Mark Bo Chen. "I'm a bit cautious of jumping in with both feet": exploring information ownership and negotiated control in AI chatbot users' communication privacy management.

School of Culture and Communication, The University of Melbourne.
Email: markchan0814@gmail.com

Abstract

Advances in artificial intelligence have garnered significant attention, with user privacy emerging as a focal point. Guided by a privacy management perspective, this exploratory study investigates how users make sense of informational privacy when interacting with their AI chatbot counterparts, drawing from Reddit data (submissions, n=193) that represent unsolicited user vignettes of chatbot-related privacy experiences. Situated in Human-Machine Communication (HMC), the study applies Communication Privacy Management (CPM) theory to analyse how information ownership and control are understood and negotiated as part and parcel of privacy management strategies in user-chatbot communication. Findings reveal users' struggle to grapple with boundary regulations in automated systems; their situational strategies of boundary making are shaped not only by users' disclosure intention and privacy concerns, but also the techno-social features of chatbots that limit the extent to which users' tactics of privacy management are practised. With a user-centric approach, this study extends CPM to HMC and contributes to our understanding of how ordinary users perceive and negotiate informational privacy in the context of everyday AI use. Theoretical and practical implications are discussed.

Keywords

Chatbot; Privacy; Artificial Intelligence; Communication Privacy Management; Human-Machine Communication; ChatGPT

Introduction

Recent years have witnessed an explosion in artificial intelligence (AI) technologies, including chatbots powered by large language models. Broadly, AI are complex techno-social assemblages (Eynon and Young, 2021), constructed through social processes that encapsulate not only the technicality, but also the knowledge, practices and negotiation in handling these systems (Guzman and Lewis, 2019). In everyday life, how users engage with AI technologies is fundamentally grounded in communication practices as relational collaborations (e.g., using natural language to communicate with chatbots) (Gunkel, 2012; Guzman, 2018). On the other hand, communication privacy is relevant to nearly all human activities (Altman, 1975; Petronio, 2002), and poses challenges in the context of AI use, particularly due to the opacity of algorithmic systems and the dynamic ways in which user data can be inferred, stored and repurposed beyond the original context (Gorwa and Veale, 2024; Lutz et al., 2019). Therefore, it is not surprising that while popular AI chatbots like ChatGPT are widely embraced in daily lives (e.g., Westfall, 2023), sentiments of uncertainty prevail, with one of the heated topics being loss of control on data and informational privacy (e.g., Sher and Benchlouch, 2023). As the hype around AI continues, communication research is required to understand, beyond the current hyperbole surrounding technological progressions, how ordinary people make sense of AI and manage privacy when they communicate directly with these machines.

AI chatbots, as epitomised by OpenAI's ChatGPT, are a type of narrow AI. Narrow AI is designed to perform a particular task, and in this sense, is seen as having limited capacity. Chatbots can extract information from user inputs and create outputs sensitive to the inputs and comprehensible to humans (Allen, 2003). Their functionalities rely on datafication (Hepp, 2020), that is, the collection and processing of large amounts of data to learn relationships between words and remember conversations and contextual dependencies to personalise responses to users. Personalisation sustains various utilitarian and social needs that motivate users to interact with AI chatbots (Brandtzaeg et al., 2022; Skjuve et al., 2024). Given the vast amounts of user data involved, these data-driven benefits can also lead to anxieties around what data are collected, how the data are processed, with whom the data are shared, and what measures are in place to protect user privacy.

Empirical research on user privacy and AI chatbots remains limited, with much literature (Ischen et al., 2019; Lim and Shim, 2022; Liu et al., 2023; Sannon et al., 2020) relying on experimental designs to measure privacy intentions in isolated environments and as numeric metrics. These designs risk priming participants to inflate their privacy concerns and overlook the relational and negotiated nature of communication privacy management (Palen and Dourish, 2003; Petronio, 2002). Furthermore, while some studies (Ischen et al., 2019; Lim and Shim, 2022) suggest that anthropomorphic design can reduce privacy fears, other perspectives (Liu et al., 2023; Sannon et al., 2020; Sundar and Kim, 2019) highlight persistent tensions in how users trust and manage information with chatbots. This underscores the need for deeper investigation into how communication privacy is understood and negotiated when users interact with conversational AI systems in the wild.

