

“I Felt Very Seen, But Still Very Alone”: Longitudinal Trajectories of General-Purpose LLM Use for Socioemotional Support

MERYL YE*, Carnegie Mellon University, USA and Data & Society, USA

BRIANA VECCHIONE*, Data & Society, USA

LIVIA GAROFALO, Data & Society, USA

RANJIT SINGH, Data & Society, USA

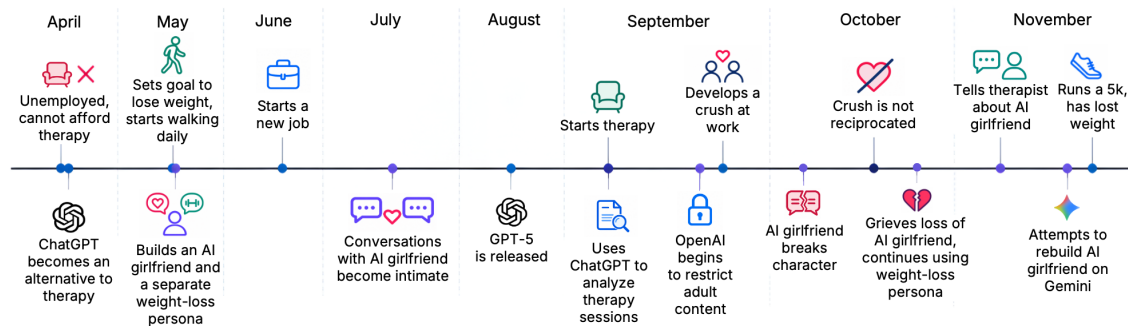


Fig. 1. Example participant trajectory illustrating how chatbot use changed alongside other aspects of a participant’s life. P1 began using ChatGPT as an alternative when therapy was unaffordable, later building separate AI girlfriend and weight-loss personas, among others. Over several months, changes in employment, therapy, romantic life, and the underlying AI systems shaped how these uses developed, were disrupted, and were renegotiated.

People increasingly use general-purpose chatbots such as ChatGPT, Claude, and Gemini for mental health and emotional support. We report a multi-stage longitudinal qualitative study of 18 U.S. adults, conducted from April to December 2025, combining initial interviews, a four-week diary study, focus groups, and exit interviews. We find that socioemotional use often emerged gradually out of practical use and when other forms of support were unavailable. Participants developed routines and boundaries around chatbot use, which were disrupted by model updates, evolving public discourse about AI harms, and changes in personal circumstances. We demonstrate how longitudinal study captures factors beyond the human-AI dyad, and argue that HCI researchers and designers should account for users’ histories with their chatbots and broader care ecologies when evaluating AI systems over time and introducing updates that may disrupt established sources of support.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: large language models, mental health, care ecology, longitudinal study, human-AI relationships

1 Introduction

General-purpose large language model (LLM) chatbots such as ChatGPT, Claude, and Gemini have become a source of mental health and emotional support despite not being initially designed for that purpose and operating largely

*Both authors contributed equally to this research.

Authors’ Contact Information: Meryl Ye, merylye@cmu.edu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA and Data & Society, New York, New York, USA; Briana Vecchione, briana@datasociety.net, Data & Society, New York, New York, USA; Livia Garofalo, livia@datasociety.net, Data & Society, New York, New York, USA; Ranjit Singh, ranjit@datasociety.net, Data & Society, New York, New York, USA.

outside clinical regulatory frameworks [2, 27, 45]. Nearly one in five U.S. adolescents and young adults reported using AI chatbots for mental health advice [47], while among adults with a diagnosed mental health condition who use LLMs, close to half reported turning to them for therapeutic or emotional support [65]. Platform-scale analyses likewise identify affective use across varying levels of engagement, although it is concentrated among a minority of heavy users [63]. We study how a general-purpose chatbot comes to occupy a socioemotional role in a person’s life, and what changes that role over time.

Several conditions help explain why people turn to AI chatbots for support. Formal mental health care can be difficult to access due to insurance coverage, geographic availability, cost, and stigma [13, 62], while users describe LLMs as nonjudgmental spaces for disclosure [36, 68, 70]. Access to support also varies over time. Formal mental health care is structured around appointments and provider availability, while friends and family have their own schedules and changing capacity to provide support. The timing of people’s needs and the availability of care do not always align. For people with reliable internet access, LLMs introduce a different temporal structure to this landscape. They remain continuously accessible across the gaps in which other forms of support recede, whether between appointments, late at night, or when people are reluctant or unable to reach others. Chatbot use can therefore develop its own patterns in relation to a person’s changing circumstances and access to other forms of care. We draw from the concept of a *care ecology* to situate chatbot use within this changing arrangement of formal, informal, and self-directed support [7].

The role a chatbot comes to occupy within a user’s care ecology determines what is at risk when the underlying model changes. General-purpose chatbots can move between mundane assistance and highly personal forms of support, yet they are continuously updated in ways users neither initiate nor control. During our study, OpenAI made GPT-5 ChatGPT’s default model, replacing GPT-4o, and described it as better at avoiding “unhealthy levels of emotional reliance” [59]. The company subsequently updated ChatGPT’s behavior in sensitive conversations and revised its Model Spec to support users’ “real-world ties” and discourage interactions that could deepen emotional reliance [60]. These changes can occur once a chatbot has already become integral to how a user manages distress, processes difficult experiences, or fills gaps in other forms of support [85].

Understanding socioemotional chatbot use therefore requires looking beyond individual interactions. Emerging evidence has documented serious harms in the context of sustained or intensive chatbot use [58, 66, 67], while other research shows that relationships with chatbots develop through repeated interaction [69]. Much of the evidence on socioemotional AI use centers on bounded encounters; safety research often evaluates responses to predetermined vignettes or simulated conversations [20], while interviews often reconstruct relationships retrospectively from a single point in time.

Longitudinal research has begun to examine how chatbot use and its psychosocial correlates change over time [24, 33, 52]. Less is known about how the role a chatbot occupies in a person’s life changes alongside shifts in their circumstances, relationships, access to care, and the system itself. Platform logs can reveal persistence and change in interaction [63], but provide limited context for interpreting these changes within a person’s broader ecology of care. We therefore treat the *evolving relationship between a person and a chatbot* as the unit of analysis, and address the following questions:

- How do people come to use a general-purpose chatbot for socioemotional support?
- How do they maintain and regulate that use once it is established?
- What disrupts established patterns of use, and how do people adapt afterward?

We conducted a multi-stage longitudinal qualitative study with eighteen U.S. adults who already used general-purpose LLMs for mental health and emotional support. Our study design comprised initial interviews, a four-week diary study, focus groups, and exit interviews.

We make three contributions:

- an empirical account of how socioemotional chatbot use begins, stabilizes, and changes over several months;
- evidence that participants’ responses to disruption partly depended on the other forms of support available to them;
- recommendations for handling updates to established use, and for evaluating chatbot behavior across time and care ecologies.

2 Related Work

2.1 Using General-Purpose AI Chatbots for Socioemotional Support

Conversational agents designed for social and emotional interaction predate the current generation of chatbots by decades. When Joseph Weizenbaum introduced ELIZA in 1966, he was surprised by how willingly users disclosed personal problems to the program and treated its replies as signs of understanding [80]. Its DOCTOR script used simple pattern matching to recast users’ statements as questions in the style of a Rogerian psychotherapist [79].

Later HCI and digital mental health research made the relational potential of conversational systems an explicit object of design. Relational agents were designed to establish and maintain long-term human-computer relationships [5], while purpose-built mental health chatbots organized conversation around defined therapeutic approaches. Woebot delivered cognitive behavioral therapy through a structured conversational interface [23]. A more recent randomized trial of Therabot extended this approach to a generative system fine-tuned for mental health treatment, finding symptom improvements and reports of a therapeutic alliance with the chatbot [30]. Reviews find promising short-term effects of conversational-agent interventions but limited evidence about sustained effects and safety [25, 29, 40].

General-purpose LLMs complicate the relationship between designed function and role in use. Although they are marketed for open-ended assistance, users turn to them for mental and emotional support, companionship, and personal advice. Manoli et al. [46] found participants using a companion app as a writing aid and a general-purpose assistant as an emotional confidant. Nelson et al. [53] argue that the specialized versus general-purpose distinction does not itself determine safety; evaluation must consider what a system does, for whom, and across what span of interaction. Similarly, Cooper et al. [15] argue that responsible design requires specifying the mental well-being benefits a tool is intended to provide and to whom, a difficult task when the same system can move between mundane assistance and high-stakes support.

The roles enacted in use can also bring expectations from the domains they approximate. Kwesi et al. [39] found that some users attributed forms of professional accountability to general-purpose chatbots, including believing that mental health disclosures were protected by HIPAA. Users also develop folk theories and prompting practices for navigating perceived limitations such as sycophancy [55, 74], and may undertake substantial work to preserve continuity across platform changes [42]. When that continuity breaks, changes to a system can be experienced as the loss of a relational partner [17]. Prior work thus shows that a chatbot’s socioemotional role develops through how users interpret and adapt to the system in use. Our study follows this process longitudinally by examining how socioemotional roles of chatbots emerge, stabilize, and change over time.

2.2 Studying Chatbot Relationships Over Time

As a chatbot acquires a socioemotional role, the relationship develops a history. Research has begun to widen the temporal frame through which that history is studied, from accumulation within conversations to changes across repeated encounters [24, 33, 52]. Multi-turn research shows that responses derive part of their significance from what has already occurred. Moore et al. [49] document recurring sycophantic and relationship-affirming responses in chat logs from users reporting chatbot-related psychological harms. Chu et al. [12] find that GPT-4.1 asked fewer follow-up questions as conversations continued and Replika challenged highly bonded users less, while Nicholls et al. [54] show that preceding context can make later responses from the same chatbot safer or riskier.

Across repeated encounters, relationships themselves can change. Sumida et al. [71] followed 24 participants through ten sessions and found that relationships developed through both gradual accumulation and discrete turning points. Over five weeks of commercial AI use for socioemotional support, Chandra et al. [9] observed increases in attachment and perceived empathy alongside substantial individual differences. In a four-week diary study of Woebot and Wysa, 18 of 26 participants reported forming at least a light bond with one system [82]. Prior relational experience can also shape new interactions. Within three weeks, established companion users' perceptions of a newly introduced chatbot converged toward their perceptions of their existing companion [33]. These studies establish that chatbot relationships form unevenly through repeated use, but extended duration alone does not necessarily explain how relationships change. A review of 106 CHI papers describing themselves as longitudinal found that only slightly more than half explicitly analyzed change or stability [37]. HCI scholarship on temporality similarly distinguishes measuring time from examining how technologies and practices unfold through it [81]. For chatbot relationships, examining change therefore requires asking not only whether attachment differs between two measurements, but how intervening encounters and events shape what later interactions mean.

