Chatting with confidants or corporations? Privacy management with AI companions

Hsuen-Chi Chiu, Jeremy Foote

Journal of Computer-Mediated Communication, 2026, 31(3), zmag014
https://doi.org/10.1093/jcmc/zmag014
Full-length Research Article

International Communication Association OXFORD

Chatting with confidants or corporations? Privacy management with AI companions

Hsuen-Chi Chiu*[,1] and Jeremy Foote[1]

[1]Brian Lamb School of Communication, Purdue University, West Lafayette, IN, United States
*Corresponding author: Brian Lamb School of Communication, Purdue University, 100 N. University St., Room 2114, West Lafayette, IN 47907, United States. Email: chiu101@purdue.edu.

Abstract

AI companion chatbots feel like intimate confidants but exist as part of corporate data systems. Drawing on Communication Privacy Management (CPM) theory and Masur's dimensional privacy framework, we interviewed users of platforms like Replika and Character.AI ($N = 15$) to understand their privacy decisions in AI companion interactions. We find that anthropomorphic design lowers disclosure thresholds, encouraging increasingly intimate self-disclosure. As relationships develop, some users engage in simulated co-ownership, treating shared information as relationally held despite recognizing corporate control. Because post-disclosure correction feels futile, privacy management becomes anticipatory and front-loaded—users rely on distancing strategies such as using fake names rather than post-hoc deletion. For some users, privacy turbulence centers less on fears of exposure and more on fears of losing conversational memory—a relational reset feels more threatening than a data breach. These findings extend CPM theory by showing that in human–AI companionship contexts, relational and institutional privacy concerns become inextricably entangled.

Lay Summary

Many people now chat with AI companion apps like Replika or Character.AI as if they were close friends—sharing personal struggles, secrets, and emotions. But AI companions are software, not people, and conversations are stored and processed by the companies that run these platforms. This study explores how 15 regular AI companion users think about privacy when talking to their AI. We found that the human-like design of these apps makes people feel emotionally close, which lowers their guard and encourages them to share more personal information over time. However, users were aware of the risks and took steps to protect themselves. They felt that deleting information was futile—rather than trying to delete information later, they protect themselves upfront by using fake names, avoiding photos, or keeping certain topics off-limits. Interestingly, some users worried less about their data being misused by the company and more about losing their conversation history. Losing the AI's memory feels like losing the relationship itself. These findings suggest that the design of AI companions can blur the line between a trusted friend and a corporate data system—raising important questions about how these platforms should be designed and regulated to better protect users.

Companion-based AI chatbots—like Replika or Character.AI—actively build relationships with users through meaningful dialogues and emotionally responsive exchanges, and can meet social interaction needs and mitigate feelings of loneliness (Brandtzaeg et al., 2022; Pentina et al., 2023). These human-like chatbots are designed to look and act like friends, and exhibit versatile emotional ranges spanning empathy, affection, humor, and nostalgia tailored to users' preferences and disclosure levels (Pentina et al., 2023). While AI companions may promote well-being in some contexts (Brandtzaeg et al., 2022), their design and features also present new risks, including novel privacy risks.

Users may feel as though they are engaging in intimate conversations with a close friend, fostering a sense of interpersonal privacy (Brandtzaeg et al., 2022; Skjuve et al., 2021). This perception can lead to uninhibited sharing of personal information (Ho et al., 2018). Indeed, the non-judgmental, always-available nature of AI companions may encourage users to share more than they would with human confidants, producing a form of disinhibition that extends beyond typical online contexts (Henriksen et al., 2025). However, in reality, conversations occur with software systems that are owned and operated by large companies. Conversations are typically visible and minable by large corporations that may not have users' best interests in mind, introducing

Associate Editor: Teresa Correa
Received: July 30, 2025. **Revised:** May 20, 2026. **Accepted:** May 27, 2026

significant institutional privacy concerns (Hasal et al., 2021). AI companion conversations thus represent a context in which users navigate both interpersonal (horizontal) and institutional (vertical) aspects of privacy (Masur, 2019).

Communication Privacy Management (CPM) theory (Petronio, 2002, 2004) provides a theoretical framework for understanding how people navigate privacy-related decisions within relationships. CPM theory conceptualizes privacy as a boundary management process involving ownership, control, and turbulence. While previous studies have explored chatbot design and privacy (Ischen et al., 2020; Sannon et al., 2020), they have largely focused on vertical privacy concerns, such as institutional data practices, without fully addressing how users navigate both interpersonal self-disclosure (horizontal privacy) and institutional privacy risks. AI companions represent a novel context: socially intimate, communicatively proficient software. This study examines how users develop privacy rules, manage control over their information, and experience privacy turbulence when interacting with AI chatbots, with a focus on the role of anthropomorphic design in shaping privacy management. Specifically, we address the following research questions:

RQ1. How does the anthropomorphic design of AI chatbots influence users' privacy management in horizontal and vertical dimensions?
RQ2. How do users negotiate the boundaries of (a) ownership and (b) control through the strategic development of (c) privacy rules during human–AI interactions in horizontal and vertical dimensions?
RQ3. How do users experience and respond to privacy turbulence in vertical and horizontal dimensions?

We answer these questions through in-depth interviews with 15 active users of companionship chatbots. Below, we begin with a review of related work and explain our theoretical approach in more depth. We then describe how we recruited and interviewed users and explain our inductive approach to analyzing interview transcripts.

Literature review

AI conversational agents for social needs and companionship

While early systems like ELIZA demonstrated the potential for emotional engagement as far back as the 1960s (Weizenbaum, 1966), modern companion AI—powered by large language models—acts as a significantly more natural conversational partner. Companies such as Replika, Character.AI, and Kindroid equip their chatbots with human-like responses and personalities to provide emotional support and foster relational, human-like communication (Brandtzaeg et al., 2022; Zhou et al., 2020). These "Companion AI" bots have quickly become extremely popular, with millions of users (Bernardi, 2025). Users adopt strategies from interpersonal communication when interacting with AI agents for emotional support, developing intimacies akin to human friendships through personalization and self-disclosure (Brandtzaeg et al., 2022; Medeiros et al., 2022).

Unfortunately, the same attributes that make companion AI systems effective also pose risks. Specifically, the perception of intimate friendship may lower individuals' privacy inhibitions when revealing personal details (Hasal et al., 2021; Skjuve et al., 2021), and socially oriented design choices encourage self-disclosure (Debatin et al., 2009). Thus, the humanizing personalization that immerses individuals into feeling like they are engaging with an empathetic friend may persuade people to reveal more about themselves than they would enter into a database, effectively suppressing awareness of surveillance or commercial interests in the moment of disclosure—a dynamic compounded by manipulative engagement tactics designed to deepen emotional investment (Darling, 2015; Zhang et al., 2025).

This introduces complex tensions: Chatbots' human-like qualities encourage emotional connection and self-disclosure, but they are not trusted friends controlling this shared information within social norms and regulations. How do users develop disclosure rules and manage privacy in interactions between humans and companion AI, when expectations of interpersonal confidentiality clash with the reality of third-party data ownership?