This exploratory study elucidates how users articulate their privacy experiences in everyday interactions with their chatbot counterparts, based on a Reddit-sourced dataset (n=193) from five sub-reddit forums (*r/ChatGPT*, *r/ClaudeAI*, *r/perplexity_ai*, *r/GeminiAI, r/CharacterAI*). Using Communalytic[1], submissions were collected in two phases, screened for relevance and then analysed thematically. In doing so, the present study moves beyond laboratory settings and evaluates how ordinary chatbot users understand information ownership and negotiated control, two key facets of privacy management. Findings were contextualised in the domain of Human-Machine Communication (HMC; Guzman, 2018; Guzman and Lewis, 2020), and interpreted using the Communication Privacy Management theory (CPM; Petronio, 2002) which considers how individuals develop rules to manage information disclosure, and (re-)negotiate these rules when boundary turbulence arises in episodes of privacy breakdown. This study extends CPM, a theory traditionally applied in interpersonal communication, to HMC, arguing that communication privacy behaviours are results of situational negotiations between users and chatbots, shaped by both technical affordances and interactional dynamics. The sections that follow begin with a review of relevant literatures and detail the methodological approach and data sources. Then, key findings are presented, followed by a discussion of their implications, a reflection on limitations, and an outline for future research.

Literature Review

Privacy Management and Communication Privacy Management Theory

Contemporary privacy scholarships draw on inspirations from diverse domains including sociology, psychology and law (Altman, 1975; Westin, 1967). Different perspectives have produced numerous tomes of insightful research but also complicate a universally applicable understanding of privacy (Solove, 2006). In communication research, a widely adopted definition comes from Westin's work (Lutz, 2023) where privacy is conceptualised as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin, 1967, p. 7). Implicit to this definition is the informational dimension of privacy, which frames privacy as a matter of information management. While other dimensions of privacy are crucial to discussions on AI technologies more broadly (see Lutz et al., 2019), this research focuses on AI chatbots and echoes Lutz's argument (2023) that privacy implications of user-chatbot interactions primarily concerns the exchange of information. This can range from metadata (e.g., IP address, timestamps) to interactional content (e.g., chat logs, uploaded documents), as part of accessing and using chatbot services.

Digital technologies mediate not just information flow, but also emotional and affective relations (e.g., Bucher, 2017). This contributes to rendering boundaries between human and technology increasingly ambiguous (Turkle, 2005), giving rise to emerging forms of human-technology intimacy (Li and Zhang, 2024) and privacy implications (Lim and Shim, 2022). For AI chatbot users, privacy concerns may be sourced from a perceived loss of control over private information. Simultaneously, utilitarian and social benefits—such as productivity (Skjuve et al., 2024), personalisation and social connectedness (Brandtzaeg et al., 2022)—motivate continued use. As some degree of disclosure is required to use technology (Palen and Dourish, 2003), users face a tension between privacy fears (pushing factors) and the benefits (pulling factors). In this light, privacy in a human-chatbot dyad is not simply about a dichotomy between disclosure and concealment, but rather the selective control of access to personal information (Altman, 1975) and "the continual management of boundaries between different spheres of action and degrees of disclosure within those spheres" (Palen and Dourish, 2003, p. 131). Given the push-and-pull dynamics as described, it can be argued that the management of personal information flow and varied degrees of disclosure undergird individual users' privacy management practices in user-chatbot interactions.