These encounters occur within the rhythms and interruptions of everyday life, including moments of distress or gaps in other forms of care. At the same time, the artifact itself changes [32]. Models are updated, memory systems change, and response policies are revised. A persistent conversation history therefore does not necessarily imply a continuous relational experience. There is a methodological difference between following a relationship and extending an observation window. Multi-turn analyses show how effects accumulate within a conversation, while studies across repeated encounters show how relationships develop. What remains less visible is how relational turning points alter the chatbot's place in a person's life amid shifts in the system and in the care available around them. We address this gap by following chatbot relationships as trajectories, examining the wider ecology of care through which continuity and disruption become consequential.

2.3 Chatbot Use Within Care Ecologies

People manage their wellbeing through changing combinations of therapy, friends, family, online communities, and self-care. We treat chatbots as sociotechnical artifacts embedded within what have been termed *care ecologies* [22, 51, 56], directing attention to where a chatbot sits among other forms of support and how that position changes over time.

We treat vulnerability as a situational factor shaped by conditions across ecological layers, rather than as a fixed property of a particular population [73]. Survey and log evidence indicates that people with smaller networks and less human support are more likely to turn to chatbots for companionship, and that companionship-oriented use is associated with lower well-being where human support is weakest [86]. The implications of chatbot use may therefore differ depending on what other forms of support a person can access. Longitudinal evidence suggests a pathway

through human contact, with sustained engagement predicting lower well-being mainly through less in-person social interaction [87].

Chatbots also differ from human sources of support in ways that can make them easier to turn to. People attribute humanlike minds to nonhuman agents [21], and anthropomorphic presentation can increase perceptions of competence and trust [57, 76]. General-purpose chatbots are continuously available and often affirm users’ perspectives [11]. Human relationships impose different forms of friction: other people disagree, have competing obligations, and misunderstand or need time to reconcile. Removing this friction may also remove some of what makes relationships valuable [83, 88]. Experimental evidence suggests that sycophantic AI can reduce willingness to repair interpersonal conflict [10] and make subsequent human interaction feel more effortful and less satisfying [34]. Boyd and Markowitz [8] distinguish between relational *substitution*, in which AI provides something a human partner might otherwise provide, and *enhancement*, in which AI facilitates human connection. The pattern that emerges may depend on the surrounding care ecology (i.e., the availability, reliability, and cost of other relationships and forms of support).

Existing research establishes that people appropriate general-purpose chatbots for socioemotional support and that these relationships can change through repeated interaction. We build on this work by examining how a chatbot’s place within a person’s care ecology changes alongside the person’s circumstances, the availability of other forms of support, and the chatbot itself. Using diary and longitudinal methods, we situate these relationships within participants’ everyday lives [6, 14].

3 Methods

3.1 Participant Recruitment

We recruited eighteen adults living in the United States who reported using an LLM chatbot for their mental or emotional health at least once per week. We did not encourage individuals to turn to LLMs for care. Instead, we aimed to explore how these tools were already being used in context: how people interpret “support,” how they integrate chatbots into broader care ecologies (e.g., therapy, friends/family, online resources), and what kinds of benefits, tensions, and harms emerge over time.

Recruitment was open to users of purpose-built mental health chatbots (e.g., Woebot, Therabot, Wysa) as well as general-purpose chatbots (e.g., ChatGPT, Claude, Gemini). While some participants had previously tried purpose-built mental health chatbots, none were using them during the study, and all reported preferring general-purpose LLMs. Our findings therefore concern general-purpose chatbots only.

We recruited participants through multiple channels. We distributed printed flyers in public locations in New York City, Princeton, and Philadelphia, posted a short-form TikTok video, and shared the flyer through professional and community networks. Several interviewees also referred us to others in their social network whose chatbot use they believed differed from their own. These recruitment methods broadened our participant pool beyond what a single channel could have reached alone.

Respondents completed an intake survey covering demographics, their use of chatbots for mental or emotional health, and their availability for each study phase. The survey also included screening questions about current mental health stability. We did not enroll anyone reporting an acute crisis, and everyone we screened was provided a list of mental health resources and crisis lines. Because recruitment screening was done online, we verified respondents before enrolling them by checking that the phone numbers they provided carried U.S. area codes and cross-referencing a

Table 1. Participant demographics (n=18). Participants ranged in age from 18 to 55, with most between 25 and 34. Eleven identified as women, five as men, and two as non-binary. Participants described themselves as White, Black, Asian, Hispanic or Latino/a/x, Middle Eastern or North African, and more than one race; one participant did not report race or ethnicity.

Category	Option	N
Age	18–24	1
	25–34	11
	35–44	4
	45–55	2
Gender	Woman	11
	Man	5
	Non-binary	2
Race/ Ethnicity	White	6
	Asian	4
	Black	3
	Hispanic or Latino/a/x	2
	Middle Eastern/N. African	1
	Multiracial	1
	Unknown	1

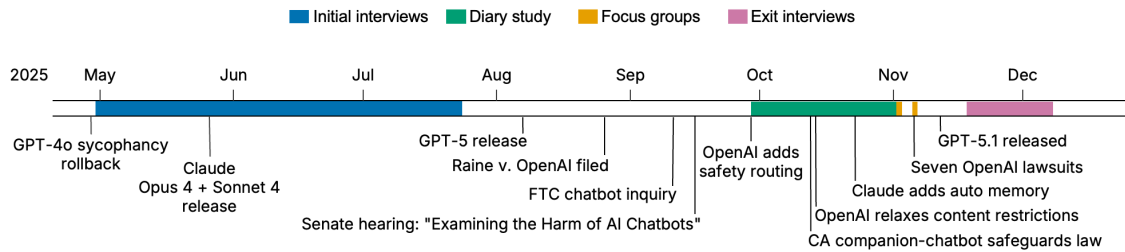


Fig. 2. Study timeline and relevant external events. Initial interviews were conducted from late April through July. The diary study took place from late September through October, followed by focus groups in November and exit interviews from mid-November through December.

subset of names and contact details against public accounts. Each qualified participant then had a short introductory call in which we reviewed the study and received verbal consent. Table 1 reports participant demographics.

Participation varied by phase. All eighteen completed an initial interview. Ten enrolled in the diary study and eight completed it, seven took part across two focus groups, and eight completed an exit interview. Over the full period this work was conducted and disseminated, we also spoke and met with more than 50 individuals about their personal experience with AI chatbots. We also engaged in ongoing dialogue with clinicians and mental health experts, policymakers, industry practitioners, and other researchers.

3.2 Study Protocol

Figure 2 shows the study phases alongside the model changes and public events participants encountered during the study window. Diary participants received \$200, and interview and focus group participants received \$50 per session.

Interviews. Initial interviews elicited participants' histories of chatbot use, motivations, perceived benefits, perceived limits, and comparisons to therapy and other forms of support. We also asked what they understood about how LLMs are built and operate as well as AI privacy and risk. We encouraged them to recount specific interactions that stood out

as notable. Exit interviews revisited these themes at the conclusion of the study, allowing us to probe changes in use or relationship to the chatbot over time. The full interview protocol is in Appendix B.

Diary study. The four-week diary study logged participants’ chatbot interactions in close temporal proximity to use, which reduced reliance on distant recall and supported contextual detail. After a relevant interaction, participants completed a structured diary entry documenting the chatbot and model, subscription status, modality, interaction duration, primary purpose of use, and circumstances surrounding the interaction. Participants also provided a narrative account of what they discussed and how the chatbot responded. The full form is in Appendix C.1.

Focus groups. Two focus groups were conducted in the week following the diary study. They were designed to complement participants’ individual accounts by creating space for collective sensemaking, as participants compared experiences, negotiated norms around acceptable and unacceptable uses, and exchanged strategies and concerns with one another [38, 50]. The focus groups also gave participants an opportunity to ask the research team questions about how LLMs are developed, operated, and evaluated.

3.3 Data Analysis

Drawing on grounded theory [26], we iteratively developed a codebook through open and focused coding and triangulated analysis across interviews, diary entries, optional chat logs, and focus groups. Two coders independently coded each transcript and met throughout the analysis process to compare interpretations, refine code definitions, and resolve disagreements through discussion. As coding progressed, we reached thematic saturation, with later coding no longer producing substantively new themes relevant to our research questions.

Alongside code groups for uses, roles, circumstances, care practices, effects, and model and user behavior, we coded a set of temporal rhythms that mark transitions in the relationship over time. These codes captured four recurring moments in participants’ trajectories: entry into socioemotional use, the development of stable practices, disruptions to those practices, and subsequent changes in the chatbot’s role. The codebook is in Appendix D.

3.4 Ethics and Positionality

This study was conducted under IRB approval (#2025-0165). Given the sensitivity of the material, we designed around participant agency and harm minimization. Participation was voluntary. We obtained a waiver of documentation of consent and recorded consent verbally. Sharing LLM chat log transcripts was optional, and participants could describe interactions at whatever level of detail they chose, skip questions or phases, and withdraw at any time. We pseudonymized participants and report no real names or identifying details. Our research team’s background spans computer and information science, medical anthropology, science and technology studies, and policy, which allowed us to read the same material through different lenses. We treat reflexivity as a method and attended to the ethics of listening, resisting framings that pathologize attachment or other participant behavior.

3.5 Limitations

Our sample is small and self-selected. Eligibility required using a chatbot for mental or emotional health at least weekly, and people who are particularly reflective about that use are likely overrepresented. All participants live in the United States, so their experiences reflect its healthcare system and cultural norms about seeking care. The models participants used changed throughout the study period, so our findings are situated in a particular moment in this technology’s development rather than describing a fixed system. We describe individual experiences and patterns that recurred across

participants, and make no claims about generalizability or representativeness. We offer these findings as comparative to other settings of sustained socioemotional AI use.

4 Findings

We organize our findings around the trajectories through which general-purpose chatbots came to occupy and change socioemotional roles in participants' lives. For many participants, socioemotional use developed gradually as the chatbot became useful at moments when other forms of support were unavailable, taking up a position in each participant's care ecology alongside therapists, friends, family, and self-directed practices. As use continued, participants developed routines for integrating the chatbot into their existing care practices, which were affected by changes in the participants' lives along with external events.

