2021," 2021. [Online]. Available: https://iris.who.int/bitstream/handle/10665/350161/9789240038349-eng.pdf?sequence=1

[88] R. Y. Pang, H. Schroeder, K. S. Smith, S. Barocas, Z. Xiao, E. Tseng, and D. Bragg, "Understanding the LLM-ification of CHI: Unpacking the Impact of LLMs at CHI through a Systematic Literature Review," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, 2025.

[89] G. W. W. Park, P. Panda, L. Tankelevitch, and S. Rintel, "The CoExplorer Technology Probe: A Generative AI-Powered Adaptive Interface to Support Intentionality in Planning and Running Video Meetings," in *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, ser. DIS '24, 2024, p. 1638–1657.

[90] H. Park and D. Ahn, "The Promise and Peril of ChatGPT in Higher Education: Opportunities, Challenges, and Design Implications," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[91] H. Park and J. Lee, "Designing a Conversational Agent for Sexual Assault Survivors: Defining Burden of Self-Disclosure and Envisioning Survivor-Centered Solutions," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21, 2021.

[92] K. Pu, D. Lazaro, I. Arawjo, H. Xia, Z. Xiao, T. Grossman, and Y. Chen, "Assistance or Disruption? Exploring and Evaluating the Design and Trade-offs of Proactive AI Programming Support," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, 2025.

[93] A. Rapp, L. Curti, and A. Boldi, "The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots," *International Journal of Human-Computer Studies*, vol. 151, 2021.

[94] A. Rapp, C. Di Lodovico, and L. Di Caro, "How do people react to ChatGPT's unpredictable behavior? Anthropomorphism, uncanniness, and fear of AI: A qualitative study on individuals' perceptions and understandings of LLMs' nonsensical hallucinations," *International Journal of Human-Computer Studies*, vol. 198, 2025.

[95] M. Ronagh Nikghalb and J. Cheng, "Interrogating AI: Characterizing Emergent Playful Interactions with ChatGPT," *Proc. ACM Hum.-Comput. Interact.*, vol. 9, no. 2, May 2025.

[96] S. I. Ross, F. Martinez, S. Houde, M. Muller, and J. D. Weisz, "The Programmer's Assistant: Conversational Interaction with a Large Language Model for Software Development," in *Proceedings of the 28th International Conference on Intelligent User Interfaces*, ser. IUI '23, 2023, p. 491–514.

[97] A. E. Scott, D. Neumann, J. Niess, and P. W. Woźniak, "Do You Mind? User Perceptions of Machine Consciousness," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, 2023.

[98] A. E. Scott, L. Tankelevitch, P. Panda, R. Vanukuru, X. Chen, and S. Rintel, "What Does Success Look Like? Catalyzing Meeting Intentionality with AI-Assisted Prospective Reflection," 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2505.14370

[99] V. Setlur and M. Tory, "How do you Converse with an Analytical Chatbot? Revisiting Gricean Maxims for Designing Analytical Conversational Behavior," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22, 2022.

[100] A. Sharma, K. Rushton, I. W. Lin, T. Nguyen, and T. Althoff, "Facilitating Self-Guided Mental Health Interventions Through Human-Language Model Interaction: A Case Study of Cognitive Restructuring," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[101] R. Shelby, S. Rismani, and N. Rostamzadeh, "Generative AI in Creative Practice: ML-Artist Folk Theories of T2I Use, Harm, and Harm-Reduction," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[102] G. Skantze, "Turn-taking in Conversational Systems and Human-Robot Interaction: A Review," *Computer Speech & Language*, vol. 67, p. 101178, 2021.

[103] M. Skjuve, A. Følstad, and P. B. Brandtzaeg, "The User Experience of ChatGPT: Findings from a Questionnaire Study of Early Users," in *Proceedings of the 5th International Conference on Conversational User Interfaces*, ser. CUI '23, 2023.

[104] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg, "A longitudinal study of human–chatbot relationships," *International Journal of Human-Computer Studies*, vol. 168, p. 102903, 2022.

[105] W. Smith, G. Wadley, S. Webber, B. Tag, V. Kostakos, P. Koval, and J. J. Gross, "Digital Emotion Regulation in Everyday Life," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22, 2022.

[106] I. Song, S. Park, S. R. Pendse, J. L. Schleider, M. De Choudhury, and Y.-H. Kim, "ExploreSelf: Fostering User-driven Exploration and Reflection on Personal Challenges with Adaptive Guidance by Large Language Models," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25, 2025.