To govern information flow, the tension between various pulling and pushing forces need to be mitigated. Petronio's Communication Privacy Management Theory (2002) provides a framework to make sense of such dialectical tension between privacy and disclosure. As a rule-based system, the theory posits that there are both risks and benefits to disclosure, and as such, individuals in dyadic relationships erect communication boundaries and establish privacy management rules for the disclosure and protection of privacy information, based on the belief that they are the owner of such information. According to CPM, these rules emerge from the "dialectical tension between openness and closedness" (Child et al., 2009, p. 2082), and are aimed at striking a balance between solitude and sociality in relational contexts. At its core, CPM rejects dichotomous thinking and recognises that disclosure and control of information are distinct user privacy management tactics, which has been extended to different technology-mediated environments including online blogging (e.g., Child et al., 2009), social media (e.g., Kang et al., 2022), e-commerce (e.g., Metzger, 2007) and smart technologies (e.g., Vitak et al., 2023). Therefore, although initially developed in

the domain of interpersonal communication, these existing cases showcase CPM's versatility and applicability in analysing technology- and privacy-related issues.

This research is inspired by CPM key principles to move beyond treating privacy as mere disclosure-withdrawal juxtaposition. It explores privacy practices as negotiated efforts of boundary management in everyday user-chatbot interactions. The next section builds on existing applications of CPM in technology-mediated communication and examines the theory's relevance to HMC. It then contextualises CPM within HMC's key focus on direct user engagement with communicative machines like chatbots.

Communication Privacy Management in User-chatbot Communication

CPM has informed various recent studies on digital technology and privacy (e.g., Child et al., 2009; Kang et al., 2022; Metzger, 2007). However, most of these cases are grounded in the computer-mediated communication (CMC) paradigm; as Lutz (2023) contends, a CMC perspective places its investigative locus on privacy relations either between individual users, or between the user and other stakeholders in the digital network (e.g., digital service providers). In contrast, HMC views machines as social actors that users communicate directly with, instead of as a mediator (Gunkel, 2012; Guzman, 2018). This perspective entails that user-chatbot communication poses different privacy implications from those explored in CMC studies, as it involves direct interactions with an autonomous system that functions as a conversational partner and a data collection interface.

Andrea L. Guzman (2018, p. 17) defines HMC as the "creation of meaning among humans and machines". Communication with machines as meaning-making endeavours echoes earlier scholarships (Gunkel, 2012; Reeves and Nass, 1996; Turkle, 2005) that interacting with human-like technologies is indeed a collaborative matter unfolding in situational communication contexts. Text-based communicative modalities are the primary interactive functions of AI chatbots, with whom users communicate directly through an interface using natural language (Hepp, 2020). To this extent, communication between the human user and the chatbot mimics that of interpersonal communication, as both parties occupy their legitimate spots in a two-way communication structure (Gunkel, 2012). This relational perspective inherent to HMC thinking acknowledges both the human user's active role in making sense of the technological other, and the machine's role in shaping the user's communication practices. CPM is premised on a relational view of privacy management as negotiated decisions and continual assessment of communication boundaries between partners (Petronio, 2002). The negotiated nature of privacy proposed by CPM suggests that privacy management strategies and rules to govern boundaries between closedness and openness are results of situational two-way collaborations that define these strategies and rules. This conceptual alignment between CPM and HMC, reinforces CPM's relevance to understanding how users develop and adapt privacy rules when interacting with relational machines.

Recent theoretical explorations (Spence, 2019) have proposed that human-human communication theories can offer productive jumping-off points to understand communication between human and machine. However, such a pragmatic approach is not without its risks (Guzman and Lewis, 2020); machine as a communicator is not the same as its human counterpart, as they lack clear social cues and contextual awareness. Furthermore, AI chatbots are complex automated systems of communication involving different techno-social layers (Hepp, 2020). To communicate with a

chatbot, users need to conceptualise a source which communication hinges upon (Guzman, 2019; cf. Reeves and Nass, 1996). In HMC it is not always straightforward what information sources (e.g., interface, hardware, software, developers, service providers) users orient themselves to (Solomon and Wash, 2014). This complicates the negotiation of privacy boundaries, as users' source orientation—whether toward the chatbot's interface or its broader system—shifts dynamically (Guzman, 2019). Consequently, privacy management in HMC involves user-driven and machine-augmented efforts that vary depending on which communication sources users believe they are engaging with.