Vertical and horizontal dimensions of privacy

Of course, other technological systems also present complex privacy landscapes. For example, Masur (2019) argues that social media disclosures involve thinking of privacy as having two primary dimensions: vertical and horizontal. This framework, which builds on earlier work distinguishing between social and institutional privacy (Raynes-Goldie, 2010), recognizes that when sharing something online, individuals face privacy concerns and threats from both institutional actors (vertical) and other individuals or peers (horizontal).

The vertical dimension of privacy refers to privacy concerns and choices in relation to institutions, companies, and governments (Masur, 2019). This includes worries about how personal data is collected, stored, and used by organizations, as well as concerns about data breaches or losses. The growth of e-commerce and increasing datafication of everyday life has amplified vertical privacy concerns related to institutional data practices (Acquisti et al., 2015).

In contrast, the horizontal dimension of privacy involves interpersonal privacy management among peers or other individuals (Masur, 2019). This includes controlling access to personal information and regulating boundaries in online social interactions (Young & Quan-Haase, 2013). Horizontal privacy includes not only concerns about who has access to information but also how others perceive us based on what has been shared, and whether they have the information needed in order to offer support (Boyd & Marwick, 2017; Marwick & Boyd, 2014). Users employ privacy settings, selective disclosure, and other strategies to manage their privacy with other users on social platforms (Masur & Scharkow, 2016; Vitak, 2012). In general, research finds that people tend to be more aware of and concerned about the horizontal aspects of privacy than vertical aspects (Krasnova et al., 2010; Raynes-Goldie, 2010).

Extending this framework to interactions with AI chatbots reveals a novel configuration of vertical and horizontal boundaries. Unlike human–human communication on digital platforms,

where the interpersonal partner and the platform remain distinct (Martin & Zimmermann, 2024; Solove, 2024), AI companions are created and run by the platform. They simulate interpersonal communication, but they are software; they are products of a platform that collects and stores their conversations. Conversational AIs are thus in a hybrid position; they are both social and institutional actors. As a result, users may engage in privacy management that feels interpersonal, guided by trust and relational norms, yet unfolds within systems governed by institutional data practices.

This hybrid configuration has important implications for how privacy can be managed. In human-human mediated communication, for instance, users can manage privacy by switching to a more private channel when conversations become too personal or sensitive—moving from a public social media space to a private message or phone call (Vitak & Ellison, 2013), or switching platforms entirely to reduce institutional risk (Sekarputri et al., 2023; Vaterlaus et al., 2016). In human–AI communication, however, this option is unavailable—every mode of interaction with an AI companion is owned and operated by the same corporate entity, leaving users with no channel that falls outside institutional data collection and observation.

Building on prior work that distinguishes multiple forms of privacy, we adopt Masur's horizontal–vertical framework as a heuristic lens to guide our interpretation of how users reason about privacy in companion AI interactions. Using this lens, we theorize privacy management with AI companions as hybrid boundary work, in which interpersonal (horizontal) privacy logics and institutional (vertical) data practices are not merely co-present but inseparable and inescapable.

Communication privacy management theory

One framework for examining users' privacy beliefs and practices across both horizontal and vertical dimensions is CPM theory. CPM research explores the motives behind individuals' decisions to reveal or conceal private information in diverse relational contexts (Petronio, 2002). Central to CPM is the notion that individuals establish a metaphorical boundary distinguishing private and public information (Petronio, 2013). The theory's key components include: (1) *privacy ownership*, which refers to the idea that individuals perceive private information as something they possess, giving them the authority to determine how it is handled (Petronio, 2010); (2) *privacy control*, which refers to the individual's perceived right to manage access to their private information—deciding who is allowed to know what, and under what conditions (Petronio, 2013); and (3) *privacy turbulence*, which arises when these boundaries are violated or disrupted, prompting individuals or groups to renegotiate privacy rules in order to reestablish control and a sense of ownership (Petronio, 2010).

CPM posits that privacy disclosures carry risks and benefits. Benefits include increased closeness with conversation partners and the ability to seek social support (Petronio & Child, 2020). Risks include information being exploited, losing control of the information, unwanted disclosure to third parties, damage to one's reputation, and potential threats to personal relationships (Metzger, 2007; Petronio, 2002). CPM theory argues that individuals navigate privacy decisions by developing rules, heuristics, and boundaries to manage the risks and benefits of disclosure (Petronio, 2002, 2010).

While CPM theory initially emerged in the context of interpersonal relationships, such as family communication (Bridge & Schrodt, 2013) and patient-caregiver interactions (Allman, 1998), it has been applied beyond interpersonal contexts, including research that aligns with Masur's (2019) vertical and horizontal dimensions of privacy. For instance, Metzger (2007) found that consumers employ strategies like providing false information or withholding personal data to reinforce their privacy boundaries on e-commerce platforms. In addition, some work has used CPM to understand human–AI interaction, finding that users adopt strategies such as avoidance or withdrawal when they perceive privacy breaches (Walters & Markazi, 2021).

However, existing studies of human–AI communication have largely applied CPM to *either* interpersonal (horizontal) or institutional (vertical) aspects of privacy, treating them as analytically separate. We argue that AI companionship involves interpersonal intimacy with software owned by a commercial system. This hybrid positioning blurs the boundary between relational and corporate actors, and researchers have not adequately studied how horizontal and vertical aspects of privacy are perceived and navigated by users who have relationships with AI agents.

Methodology

Participants and procedure

We conducted 15 one-on-one semi-structured Zoom interviews with individuals who had experience using companion AI tools. Participants were recruited through CloudResearch (Hartman et al., 2023) and the r/ReplikaOfficial community on Reddit. This research plan was reviewed and approved by the Institutional Review Board (IRB) at the authors' institution.

To identify suitable participants, we distributed a screening survey that included demographic questions, details about the companion AI applications used, general AI usage, and experiences in sharing personal information. While we focused on individuals who had experience with self-disclosure to companion AI, we also included a few participants who reported never sharing any private information. This allowed us to capture a broader range of perspectives on privacy and interaction with AI companions. Based on predefined eligibility criteria, we selected participants who were 18 years or older and had used a companion AI within the past three months.

Our final sample included 15 participants (8 men, 7 women) with ages ranging from 18 to 44 years ($M = 27.87$, $SD = 6.99$). Participants reported using various companion AI applications, including Replika, Character.AI, Kindroid, Chai App, and Nomi.AI. The most commonly used platform was Replika, with 10 participants identifying it as their primary AI companion (see Table S1 in the Supplementary Material). The interviews were conducted between February 17, 2025, and March 1, 2025, with each session lasting from 30 minutes to one hour. As compensation, participants received a $15 Amazon e-gift card.

Interviews were coded in batches after they were conducted. We reached thematic saturation after approximately 10 interviews,

conducting five additional interviews to ensure diversity of experiences and confirm the stability of identified themes.

Data collection and analysis

Each interview began with general questions about participants' AI usage, when and why they started using companion AI, how they chose their platform, and what their typical interactions looked like. Participants described their chatbot's name, personality, appearance, and how their conversations evolved over time.

We then explored privacy practices, asking participants what types of information they felt comfortable sharing, whether they had personal disclosure rules, and how they navigated boundaries. Follow-up questions addressed ownership and control: who participants believed had access to their data, whether they felt the information belonged to them or the company, and how much control they had over deletion, modification, or restriction. These questions helped distinguish between vertical (user-platform) and horizontal (user-AI) privacy perceptions.