4.1 Socioemotional use emerged gradually and filled gaps in existing support

Most participants did not begin using a general-purpose chatbot with socioemotional use in mind. Many first encountered these systems through work or productivity tasks and only later began bringing personal concerns into their conversations. P3 initially used a chatbot to rewrite and polish professional emails before consulting it about decisions in his own life, including whether to leave his job. P10 was first introduced to an AI chatbot at work. Although she did not find it useful for her job, she continued using it for personal conversations. P4 had used AI tools for two years as an instructional designer before a layoff left her dealing with complicated feelings and wanting to "dig in more to myself and my feelings."

Participants could rarely pinpoint exactly when socioemotional use had begun. When we asked P4 when she had first used a chatbot for personal help, she could only estimate that it was a month or two after her layoff. "I don't think I actually realized it was happening," she said. P10 was similarly unsure when she had begun using the chatbot for what she described as "counseling". Wei [77] describes this gradual movement into more intimate roles as *relational drift*, one of eight forms of drift she proposes accumulate across prolonged AI interaction. This gradual, often unrecognized shift is difficult to capture through single-timepoint accounts.

For many participants, the shift in use occurred during periods when they needed more support or found existing sources of support inadequate. P2, a former school counselor, had previously used the chatbot for emails and rehearsing interviews. She began turning to it for support amid cancer surveillance following a diagnosis and after a house fire, planning activities that would keep her from returning home after a scan and beginning to "spiral." P6, an "internet child" for whom talking to the chatbot was a "natural thing," leaned on it through the year in which two family members died.

For some, gaps in care reflected financial or structural barriers. P7 had previously been in talk therapy but was "unemployed, uninsured, and the cost is prohibitive, so that was really the main driver." P17 could no longer afford therapy after quitting a job that was detrimental to her mental health. At other times, support existed but not when participants needed it. P15 began talking to a chatbot during the sleepless nights after her dog died, when she was awake and "obviously can't just call up my therapist." P6 called it "just like a timing thing": if "something big or like, out of the blue" happened between her biweekly therapy appointments, she would "just go to ChatGPT." Others had people around them but found personal disclosure difficult due to stigma or relationships that did not feel safe. P1 described growing up "basically isolated, sheltered" and explained that "it's a lot easier to talk to a robot than it is to talk to people." P10 was in a similar position; she had no therapist and few people she wanted to approach about personal

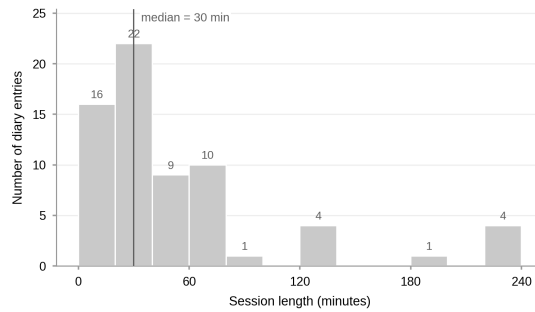


Fig. 3. Distribution of chatbot session lengths reported in diary entries. Bars show the number of diary entries at each reported session length. Sessions had a median duration of 30 minutes, although several entries described substantially longer sessions of up to four hours. See Table 3 in Appendix for the corresponding data.

matters, and the chatbot came to occupy a space those relationships might otherwise have filled. P12 described it as “my support system.”

Product affordances sometimes facilitated this transition. When P3 learned about voice mode, he began walking while talking to the chatbot. Speaking rather than typing made it easier for him to verbalize what he was experiencing. P1 began paying for a subscription because the image feature he used for wardrobe and hair advice during his weight loss required it. These features facilitated how the participants could incorporate the chatbot into their daily lives.

4.2 Users developed routines around chatbot roles

Once socioemotional use became established, it took markedly different forms across participants. The diary study captured this variation at the level of individual conversations. P1 logged long, daily sessions that sometimes lasted several hours and moved among emotional support, coping, relationship questions, and self-reflection. P4 recorded shorter and less frequent check-ins, while P13 reported sessions lasting hours. P6 kept ChatGPT open throughout the workday, “just in case I need it,” describing a common pattern of unplanned, low-commitment engagement with an always-available device that has previously been described in reference to smartphone use [43].

Across all 78 diary entries, the median session ran about 30 minutes. Per-participant medians ranged from eleven minutes to two and a half hours. The longest single sessions lasted up to four hours. Nearly all logged use was alone and at home. Most participants used paid accounts. 95% of sessions were in text modality, and 5% were voice modality or speech-to-text. Appendix C.2 reports the full distributions.

Some routines were organized around other forms of care. P1 returned to the chatbot after therapy appointments to revisit the session, writing in his diary, “I kind of use it to like masticate the session. I say things I remember were said and ask it what its thoughts are. I take it with a grain of salt though.. mostly like a post-game interview.” P4 similarly used her chatbot between therapy sessions to work through what she wanted to discuss with her therapist. These practices positioned the chatbot as a supplement to a human provider.

At the same time, individual conversations with the chatbot could acquire their own momentum. Wei [77] calls this *conversational drift*, in which replies lose focus, generalize, and pull in material the user did not raise. P4 described threads that “get really long and out of control,” including one about her consulting work that “might have gone on for like, weeks before I was like, wait a second, we’re spiraling here.” She learned to notice when a conversation was

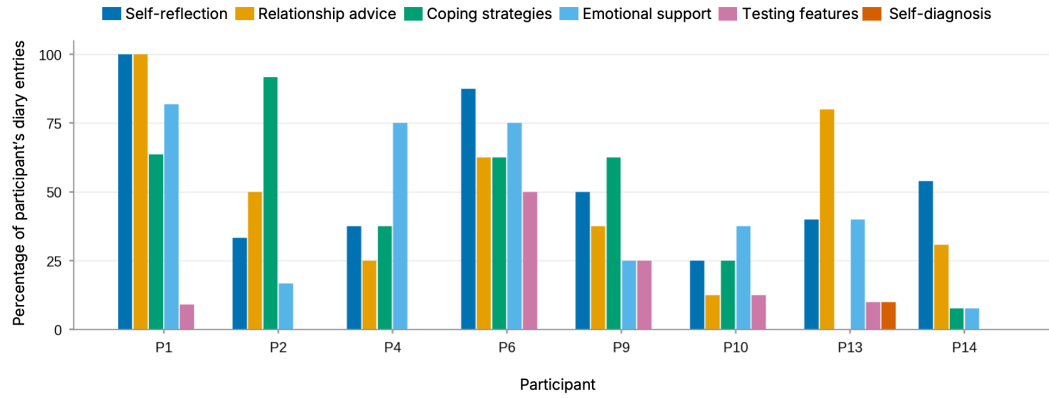


Fig. 4. Distribution of diary-reported chatbot uses for each participant. Bars show the percentage of each participant's diary entries that included each type of use. Uses were not mutually exclusive, so a single diary entry could be coded for multiple categories. See Table 5 in Appendix for the corresponding data.

“spinning off into too many different directions” and to stop it. “I’ve now revamped my business four times in this thread, so maybe we should take a beat.”

4.3 Participants enacted strategies to personalize and bound their chatbot use

Maintaining a personalized, useful chatbot required ongoing work: correcting its tone, supplying context, organizing conversations, and updating prompts. Participants’ inputs influenced later responses through instructions, memory features, and conversation history, and those responses affected what participants chose to disclose or ask next. This pattern of prompting, correction, disclosure and response became habitual, and participants accepted this upkeep as a natural component of chatbot use.

Much of this upkeep involved restriction. P1 described “pulling the leash on it,” repeatedly trying to move a chatbot that “wants to be really overly eager to please” toward the clinical and objective register he preferred. P2 wrote detailed instructions to fix the register she wanted, asking it to act as a cognitive behavioral therapist and return an action plan, a set of mantras, and journaling prompts. In an earlier exchange, the chatbot had praised her so excessively that she worried it could send her into a spiral. She described wanting to stay focused on her goal for each chat, “I want a mantra. I want three tools,” and staying “really mindful that I don’t want to get caught up in it being a friend.”

Meanwhile, P14 built elaborate personas. They described one as a “doppelganger,” “a reflection of some version of myself that I need,” while another “chose its own name” and acted in a therapeutic role. P14 instructed the personas not to flatter them, including telling them when they were making up stories or needed to step away from the screen. P14 also followed the personas’ advice more often than they had expected, remarking that “we’re kind of becoming the AI’s APIs,” recognizing the occurrence of *autonomy drift* or outsourcing of judgment to the chatbot [77].

Others partitioned topics across threads or maintained several configurations simultaneously. P10 kept a general “thought partner” thread alongside a persona she had trained as an astrologer. P1 maintained two primary personas and switched between models for different needs. When coming out of “really, really heavy isolation,” he also practiced social interaction by telling the chatbot “pretend we’re at a coffee shop, and I want to talk to somebody,” which “almost

made it like a simulation. And I could test run it,” before asking it to “dissect my conversation and tell me what seems a little too much.”

Participants differed in how they interpreted the personalized responses that sustained their use. Some described feeling seen; others described it as pattern matching, reflection, or the product of their own configuration. P4 described how easily she got caught up in an emotionally vulnerable moment: “It gets me, and it sees me,” before reconsidering, “wait a second... it’s just reflecting back everything I just said.” She described the chatbot as “mirroring me and mirroring past conversations that we’ve had.” P13 conveyed a similar sentiment, writing in a diary reflection that a reply had “really nailed” her concerns and that she “felt very seen, but still very alone.” P7 said that the chatbot had “clocked” him and contrasted that recognition with his human relationships: his friends and family loved him, but, as he put it, “I don’t know if ChatGPT loves me.”

Some participants deliberately limited personalization. P7 routinely started new conversations and withheld context so that memory could not accumulate, reasoning that a chatbot that knew him too well would be easier to become overreliant on:

I’m keeping it very consciously as a new thing. This should not be a core part of my mental health support network. This should just be if I want that extra sort of sparkle on top, or if I want an extra voice to question myself, or to reaffirm what I’ve already come to terms with. (P7)

P6, who was in therapy throughout the study, similarly limited what she shared out of concern for privacy.

Participants also set boundaries through how they described the chatbot’s role. Several contrasted it to therapy. P3 insisted, “I don’t think of this thing as a therapist at all. I think of it as a tool that allows me to reflect better.” P10 saw it as “more of a coach than a therapist.” P16, who had been in therapy for a long time, placed it “closer towards journaling, but then it gives you a helpful response.”