Therefore, key CPM concepts such as ownership and control require explication to account for the contextual dynamics in HMC. First, CPM differentiates primary ownership and co-ownership, where privacy information becomes shared after disclosure (Petronio, 2002). However, given different orientations that may exist in user-chatbot interactions, the idea of co-ownership may be perceived differently when users' source orientation shifts. In addition, "private information changes in degrees of risk based on perceived repercussions for revealing and concealing" (Petronio, 2002, p. 67). These perceived repercussions can shift when users "peel back" the layers of the chatbot that reveal how different components—from interface to backend infrastructure—are involved in collecting, storing and processing data. For example, when the chatbot is perceived primarily as a conversational partner on screen, users may feel less risky and assume that information remains within that immediate interaction. In contrast, when the source is perceived as the service provider (e.g., OpenAI), users may feel that ownership has been transferred or diluted due to a perceived change in risk degree, leading to new expectations of co-ownership and privacy management strategies. It is also important to note that user perceptions of the source do not necessarily alter the actual parameters of ownership as defined by the technical architecture surrounding data governance, meaning that their data are still subject to broader system-level processing and retention.

Second, implicit to CPM is a relational understanding of control (Petronio, 2002; Petronio et al., 2022). Boundary coordination describes the dynamic process of negotiation between relational partners determining rules around 1) whether and who to include/exclude as information co-owners; and 2) the actual content of information divulged. In user-chatbot interactions, the relationality of control lies primarily in users' proactive attempts to manage information flow in relation to constraints or possibilities entailed by the chatbot system, rather than a clean-cut negotiation with the service provider (cf. Vitak et al., 2023). Drawing inspirations from existing studies (Metzger, 2007), chatbot users may perform a kind of "soft control" by withholding or falsifying information to obfuscate personal details (Brunton and Nissenbaum, 2015) and interfere with data collection. Moreover, having some information about the relational partner is crucial to privacy management (Petronio, 2002) as it aids assessment of the perceived consequentiality of privacy disclosure. Thus, information seeking (e.g., reviewing privacy policies and regulations) can also be a control strategy that guides boundary coordination.

The present study bridges CPM with HMC thinking, as well as updates and applies CPM's core concepts—including ownership and control—to understand the possible dynamics emerging from informational privacy management in user-chatbot communication that comprises multiple communication sources users may orient to. The empirical component of this study provides rich

user perspectives on how (co-)ownership and control are made sense of and practised, which serves to address gaps in the literature outlined below.

Existing gaps and research question

Empirical studies on chatbot and user privacy adopting a CPM perspective are relatively scarce. In a between-subject factorial design experiment, where participants were exposed to one of several chatbot conditions varying in interactivity and data-sharing protocols, Sannon et al. (2020) discover that chatbots disclosing user chatlogs to third-party advertisers elicit greater privacy concerns than those sharing data only with the service provider. Liu et al. (2023) employed a similar experimental method and find that information sensitivity moderates privacy concerns: compared to a low sensitivity condition, users asked to disclose highly sensitive information reported elevated privacy concerns and lower willingness to share. These findings support CPM's premise that users view themselves as owners of private information, and violations of user privacy expectations, especially in contexts involving sensitive data, lead to increased concerns and decreased disclosure intentions. Yet, what remains less understood is how users form and negotiate privacy boundaries in everyday interactions with chatbots, as neither study provides an in-depth account of user strategies nor meaning-making practices related to privacy management in real-world settings.

In addition, as HMC is an emerging field (Guzman and Lewis, 2020), scholars have only started to explore privacy issues through an HMC lens (e.g., Ischen et al., 2019; Lutz et al., 2019). On the topic of chatbot and privacy, Ischen et al. (2019) manipulated design choices to test user responses across 3 interface types: a human-like chatbot (with a name and social cues), a machine-like chatbot (with robotic visuals and tone), and an interactive website (with no agent presence). Their finding shows that higher perceived anthropomorphism in chatbots leads to lower privacy concerns and increased disclosure intention (see also, Lim and Shim, 2022). However, this finding sits somewhat paradoxically alongside Sundar's Machine Heuristic (Sundar and Kim, 2019), which posits that users may place greater trust in systems perceived as mechanical, believing them as more neutral and therefore safer for sensitive disclosure. This misalignment warrants further studies to disentangle disclosure intentions from actual privacy behaviours, and to explore how information disclosure is practised as part and parcel of chatbot users' relational privacy management practices.