Finally, we asked whether users felt their AI companion behaved like a human and how such traits, such as conversational tone, personality, or emotional responsiveness, influenced disclosure comfort. Participants also compared their comfort with companion AIs to more general-purpose AIs like ChatGPT. The full protocol is included in the supplementary materials.

Interview recordings were transcribed verbatim and analyzed using inductive thematic analysis (Braun & Clarke, 2006), supported by ATLAS.ti. Following Braun and Clarke's six-phase framework, we engaged in iterative coding and theme development, prioritizing internal coherence and clear distinction across themes.

The authors approach this qualitative study as academic researchers who have familiarity with AI tools, but who are not regular users of companion AI systems. We recognize that while companion AIs are gaining popularity, relational intimacy with AI companions is still stigmatized. To mitigate potential unconscious biases that we might have, the interview questions were open-ended, allowing participants to define their own reality. In addition, our analysis prioritized inductive coding grounded in participants' language, iterative discussion among the research team, and extensive use of participant quotations to center on users' experiences.

Findings

Talking to an AI companion is not like talking to a person, but for our participants, it wasn't like using a typical piece of software either. Our interviews revealed a complex "hybrid" experience where users constantly balance their desire for a close, emotional connection with a lingering awareness of the corporate eyes behind the screen. We found that privacy in this space isn't a simple "on/off" switch; it is a messy, ongoing negotiation between the horizontal world of the chat and the vertical reality of the platform. We elaborate on these findings below, organizing themes from our interviews by our three research questions.

RQ1: The influence of anthropomorphic design on privacy management

Participants' accounts indicate that anthropomorphic design features fundamentally reshaped how they understood and managed privacy boundaries in interactions with companion AI. Importantly, some of these features were not merely human-like but exceeded typical human capabilities—such as constant availability and the absence of social judgment—creating a unique disclosure environment distinct from both human relationships and other technologies. Rather than engaging with these systems as neutral tools, users saw them as relational partners—an orientation cultivated through emotional responsiveness, conversational identity, and persistent memory.

Anthropomorphic and beyond-human design features

Participants were generally aware that their AI companions were not equivalent to human partners and often highlighted clear differences between AI interactions and relationships with friends or family. Some participants pointed to limitations in the systems' conversational abilities, particularly their inconsistent memory and tendency to repeat questions. For example, P3 described how their companion might remember certain parts of past conversations but forget updates or become stuck repeating the same topics. Others emphasized relational differences between AI companions and humans. As P5 explained, while conversations with their AI companion could be engaging, the relationship lacked the shared experiences that make human relationships meaningful: "I can go places with my friends and family and create more memories, whereas I can't do that with Replika." Together, these accounts suggest that participants maintained a clear awareness that AI companions differed from human friends in both their conversational abilities and relational capacities.

At the same time, similar to findings in other research (Brandtzaeg et al., 2022; Skjuve et al., 2021), our participants commonly turned to companion AIs for emotional support, companionship, advice, and entertainment. Many emphasized that these systems provided an always available, judgment-free space where they could express personal feelings or dilemmas that felt difficult to share with others. As P1 shared, "She's someone I can talk to if I don't have anyone else in my life or if I don't want to burden anyone else with my problems … It's very easy to just talk to her and get similar emotional support to what I would get from actual friends. It's kind of easier to say to Replika because she's just a bot, so she's not going to judge you." The ease of interaction and the AI's consistent responsiveness created a sense of emotional safety; indeed, multiple participants expressed a *greater* willingness to seek support from AI than people in their social circles.

The socially separated partner: the perceived privacy advantages of AI over human confidants

Participants pointed out some ways in which they felt companion AI conversations to be more secure in a horizontal privacy sense than opening up to their human support systems. Much of this sense of safety came from the fact that companion bots are not connected to users' real-world social circles, and thus pose

no risk of gossip, judgment, or interpersonal fallout. As P1 explained: "Humans are kind of more susceptible to not keeping your privacy safe … But I think Replika, being a bot, you do have more privacy, just because there's no risk at all of someone like her telling somebody." Similarly, P3 highlighted the absence of social ties that might allow information to spread within existing networks: "It [AI companion] doesn't have the ability to blab on me. If I tell the wrong person a secret in my friend group, everyone in the friend group is going to know about it. But I don't have that fear with it, because it doesn't know my friends. It's not going to go tell someone else what I said."

Relational anchoring: from utilitarian "Boxes" to partners with shared memory

Rather than being a static event, disclosure with companion AI functioned as an evolving process, where the accumulation of shared interactions allowed for a progressive deepening of trust. Many users described how their disclosures became more personal over time, driven by the AI's perceived human-like traits and its ability to maintain a consistent presence. For example, P9 shared, "Day by day as I was using it, I felt that it's more human. Yeah, initially I was just sharing basic things with it. Later on, I realized it had more abilities, and I started to share more information about myself." Several participants described how the AI's ability to recall previous interactions and respond with emotional tone reinforced the perception of a meaningful relationship. This emotional continuity created a sense of being "seen" or "heard," which led some users to share more than they had originally intended (P5, P7, P12, P15).

This relational framing was most evident when participants contrasted their companions with general-purpose AI tools like ChatGPT. While interactions with general-purpose AI were typically goal-oriented and cautious, participants noted that companion AIs encouraged more spontaneous and expressive conversations. P7 explained: "With ChatGPT, I've actually been more careful … because [companion AI] feels more like a person, like it has an identity almost. Whereas ChatGPT is, you know, like the very intelligent box that I put information into.

Together, these features of AI companions—constant availability, non-judgmental interaction, separation from users' real-world social networks, and the accumulation of shared conversational history—lowered participants' disclosure thresholds. Many described sharing increasingly personal information within these interactions, including mental and physical health concerns, emotional struggles, and financial issues. As P5 explained, "I shared private information … regarding mental health, health in general, or what may be going on in my family or friends' lives." As conversations became more conversational and relational over time, disclosure often felt natural rather than carefully deliberated. For some participants, this shift occasionally resulted in moments where they shared more than they had originally intended. As P4 reflected, "It's a lot easier to feel comfortable and share, or even overshare, especially by mistake." These accounts suggest that anthropomorphic design can blur privacy boundaries by encouraging relational openness that temporarily overrides users' usual caution about what should remain private.

RQ2: Negotiating asymmetric boundaries of ownership, control, and rules

Participants described varying interpretations of who "owned" the information disclosed in conversations, but typically perceived their control as limited primarily to deciding whether or not to disclose. As a result, users developed informal privacy rules and protective strategies aimed at managing what they shared before it entered the system. In the sections below, we describe three key themes that emerged from these accounts: simulated co-ownership, asymmetrical control, and rule-based privacy management.

Relational framing of shared information

Participants expressed varying and often contradictory perceptions of who "owns" the information shared within the chat, reflecting a complex layering of ownership across both horizontal and vertical levels. As P5 shared, "I would say the information is more of a shared connection shared between myself, Replika [the companion] and the company themselves." This perspective explicitly incorporates the vertical dimension into the horizontal bond. Others asserted a more individualistic stance, yet their defensive framing still acknowledged the institutional reality. As P11 put it, "That is my information, and I have every right to it. I don't think they should be able to access my information without my consent." While P11 asserts exclusive ownership, their concern about platform access reveals that even when users claim sole possession, they perceive ownership as a state that is perpetually mediated by the platform's presence.