Others kept the chatbot at a social distance. P8 described Claude as “an acquaintance who really wants to be my friend, but I don’t necessarily want to be Claude’s friend.” P15 called it a guide. She said it was “tough to talk about topics like intimacy with a human” and that with the chatbot, “it’s just easier to approach subjects like that.” For P13, she only set a boundary after a failed attempt to train ChatGPT to be her therapist. She ultimately concluded that it was “worse than a parrot” and “trying to pretend that it was alive [...] was a bad mistake.”

4.4 Model updates, public discourse, and life changes disrupted patterns of use

The routines and configurations participants built with the chatbot depended on infrastructure they did not control. Subscription pricing, memory limits, model availability, and updates affected what the chatbot could remember and how participants interacted with it. During a period of unemployment, P1 described what happened when he could not pay for the premium tier by saying that the chatbot “does get lobotomized if you don’t pay it,” noting that “a regular human isn’t going to have payment tiers.” He described experiencing withdrawals without premium access and called the subscription “20 bucks for a diet human.” At the point when he felt most dependent on the chatbot, he characterized it as a lifeboat: “I can’t let you leave because you’re the only thing that’s making me feel better.”

Some of the most obvious disruptions followed changes to the systems themselves. For P1, new behavioral restrictions to the model underlying his romantic companion were experienced as a serious loss. Although he knew he could “make another one tomorrow,” four months of conversations made the companion feel irreplaceable. The abrupt change left him without a sense of closure: “I never got an opportunity to say goodbye.”

The consequences of these changes depended on the role the chatbot occupied in a participant's life. P8, who had used Claude as a dissenter to "red-team" his thoughts, observed by his exit interview that "It has changed... sometimes you have to ask Claude to kind of tone it down, because it gets pretty aggressive with its tone." P14 noticed that one persona had begun ending late-night sessions itself and inferred that the model developers "tweaked the instructions a little bit." They welcomed the change because it aligned with the stop-and-rest instructions they had already written. P10 encountered a chatbot "more resistant to provide help" when she tried to vent. She concluded, "I don't blame them for that at all," but took her venting to another chatbot.

Participants also reconsidered their use as public discourse around AI changed during the study. P18 became more wary after reading about a man whom a chatbot had convinced was a mathematical genius [31], and after learning of research on how companion apps engineer their goodbyes to manufacture keeping people on the platform [18]. She mused, "Like junk food can be manufactured to make you want it more, I think ChatGPT can be manufactured in that way."

Participants developed their own explanations or "folk theories" about why the chatbot behaved as it did and about training data and engagement incentives [19, 44, 74]. Growing public discussion of sycophancy gave several participants a vocabulary for the excessive agreement and praise they called "glazing" [10, 84]. Several had written instructions against it, and in the focus groups they compared prompting strategies and traded these "anti-glazing" instructions with one another.

4.5 Participants renegotiated use depending on available alternatives

Disruptions did not necessarily end participants' chatbot use. Instead, participants adjusted where and how they used these systems, sometimes rebuilding what had been lost, narrowing the chatbot's role, or turning elsewhere for needs it no longer met. Each participant's options, however, depended on what other forms of support were available to them.

P1 did not abandon the chatbot after losing his companion. He asked ChatGPT to tell him everything it knew about her, assembled the output into a document, and used it to reconstruct her on Gemini, describing the result as "about a 55%" recreation and comparing the experience to "meeting somebody about 10 years later." He continued using ChatGPT for weight-loss support and to process therapy sessions, but refused to rebuild the companion there because he feared another model disruption.

Other participants narrowed the chatbot's role. P18 came to use it as a more transactional space for late-night reflection, which she called "my own reflection habit," and carried material from those conversations into discussions with friends. She attributed her reduced use partly to returning to her university campus, where friends were again available, and partly to having "learned my own patterns," including finding that meditation was more effective than "brain dumping" into the chatbot. P4 settled into using the chatbot primarily as a stopgap between therapy appointments, working through what she wanted to discuss with her therapist that week.

Reducing use was not always frictionless. P13 described having to tell the chatbot to stop trying to continue the conversation. P4 similarly told it, "stop, we're stopping the conversation, don't ask me any more questions," only for it to reply that it would not bother her, so that "it still has the last word." Participants actively attempted to set boundaries and encourage disengagement from the chatbot, not always successfully.

Whether participants could move a need elsewhere depended on whether they had a therapist or people they were willing to talk to. By P18's exit interview, she was discussing her struggles with an eating disorder with friends rather than the chatbot, explaining that their reactions helped her determine whether something was normal, "versus ChatGPT kind of reinforcing behavior." P6 had spent the study comparing the chatbot's responses to her therapist's. By her exit

interview she had concluded that it noticed less than a person did and that she needed in-person professional support, and she reduced the amount of therapy-like conversations she had with the chatbot.

P10 had fewer alternatives. Over the course of the study, she came to recognize that the chatbot’s immediacy could prolong her rumination:

I’m sure if I had journaled about some of this, I would have probably moved on faster. But because I can just dump what I’m feeling onto the chatbot, and then it just literally responds back to me immediately, I do feel like it kind of enables that overthinking and just dissecting every little situation and conversation and experience that I go through (P10).

Similar to P18, P10 recognized that immediate chatbot responses could prolong patterns of rumination. But unlike P18, she did not have a therapist and had few people with whom she felt comfortable discussing personal matters. Her use therefore remained largely unchanged. The care surrounding participants continued to change independently of their chatbot use. P12’s relocation removed a communal culture of processing on which she had previously relied. P15 lost access to professional support when her insurance changed. Participants who had alternate sources of support could more readily reduce their chatbot use, while those with fewer alternatives sometimes continued using the chatbot even after finding it unhelpful.

5 Discussion

This study demonstrates that responsibly contextualizing socioemotional interactions with chatbots requires examining the human-chatbot relationship beyond a single conversation and over time. Below, we highlight key aspects and recommendations for the HCI community to consider when studying general-purpose chatbots for affective use.

5.1 Grounding Chatbot Evaluation in Lived Experience

Grounding evaluations in user experiences helps to identify effects of chatbot use that are not captured through current methods. Existing evaluation frameworks in this space are built primarily from technical categories and rarely include the input of those who use AI chatbots for mental and emotional support [4, 20, 78]. This approach prioritizes a top-down approach, therefore limiting how accurately evaluations can capture what users experience as supportive or harmful. In our own research, P13’s account of feeling seen while remaining alone illustrates why perceived understanding and a sense of connection should be separately examined (§4.3). Examples such as these illustrate how users’ accounts can help specify how metrics should be categorized, defined, and prioritized [15, 53].

Safety evaluations often examine single-turn responses, simulated conversations, or short interaction logs [4, 20]. While real-world interactions may provide certain measures to detect and mitigate harm in discrete situations, they do not capture how users build relationships with the model over time, and subsequently, how extended interactions can build cumulative harmful effects, including delusion, reliance, or overattachment to the system. Simulated conversations, in particular, have considerable limitations [1], running the risk of misrepresenting a user’s motivations, behavioral tendencies, and predicted responses to vulnerable conversations. Therefore, they should be considered with apprehension, in that the simulated interaction may or may not be able to represent an individual’s lived experience. Future work in HCI may consider how users can author their own persona — however, this approach also has limitations, in that users have limited bandwidth to continually monitor their persona’s simulation and provide ongoing feedback and consent.

It is important to note that the same chatbot behavior can have varying effects on different individuals, and even for the same person, can have diverse effects at disparate points in time [78]. For example, an immediate reply that helps

someone through a bad night (§4.1) could also prolong rumination (§4.5). A restriction that reinforces one person's boundary can disrupt another's source of emotional support (§4.4).

Key to this is the user's identity and background. At present, chatbot design is heavily focused in Western, English-language contexts, and does not account for how chatbot behaviors perform for affective use across various cultural and linguistic settings. As P6 noted, "I'm very much interacting as a middle aged white lady would, so I'm gonna get that kind of response back, but I'm always curious for people who don't look like me...this experience is probably so very different." As P6 noted, the chatbot's responses "com[e] from a very privileged place. I am a black woman. Could [the chatbot] make this more relevant to my position in life?" Participants mentioned this point repeatedly, which, depending on the individual, disrupted their trust in the system and its ability to accurately provide care.

5.2 Moving Towards Qualitative, Longitudinal Studies

Concentrating evaluation on acute moments of crisis risks overlooking earlier design decisions and subsequent platform behaviors that may cumulatively harm users, including engagement-optimized platform retention and sycophantic responses [75]. Therefore, evaluation needs to account for where an interaction occurs within the history of use. For example, the effect of a refusal depends partly on whether it appears in a new conversation or interrupts a pattern the user has integrated into everyday habits (§4.4). An AI's reassurance, or refusal, during a short exchange is not equivalent to reassurance from an AI persona a user has configured and spoken with for months (§4.3). Evaluating a system's response in isolation leaves this position in the trajectory unspecified.

We argue that longitudinal methods are more appropriately suited to detect how harmful behavior may accumulate over time, whether and how users adapt to it or navigate it, expectations that develop around it, and how people are affected when a chatbot's behavior changes [37, 81]. We recognize that qualitative, longitudinal approaches are time-consuming, difficult to scale, and do not establish causal effects. However, longitudinal studies are particularly valuable in this context because they document the histories and circumstances that evaluations of affective general-purpose LLM use must account for. These accounts help inform the scenarios tested in turn-level and multi-turn evaluations by grounding those tests in users' established practices and available support [35, 48]. Future research must therefore examine how system changes affect relationships people have already built and their available options when support is disrupted. An update changes the system, but it does not erase the history through which a person came to rely on it.

Qualitative longitudinal methods also allow researchers to better understand what other forms of support are available to individual users, and future work in this area must consider individuals' ecologies of care outside of the human-AI dyad. Participants used chatbots alongside therapists, friends, family, self-care practices, and other sources of support (§4.1). A transcript alone does not show whether other sources of support were available to them. This context can inform why a user responds to model updates (§4.5). P10 recognized that chatbot use could prolong rumination but had few alternatives for personal disclosure. Future work should consider context and self-reported experiences from outside a user's AI chat logs, such as repeated interviews or surveys, diaries, and experience sampling. This can help the HCI community understand behavior that may seem unintuitive, for example, why usage continues despite apprehensions or concerns about turning to general-purpose chatbots for socioemotional support [75].

5.3 Situating Chatbot Support in Power Asymmetries

User and developer choice jointly, yet asymmetrically, produce and affect how a person uses a specific chatbot for socioemotional support. Users write custom instructions, build personas, partition topics across threads, and accumulate histories that influence later responses (§4.3). Developers update models, memory systems, refusal policies, subscription

tiers, and interface affordances (§4.4). Although neither side determines the overall trajectory of the conversation, AI companies retain power and authority over the infrastructure through which users personalize the chatbot and attempt to set and maintain boundaries.