More broadly, a recent review of conversational agents and privacy finds that much of the research focuses on how user privacy concerns influence self-disclosure to chatbots, with surveys and experimental methods—often relying on isolated variables and artificial conditions—dominating the field (Gumusel, 2024). This suggests that existing studies tend to treat privacy concerns as a static, individual-level variable, rather than as part of an ongoing process of privacy management and negotiation (Palen and Dourish, 2003; Petronio, 2002). As a result, findings are limited to quantitative insights, overlooking the situated and relational nuances of how privacy is negotiated in user–chatbot interactions. Furthermore, while these methods are valuable for hypothesis testing in controlled environments, they may lack ecological validity when applied to everyday HMC (see Spence et al., 2023), where users engage with chatbots in diverse, fluid and context-dependent ways.

Moving beyond quantitative insights and controlled conditions, the current research applies CPM's relational thinking to understand informational privacy in HMC, focusing on how users conceptualise ownership and control in their negotiated decisions around information disclosure to chatbots. It asks: **how do AI chatbot users understand and negotiate information ownership and privacy boundary control in everyday user-chatbot communication?** In addressing this question, this exploratory study contributes to the growing field of HMC and enriches existing scholarships on AI chatbot and privacy through a user-informed approach. It also provides empirical evidence to argue for the applicability of CPM in user-chatbot communication in particular and adds to our understanding of privacy disclosure and management in HMC in general.

Method

This study deploys qualitative thematic analysis (Clarke and Braun, 2017) to investigate how AI chatbot users understand informational privacy and practise privacy management strategies. Data were sourced from Reddit, a social networking platform with forums (sub-reddit) dedicated to specific topics or communities (Proferes et al., 2021). Users' active sharing of privacy-related experiences with chatbots can be seen as a form of community-driven audit that produces lay knowledge and surfaces the (in)capabilities of AI technologies in everyday contexts (Li et al., 2023, cited in Li and Zhang, 2024). A thematic analysis of such narratives contributes to uncovering detailed user perspectives surrounding privacy management in user-chatbot communication and showing how people make sense of AI chatbots in the everyday, which is key to HMC research (Guzman and Lewis, 2019).

Reddit data were chosen over direct user engagement methods (e.g., interviews) because it captures how users naturally articulate their concerns and privacy management strategies. However, it is important to note that online spaces like Reddit are socially shaped; users may tailor their posts for visibility (Shepherd, 2020). Furthermore, Reddit's user base is predominately male, skewing young (Proferes et al., 2021) and may also be over-represented by individuals with higher socio-economic status (Hargittai, 2018). Nevertheless, given the exploratory nature of this study, it is considered an acceptable trade-off. Limitations and their implications for future research are discussed in the conclusion.

Data collection

Data were retrieved via software tool Communalytic. "Privacy" was used as the keyword to retrieve relevant textual materials (called submissions[2]). Phase One was conducted in July 2024 to gather data from sub-reddit *r/ChatGPT*; a key purpose was to assess data quality and evaluate the alignment between theoretical framework and data. This phase yielded 200[3] submissions, which I read through and filtered manually, resulting in 84 relevant submissions. Irrelevant ones were excluded, such as promotional messages, news re-posts, and incomprehensive submissions. Research notes were taken to document preliminary findings. I also conducted a preliminary review of user replies associated with these filtered submissions to assess if they offered additional nuances. Findings suggested that they repeated themes present in the submissions or contained unrelated information. Therefore, replies were excluded for methodological consistency and data quality considerations. Phase Two was conducted in December 2024 to retrieve data from five

sub-reddit forums (see *Table 1* for additional details). All retrieved submissions were reviewed and filtered following the same criteria practiced in July 2024. In total, 193 submissions were included in the thematic analysis.

Table 1. Number of submissions before and after filtering

Sub-reddit name	Number of submissions retrieved	Number of submissions included in analysis
ChatGPT	200~	84
	200^	16*
ClaudeAI	129	32
perplexity_ai	28	12
GeminiAI	98	10
CharacterAI	200	39

~July 2024 dataset.
^December 2024 dataset.
*The final quantity was 100; these were then cross-checked with data from July 2024, resulting in the removal of 84 duplicates.