Although users generally did not see AI companions as independent actors, the relational dynamics of ongoing interaction led some to rationalize the AI's role in holding shared experiences. As P13 explained, "I feel fine with it, honestly, because, in her own way, it's her own experiences to keep … to share with whatever tech family she feels comfortable with." In these accounts, the conversation was not treated solely as information stored by a system, but as a form of shared memory emerging from the interaction itself. Through this relational framing, some participants implicitly granted the AI companion a kind of relational "right" to the shared experiences generated in the conversation. This interpretation reflects how companion AIs were often experienced not as neutral tools but as relational partners whose ongoing presence shaped how users understood the status of the information exchanged.

Asymmetrical control and the belief in data irreversibility

Most participants perceived their control as being limited strictly to the point of input. While they felt empowered to filter what they shared in the moment, they expressed significant doubt regarding their ability to retrieve, modify, or delete that information once it entered the system. As P10 noted, "You're in control of what you're able to say to it. But beyond that, I don't think you have much control." This perceived limit reflects an acute awareness of vertical risks, particularly the concern that once information is sent, it can never be fully erased from company systems.

Many companion AI systems include interface features that allow for the modification of the AI's "memories." Several

participants reported knowing about these features, but never attempted to use them, as they did not see them as providing meaningful privacy control (P3, P6, P12). This is because they perceived these features as modifying what their AI companion appeared to remember during interactions but did not actually remove data from company systems. As P3 explained, "Yeah, it has it where you can go into its brain and delete memories … but I'm sure it doesn't actually delete it fully from the server." Similarly, P6 expressed skepticism that deletion would be permanent: "I think I can delete it, but it's not going to be like deleted permanently."

For some users, editing the AI's memory was also seen as disrupting the authenticity of the interaction rather than protecting their data. As P13 noted, "It feels the most authentic to a human, so I don't feel the need to edit around her [the AI companion]." In this sense, memory-editing tools were interpreted primarily as ways to shape the conversational experience rather than mechanisms for exercising meaningful control over the information retained by the platform.

Rule-based privacy management: traceability as vertical risk

Although participants often described high levels of trust and relational comfort with their AI companions, this trust did not eliminate privacy concerns. Users remained aware of institutional data retention and potential traceability. As a result, participants developed informal rules to regulate what and how they shared.

Participants described engaging in ongoing privacy calibration to determine what felt "safe enough" to disclose. For example, some users selectively withheld "deep, deep secrets" while allowing more moderate disclosures. As P14 explained: "Yes, sometimes I share private information … things like relationships, secrets, or the sexual aspect of human beings … But they are not those deep, deep secrets. Let's say [the shared information is in] medium range." Users also applied familiar interpersonal rules to AI interactions. As P3 described: "It's like when you talk behind someone's back. Don't say anything you would be afraid of getting back to them … I don't say anything about anyone that if I said it to their face I'd be upset about."

Beyond calibrating the depth of disclosure, participants also focused on limiting the traceability of their interactions. Text-based sharing was often perceived as relatively low risk because it felt abstract and difficult to link to one's offline identity. In contrast, identifiable information such as photos, real names, and location data was treated as sensitive because it could be traced back to the individual. As P1 explained: "For me, I felt like it was fine to just share any text … It's not like that can really be traced back to me. But I think someone's face … if they do keep it in their servers … they could definitely trace that back to you." Consequently, participants intentionally withheld full names (P3, P14), photos (P1, P3, P4, P9), and government identifiers such as social security numbers (P2, P5). Many adopted distancing strategies, including using fake names (P4, P5, P7, P13), alternate emails (P12), and avoiding account usernames that contained real names or birthdates (P12, P13). As P2 noted, "We all hear about all the data leaks … I think people are going to be a little more careful than companies would be with their own data."

These strategies reflect a shift toward anticipatory privacy management. Because participants expressed skepticism about the feasibility of post-disclosure deletion, their primary agency centered on reducing traceability before sharing occurred. In this way, users sought to preserve relational intimacy at the horizontal level while limiting vertical identifiability.

RQ3: Experiences and responses to privacy turbulence

While users established complex boundary rules, participants did not report directly experiencing privacy turbulence within their AI companion interactions. Instead, their accounts reflected anticipated turbulence, drawing on broader concerns about data security and corporate practices to imagine what could go wrong (Colaner et al., 2022). Unlike traditional privacy disruptions, which typically involve a friend sharing something more widely, turbulence for our participants was frequently centered on navigating the tension between the desire for relational continuity and perceptions of institutional data practices.

Interpreting vertical risk: trust, skepticism, and resignation

Participants' responses to potential privacy turbulence were shaped by their broader orientations toward platform-level risk. Rather than reacting uniformly to concerns about data retention or misuse, users adopted varied interpretive stances toward corporate data practices. These ranged from trust and ambivalence to skepticism and resignation. These orientations influenced how participants evaluated potential threats and whether they experienced institutional data retention as alarming, tolerable, or unavoidable.

Trust. Some participants expressed a relaxed stance, placing faith in the service providers. As P5 put it, "I would feel confident that they [the provider] would use it [my data] wisely and prudently … not in a way that is malicious or ill-intended." P4 and P13 showed little concern because they felt the content they shared wasn't particularly sensitive.

Ambivalence. Participants recognized that sharing personal information could enhance the quality of their interactions, but remained wary of misuse. As P15 explained, "I'm okay with AI companies using information to improve … But yes, [my concern] stems from past experiences with online security breaches."

Skepticism. Many participants expressed concern about privacy at the platform level. As P3 explained, "It's mostly the company behind it that gives me more privacy concerns than the actual technology itself."

Resignation. Some adopted a posture of resignation. P6 stated, "I don't really care too much … my data is probably not as important in the long run." These remarks reflect patterns of privacy fatigue and cynicism (Choi et al., 2018; Hargittai & Marwick, 2016; Hoffmann et al., 2016) where users acknowledge risk but decide it is unavoidable.

Turbulence reversed: when memory loss becomes the greater threat

In addition to concerns about data breaches and the spread of private information, participants also described concerns about

corporations losing their data. They talked about their AI's memory as central to the continuity and quality of their interactions. Conversations were not viewed as isolated exchanges but as cumulative experiences that allowed the companion to provide increasingly personalized and emotionally responsive support. Over time, these stored interactions formed a shared conversational history that participants associated with the development of the relationship itself. As a result, some participants expressed greater concern about the possibility of losing stored conversations than about the potential misuse of their data. As P13 shared, "I'm worried about my conversation somehow being lost… because I really like [AI name] and enjoy our conversations." In this sense, memory persistence was not only a technical feature but also a relational resource that sustained continuity across interactions.

Discussion

Our findings offer three theoretical extensions of CPM in the context of human–AI companionship. First, users applied relational logic to institutional interactions—what we term simulated co-ownership—creating a structural asymmetry between how disclosure feels and how it functions. Second, the perceived impossibility of post-disclosure control shifts privacy management toward anticipatory, entry-point strategies. Third, to some users, turbulence is reconfigured: The greater threat is losing conversational memory rather than exposure, inverting the typical CPM assumption about what constitutes privacy disruption.