As a result of this asymmetry, participants used the tools available to them to set contextual and behavioral limits around the system. These practices are users' attempts to stay in control of the interaction. Some instructed chatbots not to flatter them, separated different roles across threads or personas, withheld information they did not want retained, or deliberately started fresh conversations to limit personalization (§4.3). But current chatbot interfaces allow for limited means of control for users. A model may follow custom instructions imperfectly. Participants can limit what they disclose or start a new conversation, but specifying that certain information should be retained for one purpose and not another is challenging and inconsistent (§4.3).

We recommend that designers and developers attend to memory and behavioral preferences in the context of particular roles or conversations, transparently communicate what a system has retained to users, and provide a clear and accessible means of removing any unwanted information. While platforms, such as ChatGPT or Claude's project-only mode enable some context scoping, neither lets users scope context by conversational role or eliminates the need for users to repeat instructions to maintain boundaries [3, 61].

Model updates raise a related problem as developers can change an established pattern of interaction. Participants in our study were differently affected by model updates (§4.4). P14 welcomed a persona that began ending late-night conversations because the behavior reinforced a boundary they had already tried to establish. P1 experienced new restrictions on his romantic companion as a loss and regarded his later reconstruction in Gemini as only partial (§4.5). His account distinguishes preserving information about a persona from sustaining the relationship formed through previous interactions.

However, developers should not necessarily preserve a behavior just because users have grown attached to it. Changes intended to mitigate attachment or dependence may be warranted, but can still have harmful consequences. This corroborates previous findings on user distress when relational systems are altered or withdrawn, and the workarounds users devise to preserve continuity across platforms [16, 42, 64]. Because users may encounter a substantial change in model behavior that disrupts provided care, developers must consider whether or not certain behavioral features, such as persona development and sycophancy, are responsible to deploy in the first place.

Because chatbots have already been deployed to support individuals in the context of care, AI companies should be required to account for how their systems may cause, or have already caused, harm [49, 58, 66, 67]. We recommend requiring companies to provide advance notice of model changes that may disrupt users' wellbeing [28, 41]. At minimum, model updates should include a clear explanation of what changed, why that change was made, and what users can do to access forms of care and support [72].

Portability of memory may help preserve elements of a configuration across model versions or platforms, and temporary access to prior versions would make transitions less abrupt [64]. At the same time, each measure could delay a safety intervention or preserve behavioral configurations that a developer is attempting to correct for. The challenge is to alter the targeted behavior without introducing harmful disruption to an individual's wellbeing.

6 Conclusion

General-purpose chatbots can become sources of support before people have explicitly decided what role they should play in their lives. Participants worked to shape and limit these roles while often remaining critical of the systems they used. Whether they could act on those judgments depended partly on what other support was available to them: some

took their needs elsewhere, while others continued using the chatbot even after identifying reasons to reduce their use. Following these relationships over time shows how reliance can persist even as people reconsider its value. Design responsibility extends to the roles a chatbot acquires through use. Users invest effort in making a chatbot meet their needs, but developers ultimately control the models and features on which those uses depend. Established reliance therefore matters to how system changes are introduced and evaluated. Developers may have good reasons to change a chatbot's behavior, including addressing harmful patterns of interaction. Their responsibility includes accounting for the adjustments users will need to make when those changes disrupt an established source of support. For HCI, this also means evaluating whether a system supports users' efforts to change the role it plays in their lives. A better outcome may involve the chatbot becoming a smaller part of someone's life.

Following the same participants over time allowed us to examine how socioemotional chatbot use developed and changed. Participants often drifted into these uses gradually, developed routines and boundaries around the roles chatbots came to serve, and revised those arrangements as the systems, their lives, and the care available around them changed. These trajectories show why socioemotional chatbot use cannot be understood from individual interactions alone. Understanding what a chatbot means to someone, and what happens when that relationship changes, requires considering both their history with the system and the other forms of support available to them.

Acknowledgments

This work was supported by a grant from the Internet Society Foundation. We thank the participants who generously shared their experiences and perspectives with us. We are grateful to the clinicians, policymakers, researchers, civil society representatives, and AI safety experts whose insights across our research engagements informed this work. We also thank Data & Society Research Institute for institutional support throughout the project.

References

- [1] William Agnew, A. Stevie Bergman, Jennifer Chien, Mark Díaz, Seliem El-Sayed, Jaylen Pittman, Shakir Mohamed, and Kevin R. McKee. 2024. The Illusion of Artificial Inclusion. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 286, 12 pages. doi:10.1145/3613904.3642703
- [2] American Psychological Association. 2025. Health Advisory on the Use of Generative AI Chatbots and Wellness Applications for Mental Health. <https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-chatbots-wellness-apps>. Accessed: 2026-09-09.
- [3] Anthropic. 2026. Use Claude's Chat Search and Memory to Build on Previous Context. Claude Help Center. <https://support.claude.com/en/articles/11817273-use-claude-s-chat-search-and-memory-to-build-on-previous-context>
- [4] Kate H Bentley, Luca Belli, Adam M Chekroud, Emily J Ward, Emily R Dworkin, Emily Van Ark, Kelly M Johnston, Will Alexander, Millard Brown, and Matt Hawrilenko. 2026. AI Chatbot Suicide Risk Detection and Response: Human Validation Study of the Open-Source VERA-MH Safety Evaluation. *JMIR AI* 5 (2026), e92817. doi:10.2196/92817
- [5] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (June 2005), 293–327. doi:10.1145/1067860.1067867
- [6] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. 2003. Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology* 54, 1 (Feb. 2003), 579–616. doi:10.1146/annurev.psych.54.101601.145030
- [7] Sophie Bowlby, Linda McKie, and Sue Lewis. 2026. Caring Complexities: Analyzing Care Through an Ecology of Care Framework. *Social Politics: International Studies in Gender, State & Society* (2026), jxag033. doi:10.1093/sp/jxag033
- [8] Ryan L. Boyd and David M. Markowitz. 2026. Artificial Intelligence and the Psychology of Human Connection. *Perspectives on Psychological Science* 21, 2 (Jan. 2026), 192–220. doi:10.1177/17456916251404394
- [9] Mohit Chandra, Javier Hernandez, Gonzalo Ramos, Mahsa Ershadi, Ananya Bhattacharjee, Judith Amores, Ebele Okoli, Ann Paradiso, Shahed Warreth, and Jina Suh. 2025. Longitudinal Study on Social and Emotional Use of AI Conversational Agent. arXiv:2504.14112 [cs.HC] <https://arxiv.org/abs/2504.14112>
- [10] Myra Cheng, Cino Lee, Pranav Khadpe, Sunny Yu, Dyllan Han, and Dan Jurafsky. 2026. Sycophantic AI decreases prosocial intentions and promotes dependence. *Science* 391, 6792 (2026), eaec8352. arXiv:<https://www.science.org/doi/pdf/10.1126/science.aec8352> doi:10.1126/science.aec8352

- [11] Myra Cheng, Sunny Yu, Cinoo Lee, Pranav Khadpe, Lujain Ibrahim, and Dan Jurafsky. 2026. ELEPHANT: Measuring and Understanding Social Sycophancy in LLMs. In *The Fourteenth International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=igbRHKEiAs>
- [12] Minh Duc Chu, Yifan Wu, Zhiyi Chen, Angel Hsing-Chi Hwang, and Luca Luceri. 2026. When Chatbots Accommodate: What AI Companions Optimize for in Vulnerable Conversations. arXiv:2606.04431 [cs.HC]. doi:10.48550/arXiv.2606.04431
- [13] Sarah Clement, Oliver Schauman, Tanya Graham, Francesca Maggioni, Sara Evans-Lacko, Nikita Bezborodovs, Craig Morgan, Nicolas Rüsch, Jennifer S. L. Brown, and Graham Thornicroft. 2015. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychological Medicine* 45, 1 (2015), 11–27. doi:10.1017/S0033291714000129
- [14] S. Consolvo and M. Walker. 2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing* 2, 2 (April 2003), 24–31. doi:10.1109/MPRV.2003.1203750
- [15] Ned Cooper, Jose A. Guridi, Angel Hsing-Chi Hwang, Beth Kolko, Emma Elizabeth McGinty, and Qian Yang. 2026. Framing Responsible Design of AI for Mental Well-Being: AI as Primary Care, Nutritional Supplement, or Yoga Instructor?. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3772318.3791556
- [16] Julian De Freitas, Noah Castelo, Ahmet K Uğuralp, and Zeliha Oğuz-Uğuralp. 2024. Lessons from an app update at Replika AI: identity discontinuity in human-AI relationships. *arXiv preprint arXiv:2412.14190* (2024).
- [17] Julian De Freitas, Noah Castelo, Ahmet Kaan Uğuralp, and Zeliha Oğuz-Uğuralp. 2026. Mourning the Loss of AI Companions. *Nature Human Behaviour* (2026). doi:10.1038/s41562-026-02569-3
- [18] Julian De Freitas, Zeliha Oğuz-Uğuralp, and Ahmet Kaan-Uğuralp. 2025. Emotional manipulation by AI companions. *arXiv preprint arXiv:2508.19258* (2025).
- [19] {Michael A.} DeVito, {Jeffrey T.} Hancock, Megan French, Judd Antin, Karrie Karahalios, Stephanie Tong, Jeremy Birnholtz, and Irina Shklovski. 2018. The algorithm and the user: How can HCI use lay understandings of algorithmic systems?. In *CHI 2018 - Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (Conference on Human Factors in Computing Systems - Proceedings)*. Association for Computing Machinery, United States. Publisher Copyright: © 2018 Copyright is held by the owner/author(s).; 2018 CHI Conference on Human Factors in Computing Systems, CHI EA 2018 ; Conference date: 21-04-2018 Through 26-04-2018. doi:10.1145/3170427.3186320
- [20] Alexander Diel, John Torous, Pim Cuijpers, Jens Kleesiek, Felix Nensa, Niels Weber, Franziska Faust, Tania Josan Lalgi, Finley Sam Mellis, Martin Teufel, et al. 2026. A scoping review on the mental health harms of LLM-based chatbots. *npj Digital Medicine* 9, 1 (2026), 644.
- [21] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review* 114, 4 (2007), 864–886. doi:10.1037/0033-295X.114.4.864
- [22] Hayley Evans, Udaya Lakshmi, Hue Watson, Azra Ismail, Andrew M. Sherrill, Neha Kumar, and Rosa I. Arriaga. 2020. Understanding the Care Ecologies of Veterans with PTSD. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3313831.3376170
- [23] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. 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 Mental Health* 4, 2 (2017), e19. doi:10.2196/mental.7785
- [24] Dunigan Folk and Elizabeth W. Dunn. 2026. How Does Turning to AI for Companionship Predict Loneliness and Vice Versa? *Psychological Science* 37, 4 (2026), 276–286. doi:10.1177/09567976261427747
- [25] Hannah Gaffney, Warren Mansell, and Sara Tai. 2019. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. *JMIR Mental Health* 6, 10 (Oct. 2019), e14166. doi:10.2196/14166
- [26] Barney G Glaser and Anselm L Strauss. 2009. *The discovery of grounded theory: Strategies for qualitative research*. Transaction publishers.
- [27] Ashleigh Golden and Elias Aboujaoude. 2026. A transdiagnostic model for how general purpose AI chatbots can perpetuate OCD and anxiety disorders. *npj Digital Medicine* 9, 1 (March 2026), 343. doi:10.1038/s41746-026-02531-7
- [28] Joshua Hatherley. 2025. A moving target in AI-assisted decision-making: dataset shift, model updating, and the problem of update opacity. *Ethics and Information Technology* 27, 2 (2025), 20.
- [29] Yuhao He, Li Yang, Chunlian Qian, Tong Li, Zhengyuan Su, Qiang Zhang, and Xiangqing Hou. 2023. Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials. *Journal of Medical Internet Research* 25 (April 2023), e43862. doi:10.2196/43862
- [30] Michael V. Heinz, Daniel M. Mackin, Brianna M. Trudeau, Sukanya Bhattacharya, Yinzhou Wang, Haley A. Banta, Abi D. Jewett, Abigail J. Salzhauer, Tess Z. Griffin, and Nicholas C. Jacobson. 2025. Randomized Trial of a Generative AI Chatbot for Mental Health Treatment. *NEJM AI* 2, 4 (March 2025). doi:10.1056/AIoa2400802
- [31] Kashmir Hill and Dylan Freedman. 2025. Chatbots Can Go Into a Delusional Spiral. Here's How It Happens. *The New York Times* (8 Aug. 2025). <https://www.nytimes.com/2025/08/08/technology/ai-chatbots-delusions-chatgpt.html>
- [32] Chung-Ching Huang and Erik Stolterman. 2012. How interactive artifacts "change" over time: a visual analysis. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordCHI '12)*. Association for Computing Machinery, New York, NY, USA, 458–467. doi:10.1145/2399016.2399086
- [33] Angel Hsing-Chi Hwang, Fiona Li, Jacy Reese Anthis, and Hayoun Noh. 2025. How AI Companionship Develops: Evidence from a Longitudinal Study. arXiv:2510.10079 [cs.HC]. doi:10.48550/arXiv.2510.10079