[107] L. Stapleton, S. Liu, C. Liu, I. Hong, S. Chancellor, R. E. Kraut, and H. Zhu, ""If This Person is Suicidal, What Do I Do?": Designing Computational Approaches to Help Online Volunteers Respond to Suicidality," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[108] H. Subramonyam, R. Pea, C. Pondoc, M. Agrawala, and C. Seifert, "Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[109] D. Talypova, A. Lingler, and P. Wintersberger, "User-Centered Investigation of Features for Attention Management Systems in an Online Vignette Study," in *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '23, 2023, p. 108–121.

[110] L. Tankelevitch, V. Kewenig, A. Simkute, A. E. Scott, A. Sarkar, A. Sellen, and S. Rintel, "The Metacognitive Demands and Opportunities of Generative AI," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[111] C. Torrey, S. Fussell, and S. Kiesler, "How a Robot Should Give Advice," in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '13, 2013, p. 275–282.

[112] N. van Berkel, J. Goncalves, D. Russo, S. Hosio, and M. B. Skov, "Effect of Information Presentation on Fairness Perceptions of Machine Learning Predictors," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21, 2021.

[113] S. T. Völkel and L. Kaya, "Examining User Preference for Agreeableness in Chatbots," in *Proceedings of the 3rd Conference on Conversational User Interfaces*, ser. CUI '21, 2021.

[114] S. T. Völkel, R. Schoedel, L. Kaya, and S. Mayer, "User Perceptions of Extraversion in Chatbots after Repeated Use," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22, 2022.

[115] Q. Wang, K. Saha, E. Gregori, D. Joyner, and A. Goel, "Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21, 2021.

[116] X. Wang, X. Wang, S. Park, and Y. Yao, "Mental Models of Generative AI Chatbot Ecosystems," in *Proceedings of the 30th International Conference on Intelligent User Interfaces*, ser. IUI '25, 2025, p. 1016–1031.

[117] J. D. Weisz, J. He, M. Muller, G. Hoefer, R. Miles, and W. Geyer, "Design Principles for Generative AI Applications," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[118] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, p. 36–45, 1966.

[119] Y. Wærn and R. Ramberg, "People's perception of human and computer advice," *Computers in Human Behavior*, vol. 12, no. 1, pp. 17–27, 1996.

[120] T. Xu, J. Yu, D. T. Doyle, and S. Voida, "Technology-Mediated Strategies for Coping with Mental Health Challenges: Insights from People with Bipolar Disorder," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CSCW2, Oct. 2023.

[121] A. Xygkou, C. S. Ang, P. Siriaraya, J. P. Kopecki, A. Covaci, E. Kanjo, and W.-J. She, "MindTalker: Navigating the Complexities of AI-Enhanced Social Engagement for People with Early-Stage Dementia," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[122] A. Xygkou, P. Siriaraya, A. Covaci, H. G. Prigerson, R. Neimeyer, C. S. Ang, and W.-J. She, "The "Conversation" about Loss: Understanding How Chatbot Technology was Used in Supporting People in Grief," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, 2023.

[123] K. Yatani, *Effect Sizes and Power Analysis in HCI*. Springer International Publishing, 2016, pp. 87–110.

[124] S.-F. Yeh, M.-H. Wu, T.-Y. Chen, Y.-C. Lin, X. Chang, Y.-H. Chiang, and Y.-J. Chang, "How to Guide Task-oriented Chatbot Users, and When: A Mixed-methods Study of Combinations of Chatbot Guidance Types and Timings," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22, 2022.

[125] D. W. Yoo, J. M. Shi, V. J. Rodriguez, and K. Saha, "AI Chatbots for Mental Health: Values and Harms from Lived Experiences of Depression," 2025. [Online]. Available: https://arxiv.org/abs/2504.18932

[126] D. W. Yoo, H. Woo, V. C. Nguyen, M. L. Birnbaum, K. P. Kruzan, J. G. Kim, G. D. Abowd, and M. De Choudhury, "Patient Perspectives on AI-Driven Predictions of Schizophrenia Relapses: Understanding Concerns and Opportunities for Self-Care and Treatment," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[127] Y. You, C.-H. Tsai, Y. Li, F. Ma, C. Heron, and X. Gui, "Beyond Self-diagnosis: How a Chatbot-based Symptom Checker Should Respond," *ACM Trans. Comput.-Hum. Interact.*, vol. 30, no. 4, Sep. 2023.

[128] J. Young, L. M. Jawara, D. N. Nguyen, B. Daly, J. Huh-Yoo, and A. Razi, "The Role of AI in Peer Support for Young People: A Study of Preferences for Human- and AI-Generated Responses," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[129] J. Zamora, "I'm Sorry, Dave, I'm Afraid I Can't Do That: Chatbot Perception and Expectations," in *Proceedings of the 5th International Conference on Human Agent Interaction*, ser. HAI '17, 2017, p. 253–260.