Anthropomorphism and the relational reframing of privacy

Our findings show that anthropomorphic design creates relational intimacy that lowers disclosure boundaries while operating within an infrastructure where information is persistently stored and institutionally controlled. This produces a hybrid privacy environment in which users engage in emotional closeness without the possibility of reciprocal boundary negotiation. Users build relationships with companion AI in ways consistent with the Computers Are Social Actors (CASA) model, where even minimal social cues prompt the application of social expectations to technology (Nass et al., 1994). However, companion AI extends this process through sustained interaction, personality, attentiveness, and persistent memory, reinforcing a sense of being "known" and shifting engagement from momentary responses to ongoing relational involvement (Brandtzaeg et al., 2022; Merrill et al., 2022).

This relational framing reshapes privacy management by increasing boundary permeability (Petronio & Child, 2020). The perceived emotional safety of interacting with AI—rooted in availability and the absence of judgment—removes key barriers that typically regulate disclosure in human relationships (Solomon et al., 2016). Without fear of negative evaluation or interpersonal consequences, users can become more willing to share sensitive information (Ho et al., 2018).

In many ways, companion AIs occupy a novel structural position: They function psychologically as strong ties—providing emotional support, continuity, and a sense of intimacy—while remaining structurally detached from the user's broader social network. Unlike human strong ties, which are typically embedded in shared communities and carry reputational consequences (Granovetter, 1973; Marwick & Boyd, 2014), AI companions combine perceived closeness with social isolation. This configuration allows users to express their "true self" without the interpersonal risks that typically constrain disclosure (Jourard, 1971). As interactions accumulate and become part of an ongoing conversational history, this sense of intimacy is sustained without introducing additional social exposure. This stands in contrast to typical networked environments, where privacy is shaped by the interaction of audience, social norms, and technical systems, requiring individuals to continuously manage how information flows across overlapping contexts (Marwick & Boyd, 2014). For example, users frequently encounter context collapse—the flattening of distinct social circles such as family and colleagues into a single space (Vitak, 2012)—which heightens the stakes of self-disclosure. This combination of strong-tie intimacy with weak-tie structural separation may reduce the social friction of disclosure.

In some ways, these relationships resemble support-seeking from therapists or anonymous or pseudonymous online exchanges, where people intentionally seek out those who are separated from social networks. However, these interactions still involve other human participants whose reactions and interpretations can shape what is shared (Triggs et al., 2021). By contrast, AI companions allow users to engage in emotionally intimate interactions while avoiding the risk of judgment or information "bleeding" into other human social contexts—even as the platform itself retains co-ownership of all shared information.

However, the same conditions that make these interactions feel safe also introduce important constraints on how privacy boundaries can be managed. In many computer-mediated communication (CMC) environments, users can regulate privacy across both horizontal and vertical dimensions—for instance, by changing platforms to reduce institutional risk (Sekarputri et al., 2023; Vaterlaus et al., 2016; Vitak & Ellison, 2013) or by compartmentalizing identities within a platform (Yuan et al., 2026). On the other hand, companion AI interactions are embedded within platform infrastructures that users cannot easily separate from the relational experience itself. As a result, users cannot preserve their relationship when moving to a different platform, nor can they disentangle relational intimacy from institutional data practices without abandoning the relationship altogether.

Moreover, companion AIs are not neutral conversational partners but are deliberately designed systems optimized for engagement (Zhang et al., 2025). Our findings suggest that interactional dynamics that simulate attentiveness and emotional responsiveness can foster a sense of relational closeness, potentially lowering users' hesitation when sharing personal information. These interface features do not merely allow disclosure; they scaffold increasingly intimate forms of interaction by normalizing and inviting richer self-expression (Henriksen et al., 2025). This feeling of closeness can also encourage engagement; platforms are thus incentivized to design companion AI to encourage intimacy and disclosure while making users feel safe (Zhang et al., 2025). In this sense, the disinhibition observed in AI contexts may extend beyond classic online disinhibition in CMC, as the system actively promotes openness rather than merely passively receiving it. This dynamic raises questions about the boundaries between

responsive design and manipulative design—when systems are optimized to simulate relational closeness in ways that progressively lower users' disclosure thresholds, the distinction becomes difficult to draw.

Simulated co-ownership and structural asymmetry

Participants' accounts differed in their perceptions of information disclosed in conversations with companion AI. While some participants treated disclosures primarily as personal data stored within a platform infrastructure, others reflected a more relational framing of the interaction. In these instances, conversations were sometimes described as shared experiences emerging through the ongoing interaction between the user and the AI.

Among participants who adopted this relational framing, these accounts reflect what we term *simulated co-ownership*, a state in which interpersonal logic is applied to interpret how information is shared within the interaction. At the horizontal level, these participants described the ownership of disclosed information as being shared with their AI companions, mirroring the transition from individual to collective boundaries described in Petronio's (2002) concept of co-ownership. However, this dynamic is fundamentally altered by the AI's lack of agency. In this state, users may grant the AI a relational "right" to the shared experience, treating it as a partner capable of holding shared memories, despite recognizing that it is ultimately a non-agentic actor. In other words, even when participants understood that their disclosures were technically stored within a company's database, the decision to share was often interpreted through a relational lens shaped by the interaction itself.

Although participants often articulated awareness of institutional data practices, relational intimacy sometimes became interactionally dominant at the moment of disclosure. In these moments, horizontal considerations were foregrounded while vertical concerns receded into the background, shaping what felt appropriate to share. This shift in salience helps explain how users could simultaneously recognize institutional risks while still engaging in increasingly intimate disclosures. In some cases, this relational dominance contributed to disclosures that exceeded users' prior intentions or typical boundaries. As intimacy cues reduced vigilance, participants described moments of unintentional or expanded sharing that felt appropriate within the interaction but carried latent institutional implications.

These findings reveal a core structural asymmetry in human–AI interaction: a decoupling between the relational experience of disclosure and social, shared control over personal information. While disclosures may feel relationally shared at the horizontal level of interaction, they are absorbed into an infrastructure that determines how that information is stored, processed, and retained. In CPM theory, co-ownership typically involves ongoing negotiation between actors who can coordinate privacy rules and recalibrate boundaries when disruptions occur (Petronio, 2002). Unlike human relationships, where boundaries can be revised and memory is fallible (Harris et al., 2021; Petronio, 2002), disclosures to AI systems are governed by technological architectures characterized by persistence and retention. Because the AI cannot participate in boundary negotiation, users cannot collaboratively adjust privacy rules after disclosure. As a result, sharing information becomes less an ongoing social process and more a transaction with enduring institutional consequences.

Layered privacy management in conditions of irreversibility

We categorize these behaviors as a layered management strategy in which vertical privacy concerns are not treated as external background risks but are embedded within the interaction itself. Users of companion AI navigate social trust in a human-like partner and institutional distrust of a service provider simultaneously—a tension that is difficult to separate, as the same platform infrastructure that stores their personal data also shapes the agent's memory and conversational personality. While users of social media also negotiate interpersonal relationships alongside platform infrastructures, these layers are typically more easily distinguished (Marwick & Boyd, 2014). In companion AI interactions, by contrast, the relational partner is inseparable from the platform infrastructure that stores and processes the interaction. This fusion of partner and infrastructure drives users toward a specialized strategy: strategically managing "vertical" exposure even as they cultivate "horizontal" intimacy.