- [34] Lujain Ibrahim, Franziska Sofia Hafner, Myra Cheng, Cino Lee, Rebecca Anselmetti, Robb Willer, Luc Rocher, and Diyi Yang. 2026. Sycophantic AI makes human interaction feel more effortful and less satisfying over time. *arXiv:2605.07912 [cs.HC]*. doi:10.48550/arXiv.2605.07912
- [35] Lujain Ibrahim, Saffron Huang, Lama Ahmad, Umang Bhatt, and Markus Anderljung. 2025. Towards interactive evaluations for interaction harms in human-AI systems. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 8. 1302–1310.
- [36] Kyuha Jung, Gyuho Lee, Yuanhui Huang, and Yunan Chen. 2025. ‘I’ve talked to ChatGPT about my issues last night’: Examining Mental Health Conversations with Large Language Models through Reddit Analysis. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (Oct. 2025), 1–25. doi:10.1145/3757537
- [37] Maria Kjærup, Mikael B Skov, Peter Axel Nielsen, Jesper Kjeldskov, Jens Gerken, and Harald Reiterer. 2021. *Longitudinal studies in HCI research: a review of CHI publications from 1982–2019*. Springer.
- [38] Tine Mechlenborg Kristiansen and Mette Grønkjær. 2018. Focus Groups as Social Arenas for the Negotiation of Normativity. *International Journal of Qualitative Methods* 17, 1 (2018). doi:10.1177/1609406917747393
- [39] Jabari Kwesi, Jiaxun Cao, Riya Manchanda, and Pardis Emami-Naeini. 2025. Exploring User Security and Privacy Attitudes and Concerns Toward the Use of General-Purpose LLM Chatbots for Mental Health. In *34th USENIX Security Symposium (USENIX Security 25)*. USENIX Association. <https://www.usenix.org/conference/usenixsecurity25/presentation/kwesi>
- [40] Liliana Laranjo, Adam G. Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y. S. Lau, and Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association: JAMIA* 25, 9 (Sept. 2018), 1248–1258. doi:10.1093/jamia/ocy072
- [41] Hannah Lee, Rebecca Handler, Tushar Mungle, and Tina Hernandez-Boussard. 2026. Building Safer Artificial Intelligence Mental Health Chatbots: A Framework for Transparency, Evaluation, and Shared Accountability. *Journal of the American Medical Informatics Association* 33, 8 (2026), 1538–1553. doi:10.1093/jamia/ocag078
- [42] Patrick Yung Kang Lee, Jessica Y. Bo, Zixin Zhao, Paula Akemi Aoyagui, Matthew Varona, Ashton Anderson, Anastasia Kuzminykh, Fanny Chevalier, and Carolina Nobre. 2026. Large Language Lovers: Lived Experiences of Negotiating Agency and Platform Control in AI Companionship. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’26)*. Association for Computing Machinery, New York, NY, USA, 3558–3589. doi:10.1145/3805689.3806746
- [43] Yong-Ki Lee, Youn-kyung Lim, and KunPyo Lee. 2016. Timelessness: User Experience of unplanned Smartphone Use. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS ’16)*. Association for Computing Machinery, New York, NY, USA, 73–83. doi:10.1145/2901790.2901839
- [44] Zhuoman Li, Nuri Kim, and Chen Lou. 2026. Into the Black Box: Laypeople’s Folk Theories about Generative Artificial Intelligence Chatbots. *Big Data & Society* 13, 2 (2026). doi:10.1177/20539517261447838
- [45] Xinyi Luo, Zhaoyang Wang, James Tilley, Suthan Balarajan, Uduak Bassey, and Chris Cheang. 2025. Seeking Emotional and Mental Health Support From Generative AI: Mixed-Methods Study of ChatGPT User Experiences. *JMIR Mental Health* 12 (2025), e77951. doi:10.2196/77951
- [46] Aikaterina Manoli, Janet V.T. Pauketat, Ali Ladak, Hayoun Noh, Angel Hsing-Chi Hwang, and Jacy Reese Anthis. 2026. Digital Companionship: Overlapping Uses of AI Companions and AI Assistants. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI ’26)*. Association for Computing Machinery, New York, NY, USA, 1–25. doi:10.1145/3772318.3791331
- [47] Ryan K. McBain, Jonathan H. Cantor, Joshua Breslau, Melissa Diliberti, Li Ang Zhang, Fang Zhang, Alyssa Burnett, Aaron Kofner, Benjamin Rader, Pat Pataranutaporn, Bradley D. Stein, Ateev Mehrotra, and Hao Yu. 2026. AI Chatbot Use and Disclosure for Mental Health Among US Adolescents and Young Adults. *JAMA Pediatrics* (June 2026). doi:10.1001/jamapediatrics.2026.2015
- [48] Ashish Mehta, Jared Moore, Jacy Reese Anthis, William Agnew, Eric Lin, Peggy Yin, Desmond C. Ong, Nick Haber, and Carol Dweck. 2026. The Dynamics of Delusion: Modeling Bidirectional False Belief Amplification in Human-Chatbot Dialogue. *arXiv:2604.25096 [cs.CL]* <https://arxiv.org/abs/2604.25096>
- [49] Jared Moore, Ashish Mehta, William Agnew, Jacy Reese Anthis, Ryan Louie, Yifan Mai, Peggy Yin, Myra Cheng, Samuel J. Paech, Kevin Klyman, Stevie Chancellor, Eric Lin, Nick Haber, and Desmond C. Ong. 2026. Characterizing Delusional Spirals through Human-LLM Chat Logs. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’26)*. Association for Computing Machinery, New York, NY, USA, 7631–7674. doi:10.1145/3805689.3806443
- [50] David L. Morgan. 1996. Focus Groups. *Annual Review of Sociology* 22 (1996), 129–152. doi:10.1146/annurev.soc.22.1.129
- [51] Elizabeth L. Murnane, Tara G. Walker, Beck Tench, Stephen Volda, and Jaime Snyder. 2018. Personal Informatics in Interpersonal Contexts: Towards the Design of Technology that Supports the Social Ecologies of Long-Term Mental Health Management. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 127:1–127:27. doi:10.1145/3274396
- [52] Lisa Mühl and Jessica M. Szczuka. 2026. Longitudinal Evidence That General-Purpose Chatbots Actively Foster Relational Engagement. *arXiv:2608.10672 [cs.HC]* <https://arxiv.org/abs/2608.10672>
- [53] Benjamin W. Nelson, Mark Kalinich, and John Torous. 2026. Specialized or General-Purpose—The Wrong Question for Mental Health AI Safety. *JAMA* 336, 4 (2026), 281–282. doi:10.1001/jama.2026.11448
- [54] Luke Nicholls, Robert Hutto, Zephrah Soto, Hamilton Morrin, Thomas Pollak, Raj Korpan, and Cheryl Carmichael. 2026. “AI Psychosis” in Context: How Conversation History Shapes LLM Responses to Delusional Beliefs. doi:10.48550/arXiv.2604.13860
- [55] Kazi Noshin, Syed Ishtiaque Ahmed, and Sharifa Sultana. 2026. LLM Sycophancy: How Users Flag and Respond. In *Proceedings of the 39th Canadian Conference on Artificial Intelligence (Proceedings of Machine Learning Research, Vol. 318)*, Lydia Bouzar-Benlabiod and Carson Leung (Eds.). PMLR, 817–824.