[130] N. Zargham, L. Reicherts, M. Bonfert, S. T. Voelkel, J. Schoening, R. Malaka, and Y. Rogers, "Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma," in *Proceedings of the 4th Conference on Conversational User Interfaces*, ser. CUI '22, 2022.

[131] A. W. Zhang, T.-H. Lin, X. Zhao, and S. Sebo, "Ice-Breaking Technology: Robots and Computers Can Foster Meaningful Connections between Strangers through In-Person Conversations," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, 2023.

[132] Z. Zhang, C. Shen, B. Yao, D. Wang, and T. Li, "Secret Use of Large Language Model (LLM)," *Proc. ACM Hum.-Comput. Interact.*, vol. 9, no. 2, May 2025.

APPENDIX

The appendix of the thesis constitutes the papers:

1. **Joel Wester**, Henning Pohl, Simo Hosio, Niels van Berkel. *"This Chatbot Would Never…": Perceived Moral Agency of Mental Health Chatbots*. In Proceedings of the 2024 ACM on Human-Computer Interaction (CSCW1). https://doi.org/10.1145/3637410

2. **Joel Wester**, Sander de Jong, Henning Pohl, Niels van Berkel. *Exploring People's Perceptions of LLM-generated Advice*. Computers in Human Behavior: Artificial Humans, 2024. https://doi.org/10.1016/j.chbah.2024.100072

3. **Joel Wester**, Bhakti Moghe, Katie Winkle, Niels van Berkel. *Facing LLMs: Robot Communication Styles in Mediating Health Information between Parents and Young Adults*. In Proceedings of the 2024 ACM on Human-Computer Interaction (CSCW2). https://doi.org/10.1145/3687036

4. **Joel Wester**, Sander de Jong, Henning Pohl, Niels van Berkel. *Using LLMs for Self-Care: User and Counsellor Perspectives*. International Journal of Human-Computer Studies. https://doi.org/10.1016/j.ijhcs.2025.103589

5. **Joel Wester**, Tim Schrills, Henning Pohl, Niels van Berkel. *"As an AI language model, I cannot": Investigating LLM Denials of User Requests*. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI'24) https://doi.org/10.1145/3613904.3642135

Paper 1

"This Chatbot Would Never…": Perceived Moral Agency of Mental Health Chatbots

Abstract

Despite repeated reports of socially inappropriate and dangerous chatbot behaviour, chatbots are increasingly used as mental health services in providing support for young people. In sensitive settings as such, the notion of perceived moral agency (PMA) is crucial, given its critical role in human-human interactions. In this paper, we investigate the role of PMA in human-chatbot interactions. Specifically, we seek to understand how PMA influence the perception of trust, likeability, and perceived safety of chatbots for mental health across two distinct age groups. We conduct an online experiment (N = 279) to evaluate chatbots with low and high PMA as targeted towards teenagers and adults. Our results indicate increased trust, likeability, and perceived safety in mental health chatbots displaying high PMA. A qualitative analysis revealed four themes, assessing participants' expectations of mental health chatbots in general, as well as targeted towards teenagers: Anthropomorphism, Warmth, Sensitivity, and Appearance manifestation. We show that PMA plays a crucial role in influencing the perceptions of chatbots and provide recommendations for designing socially appropriate mental health chatbots.

Joel Wester, Henning Pohl, Simo Hosio, Niels van Berkel. *"This Chatbot Would Never…": Perceived Moral Agency of Mental Health Chatbots.* In Proceedings of the 2024 ACM on Human-Computer Interaction (CSCW1). https://doi.org/10.1145/3637410

Paper 2

Exploring people's perceptions of LLM-generated advice

Abstract

When searching and browsing the web, more and more of the information we encounter is generated or mediated through large language models (LLMs). This can be looking for a recipe, getting help on an essay, or looking for relationship advice. Yet, there is limited understanding of how individuals perceive advice provided by these LLMs. In this paper, we explore people's perception of LLM-generated advice, and what role diverse user characteristics (i.e., personality and technology readiness) play in shaping their perception. Further, as LLM-generated advice can be difficult to distinguish from human advice, we assess the perceived creepiness of such advice. To investigate this, we run an exploratory study (N = 91), where participants rate advice in different styles (generated by GPT-3.5 Turbo). Notably, our findings suggest that individuals who identify as more agreeable tend to like the advice more and find it more useful. Further, individuals with higher technological insecurity are more likely to follow and find the advice more useful, and deem it more likely that a friend could have given the advice. Lastly, we see that advice given in a 'skeptical' style was rated most unpredictable, and advice given in a 'whimsical' style was rated least malicious—indicating that LLM advice styles influence user perceptions. Our results also provide an overview of people's considerations on likelihood, receptiveness, and what advice they are likely to seek from these digital assistants. Based on our results, we provide design takeaways for LLM-generated advice and outline future research directions to further inform the design of LLM-generated advice for support applications targeting people with diverse expectations and needs.