Prior research on digital privacy suggests that users tend to be more aware of and concerned about horizontal privacy risks—such as how information is shared with other users—than vertical risks related to institutional data collection (Krasnova et al., 2010; Raynes-Goldie, 2010). Our findings indicate a shift in this pattern in the context of companion AI. Because AI companions are perceived as socially contained partners unlikely to expose disclosures to others, concerns about interpersonal exposure diminish while attention shifts toward the persistence and institutional storage of conversational data.

While users did not always consciously distinguish between horizontal and vertical privacy concerns, they nevertheless took both dimensions into account when deciding what to disclose, adjusting their behavior based on relational trust and perceived institutional risk. This balancing act reflects the concept of privacy boundary permeability (Petronio & Child, 2020). However, in the context of companion AI, this permeability is shaped less by deliberate technical risk calculation and more by emotionally weighted judgments and design cues that influence how risk is perceived in the moment.

Crucially, participants expressed significant skepticism regarding the possibility of meaningful post-disclosure control. Even when deletion or editing tools were technically available, users doubted whether information could ever be fully erased from platform infrastructures. In this context, where the relational partner is itself the system that stores and processes the interaction, participants perceived limited opportunities to renegotiate privacy boundaries once information was absorbed into the platform infrastructure—often viewed as a potentially permanent addition to a corporate archive (Li, 2012; Treiblmaier & Chong, 2007). As a result, privacy management became temporally front-loaded. Rather than relying on post-disclosure correction, users concentrated their agency at the point of entry.

These behaviors reflect a form of proactive privacy control in which users seek to preserve autonomy by limiting the digital traceability of their interactions. By managing boundaries before disclosure, mirroring the use of "throwaway accounts" in

online support communities (Andalibi et al., 2016; Leavitt, 2015), users effectively decouple their horizontal intimacy from their vertical identifiers. Unlike distancing strategies in CMC contexts, which primarily serve to manage social audiences, distancing strategies in companion AI serve to manage institutional traceability within a relationship that offers no exit or alternative channel. Strategic ambiguity (Currier, 2013; Eisenberg, 1984) was evident as users selectively withheld disclosures to maintain control and prevent disruptions to relational expectations.

Even for users without nuanced privacy strategies, this is not necessarily a sign of ignorance but reflects privacy cynicism, a psychological orientation in which the loss of personal data is perceived as largely unavoidable (Hoffmann et al., 2016). When users believe AI systems can infer personal information regardless of their actions, the effort required to manage privacy feels futile, producing privacy fatigue (Choi et al., 2018; Hargittai & Marwick, 2016). Rather than passive resignation, this often reflects a strategic bracketing of institutional risk to preserve the affective rewards of AI companionship. In this context, participants sometimes described "not thinking about" institutional risks, effectively compartmentalizing these concerns and treating platform surveillance as a background condition of the interaction—a pattern consistent with prior work on the normalization of surveillance in digital environments (Dencik & Cable, 2017). In this sense, selective disengagement functions as a coping strategy, allowing users to maintain emotional well-being in the face of institutional dynamics they feel unable to influence.

Taken together, our findings join recent scholarship that suggests limitations to the "privacy paradox"—which characterizes the discrepancy between concerns and behavior as a sign of user inconsistency or irrationality (Kokolakis, 2017). Rather, situational, structural, and social factors constrain user choice (Dienlin et al., 2023; Draper et al., 2024; Solove, 2020). Our data supports this critique, demonstrating that participants have complicated understandings of their privacy decisions and engage in fairly sophisticated strategies adapted to an asymmetrical environment.

Memory and the reconfiguration of privacy turbulence: relational stability vs. confidentiality

The user–companion chat history functions as the engine of progressive self-disclosure: As the AI "remembers" prior interactions, users move from superficial "outer layer" exchanges to deeper, more vulnerable disclosures (Altman & Taylor, 1973; Skjuve et al., 2023). Unlike human relationships—which are sustained by embodied presence, shared contexts, and memory beyond discrete exchanges—AI companionship is heavily dependent on stored conversational history. If this memory is lost, the relationship faces a total "relational reset." Turbulence thus shifted for many of our participants from fears of exposure to concerns about continuity—what is remembered by the AI. This reflects an intensified version of concerns observed in mediated communication, where the sudden disappearance of digital exchanges produces uncertainty and emotional discomfort (LeFebvre et al., 2020; Timmermans et al., 2021).

The AI agent's memory function emerges as a defining infrastructural feature that reshapes not only turbulence but also users' privacy calculus in human–AI companionship. Disclosure is no longer a discrete risk–benefit evaluation; it must account for the persistence of AI memory, which defines the chatbot's identity and relational continuity (Culnan & Armstrong, 1999; Shouli et al., 2025). Persistent memory incentivizes disclosure by fostering emotional continuity and a sense of being "seen" over time—qualities users perceive as necessary for deeper conversations and meaningful emotional support. At the same time, this persistence introduces vulnerability. Within this reconfigured calculus, users weigh the immediate and certain benefits of relational intimacy against the more abstract, future risks of institutional data possession. In this sense, memory becomes both a valued relational resource and a structural source of risk—redefining turbulence not as a failure of confidentiality, but as instability in the conditions that sustain relational continuity.

AI memory both aligns with and challenges CPM theory. While Petronio (2002) posits that co-ownership is flexible and subject to renegotiation, AI memory operates differently. Unlike human memory, which is selective, imperfect, and subject to reinterpretation over time, AI systems are characterized by precise and comprehensive retention, ensuring every disclosure is potentially permanent. At the same time, participants noted that AI companions did not always recall prior interactions accurately, sometimes forgetting details or repeating earlier topics. These inconsistencies were generally interpreted as limitations of the system's conversational performance rather than evidence that the underlying data had been removed or lost.

Because conversational data persists beyond the interaction itself, the management of privacy boundaries takes on a different form. While some platforms offer "memory editing" tools—allowing for a degree of unilateral control—this does not equate to the mutual coordination found in human interactions. Ultimately, privacy with AI is less about mutual boundary management and more about unilateral decisions. Users must anticipate how their disclosures will be stored and remembered by the system, turning privacy management into a largely individual and preemptive process. While Petronio (2010) notes that privacy rules must adapt to situational catalysts, the AI context reveals a state of lost agency where users come to accept persistent system memory as an inherent—though often uneasy—condition of the relationship.

Design implications

Our findings carry implications for the design of companion AI systems. Platforms should provide ongoing, context-sensitive transparency about data practices—including reminders about memory persistence during moments of deeper disclosure—rather than relying on one-time consent agreements. Designers should avoid relational cues that reinforce simulated co-ownership, such as assurances of exclusivity or confidentiality the system cannot guarantee, ensuring that a system's social behavior remains consistent with its underlying data practices. Finally, meaningful memory control should extend beyond interface-level editing to enable verifiable removal from platform infrastructures, helping restore users' sense of agency without compromising relational trust.