The 5 chatbot services were chosen for their public accessibility, popularity and active user communities[4]. As conversational systems, they represent a specific sub-set of chatbots underpinned by large language models (Guo et al., 2023) which require vast amount of data for training and iteration (Hepp, 2020). Public documents[5] show that model training draws on three main data types, including Internet content, third-party licensed datasets, and user-/crowd worker-provided information. All 5 services offer users basic privacy safeguards such as data deletion options, privacy settings, and published data policies. Limited protective measures reflect an institutional emphasis on data accessibility and value (Gorwa and Veale, 2024). In data analysis, these operational features of the selected chatbots were considered when examining how users referenced and navigated specific privacy settings and data policies in their submissions.

Procedure of analysis

To conduct the analysis, data (n=193) were compiled and uploaded into NVivo 14. Qualitative thematic analysis (Clarke and Braun, 2017) serves as a flexible methodological tool, as it facilitates both a deductive approach guided by the theoretical framework and an inductive approach to uncover emerging themes specific to the research context. First, I developed an initial coding scheme based on two sources: 1) key CPM concepts such as ownership, control, boundary coordination (Petronio, 2002) and key HMC concepts such as source orientation (Guzman, 2019); 2) notes taken during Phase One. The data were then coded iteratively through constant comparative analysis (Corbin and Strauss, 2008). This means that codes were continually revised and elaborated: new codes were added when necessary, and existing codes were refined or collapsed to address overlaps. Second, submissions containing rich, detailed descriptions of user experiences were exported into Excel for further analysis. Patterns were identified and linked to the research question.

Ethical considerations

I followed established internet research guidelines (British Psychological Society, 2021; Franzke et al., 2020) and assessed ethical issues related to Reddit data (Proferes et al., 2021). A consensus is that online platforms like Reddit are "informal spaces that users often perceive as private but may strictly speaking be publicly accessible" (Franzke et al., 2020, p. 69). Sub-reddit forums like those outlined above do not generally include sensitive information, nor do they bear significant risks of exposing vulnerable individuals or pose immediate harm towards a particular group. Given the number of submissions involved, it was not practical to gain informed consent from each user. These ethical considerations shaped my practices where several strategies were adopted to protect user privacy.

First, after data retrieval, files were downloaded and removed from Communalytic. Second, when reviewing, filtering and analysing submissions, I only looked at the titles and actual content. Any information identifiable to a user (e.g., username/Reddit ID) or a submission (e.g., URL links) was stored in a separate file. This file was used only for verification on Reddit, when submissions contained rich user perspectives and were selected for detailed analysis. Third, I used composite accounts (Markham, 2012) that blended similar statements and themes from multiple users. These accounts, designed to replace direct quotations and to prevent re-identification, are italicised in text.

Findings

Ownership boundaries and associated uncertainties

A prominent theme emerging from the data was users' sense of ownership towards their information. The scope appeared to have significant breadth, covering 3 major domains:

1) access pre-requisites like email address, date of birth, and credit card specifics.
2) tracked information like location, interaction session duration, and Internet Protocol address and other cookie-related details.
3) interaction details that users and chatbots co-create, such as chatlogs and conversation history.

Despite an overall perceived sense of ownership, users tended to express uncertainty in grappling with the extent to which private information is shared with what/whom. Some speculated that their information *might be retained on the server-side or linked to hidden identifiers*, while others feared that *uploaded content could be accessed by anyone with a URL*. These uncertainties were described as *major privacy concerns* and *security failures in the design of the systems*.

One repeated theme in relation to uncertainty of ownership was the opaque and layered nature of chatbot systems. Users raised concerns about whether their interactions with chatbots *were ephemeral*. Some questioned whether it was possible to engage with the system *without leaving a data trace*, asking if their inputs could be *excluded from training datasets*, or if the system could *remain unchanged after their sessions*. What also stands out is that some users demonstrated a notable degree of technical literacy, referencing servers, URLs and training pipelines, suggesting they were not passive users, but actively engaged with and questioned the technological structures shaping their interactions.