Joel Wester, Sander de Jong, Henning Pohl, Niels van Berkel. *Exploring People's Perceptions of LLM-generated Advice*. Computers in Human Behavior: Artificial Humans, 2024. https://doi.org/10.1016/j.chbah.2024.100072

Paper 3

Facing LLMs: Robot Communication Styles in Mediating Health Information between Parents and Young Adults

Abstract

Young adults may feel embarrassed when disclosing sensitive information to their parents, while parents might similarly avoid sharing sensitive aspects of their lives with their children. How to design interactive interventions that are sensitive to the needs of both younger and older family members in mediating sensitive information remains an open question. In this paper, we explore the integration of large language models (LLMs) with social robots. Specifically, we use GPT-4 to adapt different Robot Communication Styles (RCS) for a social robot mediator designed to elicit self-disclosure and mediate health information between parents and young adults living apart. We design and compare four literature-informed RCS: three LLM-adapted (Humorous, Self-deprecating, and Persuasive) and one manually created (Human-scripted), and assess participant perceptions of Likeability, Usefulness, Helpfulness, Relatedness, and Interpersonal Closeness. Through an online experiment with 183 participants, we assess the RCS across two groups: adults with children (Parents) and young adults without children (Young Adults). Our results indicate that both Parents and Young Adults favoured the Human-scripted and Self-deprecating RCS as compared to the other two RCS. The Self-deprecating RCS furthermore led to increased relatedness as compared to the Humorous RCS. Our qualitative findings reveal challenges people have in disclosing health information to family members, and who normally assumes the role of family facilitator-two areas in which social robots can play a key role. The findings offer insights for integrating LLMs with social robots in health-mediation and other contexts involving the sharing of sensitive information.

Joel Wester, Bhakti Moghe, Katie Winkle, Niels van Berkel. *Facing LLMs: Robot Communication Styles in Mediating Health Information between Parents and Young Adults*. In Proceedings of the 2024 ACM on Human-Computer Interaction (CSCW2). https://doi.org/10.1145/3687036

Paper 4

Using LLMs for Self-Care: User and Counsellor Perspectives

Abstract

People are increasingly relying on technology for self-care, including, more recently, seeking help through conversational interfaces driven by large language models (LLMs). Yet, how interaction with LLMs has impacted people's self-care processes is not well understood. Therefore, we collected 405 user stories posted on Reddit about using LLMs for self-care. We identified four key themes on how people use LLMs for this purpose: Letting go, Finding comfort, Building up, and Reflecting on. We interviewed twelve counsellors to capture their perspectives on this practice, given their professional expertise and understanding of healthy self-care practices. Our results show that counsellors recognised several benefits, such as using LLMs as stepping stones or springboards towards improved self-care. They also highlighted several areas of concern, such as unintended consequences that might negatively affect users. We discuss the dissonance around how the early adopters of LLMs appropriate this technology to care for themselves, how counsellors see such usage, and outline implications of using LLMs as a technology for self-care.

Joel Wester, Sander de Jong, Henning Pohl, Niels van Berkel. *Using LLMs for Self-Care: User and Counsellor Perspectives*. International Journal of Human-Computer Studies.
https://doi.org/10.1016/j.ijhcs.2025.103589

Paper 5

"As an AI language model, I cannot": Investigating LLM Denials of User Requests

Abstract

Users ask large language models (LLMs) to help with their homework, for lifestyle advice, or for support in making challenging decisions. Yet LLMs are often unable to fulfil these requests, either as a result of their technical inabilities or policies restricting their responses. To investigate the effect of LLMs denying user requests, we evaluate participants' perceptions of different denial styles. We compare specific denial styles (baseline, factual, diverting, and opinionated) across two studies, respectively focusing on LLM's technical limitations and their social policy restrictions. Our results indicate significant differences in users' perceptions of the denials between the denial styles. The baseline denial, which provided participants with brief denials without any motivation, was rated significantly higher on frustration and significantly lower on usefulness, appropriateness, and relevance. In contrast, we found that participants generally appreciated the diverting denial style. We provide design recommendations for LLM denials that better meet peoples' denial expectations.

Joel Wester, Tim Schrills, Henning Pohl, Niels van Berkel. *"As an AI language model, I cannot": Investigating LLM Denials of User Requests*. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI'24) https://doi.org/10.1145/3613904.3642135