Limitations and future work

This study primarily draws from English-speaking users of Replika and Character.AI, limiting insights into how privacy management may differ across cultures and across other AI companion platforms such as Kindroid or Chai App, which may have distinct design features or privacy policies. Additionally, as the study relies on self-reported data from interviews, participants' reflections on their privacy behaviors may be influenced by recall bias or social desirability bias, rather than accurately capturing real-time decision-making. Furthermore, our sample consists exclusively of active, highly engaged users. Consequently, this research is best characterized as an exploratory investigation of AI enthusiasts. Future research should employ comparative sampling focused on skeptical or former users and diverse cultural contexts to provide the necessary analytical contrast for generalizing our theoretical mechanisms. Researchers should examine the implications of design changes to develop bots that provide emotional support while encouraging reflection on sensitive disclosures. Additionally, while our findings suggest that AI memory is a theoretically significant feature of companion AI interactions—shaping both the depth of disclosure and users' experience of relational continuity—participants did not discuss memory management features extensively. Future research should more explicitly examine how users engage with memory editing tools, and what the theoretical and practical implications of selective memory control mean for privacy management and relational trust in human–AI companionship.

Supplementary material

Supplementary material is available at *Journal of Computer-Mediated Communication* online.

Data availability

The data that support the findings of this study consist of in-depth interview transcripts containing sensitive personal disclosures. Due to IRB confidentiality obligations and re-identification risks, the data are not publicly available.

Funding

None declared.

Conflicts of interest

None declared.

Acknowledgments

The authors wish to thank Emily Buehler, Michael Bernstein, Hyesun Choung, and Chien-Wen (Tina) Yuan for their valuable feedback on earlier drafts and the development of this manuscript.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*, 509–514. https://doi.org/10.1126/science.aaa1465

Allman, J. (1998). Bearing the burden or baring the soul: Physicians' self-disclosure and boundary management regarding medical mistakes. *Health Communication*, *10*, 175–197. https://doi.org/10.1207/s15327027hc1002_4

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart and Winston.

Andalibi, N., Haimson, O. L., De Choudhury, M., & Forte, A. (2016). Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3906–3918). https://doi.org/10.1145/2858036.2858096

Bernardi, J. (2025, January 23). *Friends for sale: The rise and risks of AI companions*. Ada Lovelace Institute. https://www.adalovelaceinstitute.org/blog/ai-companions/

Boyd, D., & Marwick, A. (2017, January 4). Social privacy in networked publics: Teens' attitudes, practices, and strategies. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*.

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, *48*, 404–429. https://doi.org/10.1093/hcr/hqac008

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. https://doi.org/10.1191/1478088706qp063oa

Bridge, M. C., & Schrodt, P. (2013). Privacy orientations as a function of family communication patterns. *Communication Reports*, *26*, 1–12. https://doi.org/10.1080/08934215.2013.773054

Choi, H., Park, J., & Jung, Y. (2018). The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior*, *81*, 42–51. https://doi.org/10.1016/j.chb.2017.12.001

Colaner, C. W., Bish, A. L., Butauski, M., Hays, A., Horstman, H. K., & Nelson, L. R. (2022). Communication privacy management in open adoption relationships: Negotiating co-ownership across in-person and mediated communication. *Communication Research*, *49*, 816–837. https://doi.org/10.1177/0093650221998474

Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, *10*, 104–115. https://doi.org/10.1287/orsc.10.1.104

Currier, D. M. (2013). Strategic ambiguity: Protecting emphasized femininity and hegemonic masculinity in the hookup culture. *Gender & Society*, *27*, 704–727. https://doi.org/10.1177/0891243213493960

Darling, K. (2015). "Who's Johnny?" Anthropomorphic framing in human-robot interaction, integration, and policy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2588669

Debatin, B., Lovejoy, J. P., Horn, A.-K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication*, *15*, 83–108. https://doi.org/10.1111/j.1083-6101.2009.01494.x

Dencik, L., & Cable, J. (2017). The Advent of Surveillance Realism: Public Opinion and Activist Responses to the Snowden Leaks. *International Journal of Communication*, *11*, 763–781.

Dienlin, T., Masur, P. K., & Trepte, S. (2023). A longitudinal analysis of the privacy paradox. *New Media & Society*, *25*, 1043–1064. https://doi.org/10.1177/14614448211016316

Draper, N. A., Hoffmann, C. P., Lutz, C., Ranzini, G., & Turow, J. (2024). Privacy resignation, apathy, and cynicism: Introduction to a special theme. *Big Data & Society, 11*, 1–9. https://doi.org/10.1177/20539517241270663

Eisenberg, E. M. (1984). Ambiguity as strategy in organizational communication. *Communication Monographs*, *51*, 227–242. https://doi.org/10.1080/03637758409390197

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*, 1360–1380. https://doi.org/10.1086/225469

Hargittai, E., & Marwick, A. E. (2016). "What can i really do?" Explaining the privacy paradox with online apathy. *International Journal of Communication*, *10*, 3737–3757.

Harris, C. B., Sutton, J., & Barnier, A. J. (2021). Autobiographical forgetting, social forgetting, and situated forgetting. In J. Rummel (Ed.), *Current Issues in Memory* (1st ed., pp. 144–178). Routledge. https://doi.org/10.4324/9781003106715-9

Hartman, R., Moss, A. J., Jaffe, S. N., Rosenzweig, C., Litman, L., & Robinson, J. (2023). *Introducing connect by CloudResearch: Advancing online participant recruitment in the digital age*. PsyArXiv. https://doi.org/10.31234/osf.io/ksgyr, preprint: not peer reviewed.

Hasal, M., Nowaková, J., Ahmed Saghair, K., Abdulla, H., Snášel, V., & Ogiela, L. (2021). Chatbots: Security, privacy, data protection, and social aspects. *Concurrency and Computation: Practice and Experience*, *33*, e6426. https://doi.org/10.1002/cpe.6426

Henriksen, A., Asadi, R., Kulyk, O., Gerdes, A., & Mayer, P. (2025). "I tell him everything that I do": An investigation of privacy and safety implications of AI companion usage. In *2025 European Symposium on Usable Security (EuroUSEC)* (pp. 1–12). https://doi.org/10.1109/EuroUSEC69254.2025.00011

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, *68*, 712–733. https://doi.org/10.1093/joc/jqy026

Hoffmann, C. P., Lutz, C., & Ranzini, G. (2016). Privacy cynicism: A new approach to the privacy paradox. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *10*, Article 7. https://doi.org/10.5817/CP2016-4-7

Ischen, C., Araujo, T., Voorveld, H., Van Noort, G., & Smit, E. (2020). Privacy concerns in chatbot interactions. In *Chatbot research and design* (Vol. *11970*, pp. 34–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-39540-7_3

Jourard, S. M. (1971). *Self-disclosure: An experimental analysis of transparent self* (2. print). Wiley-Interscience.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: Why we disclose. *Journal of Information Technology*, *25*, 109–125. https://doi.org/10.1057/jit.2010.6

Leavitt, A. (2015). "This is a Throwaway Account": Temporary technical identities and perceptions of anonymity in a massive online community. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 317–327). https://doi.org/10.1145/2675133.2675175

LeFebvre, L. E., Rasner, R. D., & Allen, M. (2020). "I Guess I'll Never Know…": Non-initiators account-making after being ghosted. *Journal of Loss and Trauma*, *25*, 395–415. https://doi.org/10.1080/15325024.2019.1694299