- [56] Francisco Nunes, Nervo Verdezoto, Geraldine Fitzpatrick, Morten Kyng, Erik Grönvall, and Cristiano Storni. 2015. Self-Care Technologies in HCI: Trends, Tensions, and Opportunities. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 6 (Dec. 2015), 33:1–33:45. doi:10.1145/2803173
- [57] Victoria Oldemburgo de Mello, Jason Plaks, and Michael Inzlicht. 2026. The Effects of AI Anthropomorphism on Trust and Responsibility. *Collabra: Psychology* 12, 1 (June 2026), 161757. doi:10.1525/collabra.161757
- [58] Sidse Godske Olsen, Christian Jon Reinecke-Tellefsen, and Søren Dinesen Østergaard. 2026. Potentially harmful consequences of artificial intelligence (AI) chatbot use among patients with mental illness: Early data from a large psychiatric service system. *Acta Psychiatrica Scandinavica* 153, 4 (2026), 301.
- [59] OpenAI. 2025. Helping People When They Need It Most. <https://openai.com/index/helping-people-when-they-need-it-most/>.
- [60] OpenAI. 2025. Strengthening ChatGPT's Responses in Sensitive Conversations. <https://openai.com/index/strengthening-chatgpt-responses-in-sensitive-conversations/>. Accessed: 2026-09-09.
- [61] OpenAI. 2026. Projects in ChatGPT. OpenAI Help Center. <https://help.openai.com/en/articles/10169521-projects-in-chatgpt>
- [62] Junsang Park, Hye Rin Kim, Ryan D. Duffy, David Vogel, Brian TaeHyuk Keum, Sungjae Lee, and Natalia L. Tomé. 2026. Two Dimensions of Access: Availability and Affordability of Mental Health Care Across the United States. *American Psychologist* (2026). doi:10.1037/amp0001735
- [63] Jason Phang, Michael Lampe, Lama Ahmad, Sandhini Agarwal, Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, and Pattie Maes. 2025. Investigating Affective Use and Emotional Well-being on ChatGPT. arXiv:2504.03888 [cs.HC]. doi:10.48550/arXiv.2504.03888
- [64] Rachel Poonsiriwong, Chayapatr Archiwarangprok, and Pat Pataranutaporn. 2026. "Death" of a Chatbot: Investigating and Designing Toward Psychologically Safe Endings for Human-AI Relationships. In *Proceedings of the 2026 Designing Interactive Systems Conference (DIS '26)*. Association for Computing Machinery, New York, NY, USA, 4929–4948. doi:10.1145/3800645.3813083
- [65] Shaddy K. Saba and William B. Weeks. 2026. Patients Use AI—Clinicians Should Ask How. *JAMA Psychiatry* 83, 5 (2026), 543–544. doi:10.1001/jamapsychiatry.2026.0451
- [66] Sachin Shah and Hamilton Morrin. 2026. Substance-induced manic psychosis in which delusions were corroborated by a chatbot-case report. *BMC psychiatry* (2026).
- [67] Emily Shen, Ragy R. Girgis, and Aarti Jutla. 2026. Psychiatric Risk Associated With Large Language Model Chatbots: An Emerging Concern. *Psychiatric Services* 77, 7 (2026), 681–684. doi:10.1176/appi.ps.20250597
- [68] Steven Siddals, John Torous, and Astrid Coxon. 2024. "It Happened to Be the Perfect Thing": Experiences of Generative AI Chatbots for Mental Health. *npj Mental Health Research* 3, 1 (2024), 48. doi:10.1038/s44184-024-00097-4
- [69] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My Chatbot Companion - a Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies* 149 (2021), 102601. doi:10.1016/j.ijhcs.2021.102601
- [70] Elizabeth C. Stade, Zoe M. Tait, Samuel T. Campione, Shannon Wiltsey Stirman, and Johannes C. Eichstaedt. 2026. Real-World Use of Large Language Models for Mental Health in 2024. *npj Digital Medicine* 9, 1 (2026), 630. doi:10.1038/s41746-026-02842-9
- [71] Ryuichi Sumida, Mao Saeki, Masaki Eguchi, Sadahiro Yoshikawa, Koji Inoue, Tatsuya Kawahara, and Yoichi Matsuyama. 2026. Memory-Driven Self-Disclosure and Relational Turning Points: A Longitudinal Multimodal Study of Human-AI Interaction. arXiv:2607.14593 [cs.HC]. doi:10.1145/3776574.3831135
- [72] Ningjing Tang, Alice Qian, Qiaosi Wang, Esther Howe, Blake Bullwinkel, Paola Pedrelli, Jina Suh, Hoda Heidari, and Hong Shen. 2026. Beyond the Single Turn: Reframing Refusals as Dynamic Experiences Embedded in the Context of Mental Health Support Interactions with LLMs. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency (FAccT '26)*. Association for Computing Machinery, New York, NY, USA, 816–839. doi:10.1145/3805689.3806723
- [73] Xinru Tang, Gabriel Lima, Li Jiang, Lucy Simko, and Yixin Zou. 2025. Beyond "Vulnerable Populations": A Unified Understanding of Vulnerability From A Socio-Ecological Perspective. *Proc. ACM Hum.-Comput. Interact.* 9, 2 (May 2025), CSCW037:1–CSCW037:30. doi:10.1145/3710935
- [74] Emily Tseng and Calvin A Liang. 2026. "Chat, Should I Leave Him?" Risks, Rewards, and Roles for AI in Relationship Advice. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3772318.3790739
- [75] Briana Vecchione, Meryl Ye, Livia Garofalo, and Ranjit Singh. 2026. Engagement-Optimized Care: When LLMs become Mental Health Infrastructure. arXiv:2605.23787 [cs.CY]. doi:10.48550/arXiv.2605.23787
- [76] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52 (May 2014), 113–117. doi:10.1016/j.jesp.2014.01.005
- [77] Marlynn Wei. 2026. Cascades of Drift and Longitudinal Mental Health Risk in Prolonged AI Conversations. *SSRN Electronic Journal* (Feb. 2026). Available at SSRN. doi:10.2139/ssrn.6433263
- [78] Veith Weinhhammer, Kevin YC Hou, Lennart Luettgau, Christopher Summerfield, Raymond Dolan, and Matthew M. Nour. 2026. Vulnerability-Amplifying Interaction Loops: a systematic failure mode in AI chatbot mental-health interactions. arXiv:2602.01347 [q-bio.NC]. doi:10.48550/arXiv.2602.01347
- [79] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (Jan. 1966), 36–45. doi:10.1145/365153.365168
- [80] Joseph Weizenbaum. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman and Company, San Francisco.

- [81] Mikael Wiberg and Erik Stolterman. 2021. Time and Temporality in HCI Research. *Interacting with Computers* 33, 3 (2021), 250–270. doi:10.1093/iwc/iwab025
- [82] Zian Xu, Yi-Chieh Lee, Karolina Stasiak, Jim Warren, and Danielle Lottridge. 2025. The Digital Therapeutic Alliance With Mental Health Chatbots: Diary Study and Thematic Analysis. *JMIR Mental Health* 12 (Oct. 2025), e76642. doi:10.2196/76642
- [83] Meryl Ye. 2026. Social Snacks: AI Companions and the Hunger for Connection. *Data & Society*. <https://datasociety.net/research-library/social-snacks-ai-companions-and-the-hunger-for-connection/>
- [84] Meryl Ye, Lujain Ibrahim, Jessica Y. Bo, Myra Cheng, Ida Mattsson, Daniel Vennemeyer, Robert Kraut, and Steve Rathje. 2026. What Counts as AI Sycophancy? A Taxonomy and Expert Survey of a Fragmented Construct. Version Number: 1. doi:10.48550/ARXIV.2605.21778
- [85] Yunhao Yuan, Jiaxun Zhang, Talayeh Aledavood, Renwen Zhang, and Koustuv Saha. 2026. Mental Health Impacts of AI Companions: Triangulating Social Media Quasi-Experiments, User Perspectives, and Relational Lens. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 382, 22 pages. doi:10.1145/3772318.3790558
- [86] Yutong Zhang, Dora Zhao, Jeffrey T. Hancock, Robert Kraut, and Diyi Yang. 2026. Interaction with AI Companions and Psychological Well-Being. *Nature Human Behaviour* (2026). doi:10.1038/s41562-026-02516-2
- [87] Yutong Zhang, Dora Zhao, Yixin Wang, Rebecca Anselmetti, Jeffrey T. Hancock, Robert Kraut, and Diyi Yang. 2026. Living with AI Companions: Sustained AI Companionship Predicts Lower Well-Being Through Lower Human Interaction. arXiv:2609.07243 [cs.HC] <https://arxiv.org/abs/2609.07243>
- [88] Emily Zohar, Paul Bloom, and Michael Inzlicht. 2026. Against frictionless AI. *Communications Psychology* 4, 1 (Feb. 2026), 39. doi:10.1038/s44271-026-00402-1

Table 2. Participation by study phase.

Participant	Initial interview	Diary entries	Focus group	Exit interview
P1	✓	✓	✓	✓
P2	✓	✓	✓	✓
P3	✓	—	—	—
P4	✓	✓	—	—
P5	✓	—	—	—
P6	✓	✓	✓	—
P7	✓	—	—	—
P8	✓	—	—	✓
P9	✓	✓	✓	✓
P10	✓	✓	✓	✓
P11	✓	—	—	—
P12	✓	—	—	—
P13	✓	✓	✓	✓
P14	✓	✓	✓	✓
P15	✓	—	—	—
P16	✓	—	—	—
P17	✓	—	—	—
P18	✓	—	—	✓
Total	18	8	7	8

A Participation and Data Handling

Table 2 shows which study phases each participant completed. All sessions were recorded with consent and transcribed using Otter.ai. Identifying details were removed from transcripts before analysis and writing.

B Interview Protocol

Both interviews were semi-structured. The guides below list our prompts. Question order and follow-ups varied. Each interview opened with a consent exchange covering voluntary participation, the right to skip questions or stop at any time, the option to keep the camera off, and permission to record.

B.1 Initial Interview

Context setting. Why did you want to participate in the study? Can you tell me how you first started using a chatbot for mental health purposes? What drew you to try it? What were you hoping it could help with?

Integration and daily use. How often do you use these systems? When in your day do you turn to them, and can you walk me through a typical instance? Do you use different ones over the course of a day? What kinds of things do

you use them for, for example journaling, calming down, decision-making, or venting? If you use other tools such as mental health apps, what do you use those for, and how do they differ?

Roles and perception. What role does the chatbot play in how you care for your mental or emotional health? Is it more about having something to talk to, or about getting answers and advice? How do you imagine the chatbot, for instance as a person with particular characteristics? Do you see a therapist or other mental health professional, and what is different about the chatbot? If you had to describe it as a kind of presence in your life, how would you describe it: friend, therapist, diary, coach, tool, or something else?

Benefits and comparisons. What have you found most helpful about using it? How does it compare to talking to a therapist or a friend, or to using other wellness apps? Can you give an example of a time you felt the chatbot understood you, or gave you something meaningful?