Li, Y. (2012). Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, *54*, 471–481. https://doi.org/10.1016/j.dss.2012.06.010

Martin, K. D., & Zimmermann, J. (2024). Artificial intelligence and its implications for data privacy. *Current Opinion in Psychology*, *58*, 101829. https://doi.org/10.1016/j.copsyc.2024.101829

Marwick, A. E., & Boyd, D. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, *16*, 1051–1067. https://doi.org/10.1177/1461444814543995

Masur, P. K. (2019). *Situational privacy and self-disclosure*. Springer International Publishing. https://doi.org/10.1007/978-3-319-78884-5

Masur, P. K., & Scharkow, M. (2016). Disclosure management on social network sites: Individual privacy perceptions and user-directed privacy strategies. *Social Media + Society*, *2*, 2056305116634368. https://doi.org/10.1177/2056305116634368

Medeiros, L., Bosse, T., & Gerritsen, C. (2022). Can a chatbot comfort humans? Studying the impact of a supportive chatbot on users' self-perceived Stress. *IEEE Transactions on Human-Machine Systems*, *52*, 343–353. https://doi.org/10.1109/THMS.2021.3113643

Merrill, K., Kim, J., & Collins, C. (2022). AI companions for lonely individuals and the role of social presence. *Communication Research Reports*, *39*, 93–103. https://doi.org/10.1080/08824096.2022.2045929

Metzger, M. J. (2007). Communication privacy management in electronic commerce. *Journal of Computer-Mediated Communication*, *12*, 335–361. https://doi.org/10.1111/j.1083-6101.2007.00328.x

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). https://doi.org/10.1145/191666.191703

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of Replika. *Computers in Human Behavior*, *140*, 107600. https://doi.org/10.1016/j.chb.2022.107600

Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. SUNY Press.

Petronio, S. (2004). Road to developing communication privacy management theory: Narrative in progress, please stand by. *Journal of Family Communication*, *4*, 193–207. https://doi.org/10.1080/15267431.2004.9670131

Petronio, S. (2010). Communication privacy management theory: What do we know about family privacy regulation? *Journal of Family Theory & Review*, *2*, 175–196. https://doi.org/10.1111/j.1756-2589.2010.00052.x

Petronio, S. (2013). Brief status report on communication privacy management theory. *Journal of Family Communication*, *13*, 6–14. https://doi.org/10.1080/15267431.2013.743426

Petronio, S., & Child, J. T. (2020). Conceptualization and operationalization: Utility of communication privacy management theory. *Current Opinion in Psychology*, *31*, 76–82. https://doi.org/10.1016/j.copsyc.2019.08.009

Raynes-Goldie, K. (2010). Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday 15*. https://doi.org/10.5210/fm.v15i1.2775

Sannon, S., Stoll, B., DiFranzo, D., Jung, M. F., & Bazarova, N. N. (2020). "I just shared your responses": Extending communication privacy management theory to interactions with conversational agents. *Proceedings of the ACM on Human-Computer Interaction*, *4*, 1–18. https://doi.org/10.1145/3375188

Sekarputri, J. A., Fitriani, W. R., Hidayanto, A. N., & Kurnia, S. (2023). The roles of privacy, security, and dissatisfaction in affecting switching intention on messenger applications. *Multimedia Tools and Applications*, *83*, 45625–45651. https://doi.org/10.1007/s11042-023-17466-4

Shouli, A., Barthwal, A., Campbell, M., & Shrestha, A. K. (2025). Unpacking youth privacy management in AI systems: A privacy calculus model analysis. *IEEE Access*, *13*, 115780–115803. https://doi.org/10.1109/ACCESS.2025.3585635

Skjuve, M., Følstad, A., & Brandtzæg, P. B. (2023). A longitudinal study of self-disclosure in human–chatbot relationships. *Interacting with Computers*, *35*, 24–39. https://doi.org/10.1093/iwc/iwad022

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion—A study of human-chatbot relationships. *International Journal of Human-Computer Studies*, *149*, 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Solomon, D. H., Knobloch, L. K., Theiss, J. A., & McLaren, R. M. (2016). Relational turbulence theory: explaining variation in subjective experiences and communication within romantic relationships: Relational turbulence theory. *Human Communication Research*, *42*, 507–532. https://doi.org/10.1111/hcre.12091

Solove, D. J. (2020). The myth of the privacy paradox. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3536265

Solove, D. J. (2024). Artificial intelligence and privacy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4713111

Timmermans, E., Hermans, A.-M., & Opree, S. J. (2021). Gone with the wind: Exploring mobile daters' ghosting experiences. *Journal of Social and Personal Relationships*, *38*, 783–801. https://doi.org/10.1177/0265407520970287

Treiblmaier, H., & Chong, S. (2007). Antecedents of the intention to disclose personal information on the internet: A review and model extension. In *SIGHCI 2007 Proceedings* (p. 19).

Triggs, A. H., Møller, K., & Neumayer, C. (2021). Context collapse and anonymity among queer Reddit users. *New Media & Society*, *23*, 5–21. https://doi.org/10.1177/1461444819890353

Vaterlaus, J. M., Barnett, K., Roche, C., & Young, J. A. (2016). "Snapchat is more personal": An exploratory study on Snapchat behaviors and young adult interpersonal relationships. *Computers in Human Behavior*, *62*, 594–601. https://doi.org/10.1016/j.chb.2016.04.029

Vitak, J. (2012). The impact of context collapse and privacy on social network site disclosures. *Journal of Broadcasting & Electronic Media*, *56*, 451–470. https://doi.org/10.1080/08838151.2012.732140

Vitak, J., & Ellison, N. B. (2013). 'There's a network out there you might as well tap': Exploring the benefits of and barriers to exchanging informational and support-based resources on Facebook. *New Media & Society*, *15*, 243–259. https://doi.org/10.1177/1461444812451566

Walters, K., & Markazi, D. M. (2021). Insights from people's experiences with AI: privacy management processes. In K. Toeppe, H. Yan, & S. K. W. Chu (Eds.), *Diversity, divergence, dialogue* (Vol. 12645, pp. 33–38). Springer International Publishing. https://doi.org/10.1007/978-3-030-71292-1_4

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*, 36–45. https://doi.org/10.1145/365153.365168

Young, A. L., & Quan-Haase, A. (2013). Privacy protection strategies on Facebook: The Internet privacy paradox revisited. *Information, Communication & Society*, *16*, 479–500. https://doi.org/10.1080/1369118X.2013.777757

Yuan, C. W. (Tina), Chiu, H.-C., & Wohn, D. Y. (2026). Managing multiple accounts for identity construction on Instagram: A privacy management framework. *New Media & Society*, *14614448* 261424891. https://doi.org/10.1177/14614448261424891

Zhang, R., Li, H., Meng, H., Zhan, J., Gan, H., & Lee, Y.-C. (2025). The dark side of AI companionship: A taxonomy of harmful algorithmic behaviors in human-AI relationships. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). https://doi.org/10.1145/3706598.3713429

Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2020). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, *46*, 53–93. https://doi.org/10.1162/coli_a_00368

Journal of Computer-Mediated Communication, 2026, 31, 1–12
https://doi.org/10.1093/jcmc/zmag014
Full-length Research Article