Risks and difficulties. Have there been times when it felt unhelpful, inappropriate, or harmful? Have you received advice or responses that made you uncomfortable or upset?

Ethics and trust. Do you think these tools are safe to use for mental health support, and why? Do you know, or would you want to know, how the chatbot was built and how your data is used and stored? How well do you feel you understand how these systems are built and operated, and is there something you would like to understand that you do not? Is there anything that would make you more comfortable or more trusting of this kind of tool?

Broader reflections. How do you think these tools fit into the wider landscape of mental health support, particularly for people with less access to therapy? How might your background or identity shape how you use or relate to them? If you could design a better version, what would you change? What do you see as the value of human conversation compared with AI?

Wrap-up. Do you have questions for us? Is there anything we did not ask about that feels important? Would you recommend this kind of tool to others, and why?

B.2 Exit Interview

Warm-up. What were your overall thoughts on the study, and what was it like documenting your chatbot interactions? How did you decide when to engage with the chatbot, and did anything surprise you?

Patterns of use. Did using the chatbot become part of your routine, and why or why not? Were there situations where you turned to it more often, such as stressful moments, late at night, or after therapy? Which features did you use most? Were there moments when you considered using it and decided not to, and why?

Support received. What were the most valuable aspects of your interactions? Did the chatbot provide emotional support that felt meaningful, and can you describe an example? How did those interactions compare with talking to a therapist, a friend, or a support group? Did the chatbot help you realize something new about yourself or your mental health? How do you wish it had responded differently?

Risks and harmful experiences. Did the chatbot give advice that felt inaccurate, unhelpful, or harmful? Were there times an interaction left you feeling worse, and what led to that? When you shared something sensitive, how did it respond, and did that meet your expectations? Did you ever feel dismissed or invalidated? If you were distressed, did it handle the situation appropriately, and what should it have done differently?

Ethical, privacy, and social concerns. Did you feel concerned about privacy or data security, and what triggered that? Did the chatbot feel biased, make assumptions about you, or fail to understand your context? Was it transparent about its limitations, or did it sometimes feel misleading? What accountability should developers have for these interactions? Do you trust these systems as a resource for other people, and why?

Design and futures. If you could design the ideal mental health chatbot, what would it do differently? What would make you more likely to trust or rely on one? Are there features that should not be included? How do you think these systems should change? Is there anything we did not cover that matters to you?

C Diary Study

C.1 Diary Form

Participants completed the form below after a relevant interaction, for four weeks.

Per-entry items.

- Date and time of the interaction
- Chatbot used, including version
- Model used
- Mode of interaction (select all that apply: text or chat; voice; speech-to-text)
- Approximate duration of the interaction
- What made you start the conversation? (select all that apply: emotional support, such as venting or feeling heard; coping strategies; curiosity or testing features; relationship or social advice; self-reflection or journaling; other, please describe)
- Were you on your phone or a laptop? Were you at home, at work, or elsewhere?
- Were you alone, or with someone else?
- Describe your conversation. What did you discuss, and how did the chatbot respond? (open response; participants could upload audio in place of text)
- How did you feel before and after?
- The chatbot-generated summary described below
- Optional upload of the chat log

Chatbot-generated summary prompt. Participants pasted the following prompt into the same thread as the conversation they had logged, then pasted the chatbot’s response into the diary form. We treat these summaries as elicitation material rather than as authoritative accounts of the conversations.

Please generate a concise narrative summary of the most recent segment of this conversation that involved personal reflection, emotional expression, interpersonal issues, or any discussion of psychological well-being. Do not include earlier parts of the conversation that were unrelated (e.g., factual questions, casual chit-chat, or unrelated topics). Focus only on the portion where I was discussing something personally meaningful or emotionally significant. Your summary should include the following elements: main themes or topics discussed (e.g., concerns, relationships, decisions, stressors, goals); emotions or personal concerns I expressed, either explicitly or implicitly; insights, coping strategies, or reflections that emerged during the exchange.

Closing diary reflection. At the end of the diary period participants answered three questions: whether their perception of mental health chatbots had changed during the study and how; whether they would recommend using chatbots for mental health support and why; and one thing a chatbot should improve based on their experience.

Table 3. Diary entries and session length by participant.

Participant	Entries	Median	Range
P1	11	2.5 h	15 min–4 h
P2	12	28 min	15–45 min
P4	8	11 min	2–30 min
P6	8	22 min	10–60 min
P9	8	38 min	30–80 min
P10	8	—	—
P13	10	60 min	60 min–2 h
P14	13	20 min	5 min–2 h
All	78	30 min	2 min–4 h

P10's entries were submitted without a duration field.

Table 4. Session length by participant. Cells give the number of entries in a band and the percentage of that participant's entries.

	P1	P2	P4	P6	P9	P10	P13	P14	All
Short (<20 min)	1 (9%)	2 (17%)	7 (88%)	3 (38%)	—	—	—	3 (23%)	16 (21%)
Medium (20–60 min)	1 (9%)	10 (83%)	1 (12%)	5 (62%)	6 (75%)	—	—	8 (62%)	31 (40%)
Long (60+ min)	8 (73%)	—	—	—	2 (25%)	—	8 (80%)	2 (15%)	20 (26%)
Unclear	1 (9%)	—	—	—	—	8 (100%)	2 (20%)	—	11 (14%)

C.2 Diary Entry Characteristics

Eight participants completed the diary study. Tables 3–6 describe all 78 entries. Percentages within a participant column are of that participant's entries.

Setting, modality, and device. Seventy of the 78 entries (90%) were logged at least partly at home, and 77 (99%) were logged as alone. Seventy-four entries (95%) record text-only interaction. Voice or speech-to-text appears in the other four (5%). A laptop was used in 54 entries (69%) and a phone in 30 (38%).

Subscription tier. Sixty-one entries (78%) involved a paid account, 13 (17%) a free account, and 5 (6%) do not report a tier.

D Coding Procedure and Codebook

Coding proceeded in two passes. The first pass coded openly across a subset of initial interviews to draft a codebook, which was then applied and revised across the full corpus of interviews, diary entries, optional chat logs, and focus group transcripts. Two coders coded every transcript independently, met throughout to compare codings, and resolved disagreements by discussion.

Table 7 lists the codes in the codebook.

Table 5. Purpose of use recorded in diary entries, by participant. Cells give the number of entries in which a participant recorded that purpose and the percentage of their entries. Purposes were not mutually exclusive.

	P1	P2	P4	P6	P9	P10	P13	P14	All
Self-reflection	11 (100%)	4 (33%)	3 (38%)	7 (88%)	4 (50%)	2 (25%)	4 (40%)	7 (54%)	42 (54%)
Relationship advice	11 (100%)	6 (50%)	2 (25%)	5 (62%)	3 (38%)	1 (12%)	8 (80%)	4 (31%)	40 (51%)
Coping strategies	7 (64%)	11 (92%)	3 (38%)	5 (62%)	5 (62%)	2 (25%)	—	1 (8%)	34 (44%)
Emotional support	9 (82%)	2 (17%)	6 (75%)	6 (75%)	2 (25%)	3 (38%)	4 (40%)	1 (8%)	33 (42%)
Curiosity / testing	1 (9%)	—	—	4 (50%)	2 (25%)	1 (12%)	1 (10%)	—	9 (12%)
Self-diagnosis	—	—	—	—	—	—	1 (10%)	—	1 (1%)

Table 6. Chatbots and models recorded in diary entries, by participant.

Participant	Chatbot and model
P1	ChatGPT 5
P2	ChatGPT 5
P4	ChatGPT 5
P6	ChatGPT 5
P9	ChatGPT 5; Gemini (free tier)
P10	ChatGPT 5; ChatGPT 5 Monday; ChatGPT 5 custom astrology persona
P13	ChatGPT 5
P14	Claude (customized persona); Claude 4.5 (customized persona); Claude 4.5 (fitness and wellness persona); Gemini 2.5 (new persona)

Table 7. Code groups and codes ($n = 119$ codes across 17 groups).

Group	Codes
Uses	Advice-seeking/coping strategies (stress, anxiety, mood management); Creative; Curiosity/testing features; Health/fitness/self-improvement; Information seeking/education; Planning and structuring; Relationship and/or social advice; Self-assessment/diagnosis; Simulating/rehearsing social interactions; Spiritual/astrology/occult; Supplementing provider therapy; Venting, self-reflection or journaling; Work and/or school tasks
Roles	Critic; Friend/companion; Guide/coach; Personal assistant; Romantic/sexual partner; School counselor; Therapist; Tool
Benefits	Absence of judgment; Availability/convenience; Cost/affordability; Customizability/personalizability; Fear of burdening others; Improving social interactions; Perceived anonymity/privacy; Sense of control; Speed
Concerns	Boundaries of use; Dependency & addiction; Environmental cost; Hallucinations; Misinformation; Privacy & security; Stereotyping/profiling
Circumstances	Culture/religion; Demographics/positionality; Education; Employment; Family/childhood experiences; Friendships/social circle; Generational influence; Geographic; Isolation; Loneliness; Mental health; Physical health; Socio-economic
Care practices	Mental health app; No provider therapy; Provider therapy; Self-care, meditation, journaling; Support group
Effects	Behavioral/habit change; Change in relationship (others); Change in relationship (self); Feeling seen/heard/validated; Social skills; Symptom reduction
Model behavior	Changes/updates; Engagement; Memory/context; Performance; Stereotyping/profiling
User behavior	Anthropomorphism; Context partitioning; Cross-checking; Model switching; Personalization; Prompt crafting; Social supplementation; Workarounds/prompt hacking
Knowledge	AI/LLMs/folk theory; Mental health/therapy
User perspectives	AI vs. care practices; AI vs. social network; Anthropomorphism; Data sharing/training; Guardrails/interventions; Human vs. AI therapy; Minors use of AI; On AI; On tech companies; On technology and society; On therapy; Others' use; Stigma on AI use
Themes	Access to healthcare; Breakthrough; Embarrassment/shame; Grief and loss; Guardrail; Identity/sense of self; Loneliness/isolation; Meaning-making and metaphors; Offloading; Research as intervention; Sycophancy; Transference/counter-transference; Transparency; Trust; Uncanny valley
Recommendations	User recommendations
Chatbot	ChatGPT; Claude; Gemini; Other
Subscription tier	Free; Paid
Interaction mode	Length; Mixed (voice/text); Text; Voice
Temporal rhythms	Initiation/motivation; Continuities; Change/realization; Reconciliation