In these cases, it seemed that users initially set up privacy boundaries with the chatbot as an information co-owner (thus granting co-ownership), which was the immediate communication source. Data exchange and processing was deemed acceptable to the extent that information remains within the given communication context. This also helps to explain why users considered interaction details such as chatlogs and conversation history as privately owned, even though private information is not necessarily always disclosed. However, as other layers beyond the immediate source manifested (e.g., the system, the language model, the company, other third parties), users began to perceive that their information had moved beyond the original expected scope of interactions with the interlocutor. This triggered a sense of violated ownership rights—a form of boundary turbulence (Petronio, 2002)—leading to discomfort and unease.

However, not all users shared the same level of uncertainty in their understanding of ownership violations. The most notable case was—echoing existing studies (Draper and Turow, 2019; Hargittai and Marwick, 2016)—the resignation trope. These users tended to disregard the importance of data sensitivity as they felt little ability to control their own. This mentality led to lower privacy concerns and an overt focus on benefits to rationalise the lack of clarity around the system's data practices. For example, some users acknowledged privacy risks associated with chatbot user, such as *data retention and third-party access,* but they also expressed *a willingness to accept these risks in exchange for functionality or innovation*. For some, the potential of *real-time internet access* or *personalised assistance* outweighed such risks. Others normalised data sharing, comparing it to *everyday practices like location tracking or app permissions*. As one user put it, *privacy is important, but the possibilities are just too exciting to ignore*.

CPM posits that people engage in a mental risk-benefit calculus to determine the degree of privacy disclosure as an inherent part to privacy management practices (Petronio, 2002). As these cases suggest, in user-chatbot communication users may engage in tilting the balance towards benefits gained by downplaying risks, so that privacy disclosure is justified on an intrapersonal level. In this light, primary ownership becomes a personal sacrifice and obscured by the multiple layers of information exchange that a chatbot systems entails.

Negotiating control through privacy boundary making

Information control is fundamental to active privacy management practices (Altman, 1975; Palen and Dourish, 2003) and is viewed as tactics to balance the dialectical tension between openness and closedness (Petronio, 2002; Child et al., 2009). Uncertainties around ownership boundaries emerged as a key characteristic of data privacy management in user-chatbot interactions. CPM tenets suggest that risk and uncertainty perceptions contribute to amplifying such tension and subsequently motivating people to develop mitigation strategies to restore the balance. However, while uncertainty served as a motivation that prompted some users to introduce protective measures to maximise benefits gained while minimise risks of privacy loss, technological restrictions also interfered with users' information management intentions and practices. Boundary coordination in user-chatbot interactions became a negotiated effort and interplay between human and machine agency.

To start with, users engaged in information seeking as a strategy to aid disclosure decision-making, as gathering adequate information about the relational partner helps to assess risks and inform

disclosure depth (Petronio, 2002); for example, *going through privacy policies before setting up the account, for peace of mind*. In fact, privacy policies of chatbots were frequently referred to in users' articulation of privacy management, which formed part of users' knowledge base. Yet existing studies (Ragab et al., 2024) suggest the purpose of privacy policies is not always aligned with chatbot users' interests; terms and condition of data usage is left intentionally vague and open to interpretations. This observation is also evident in the current study. Some users welcomed recent improvements to privacy controls—such as *clearer opt-out options* or *data retention limits*—and expressed *a newfound willingness to use chatbots for highly specific tasks*. However, this optimism was tempered by 1) the ambiguity in *policy definitions of data collection and processing* or *incomplete explanations in FAQs*; and 2) the lack of *sufficient alternatives to opt out without giving up certain benefits*. Therefore, users were "a bit cautious of jumping in with both feet".

The proactive approach to reading privacy policies echoes CPM's concept of boundary ownership, in the sense that it involves users' sense-making of the rules and terms that govern the control and management of personal information (Petronio et al., 2022). However, users often found themselves at the mercy of intentionally vague policies, highlighting a mismatch between user expectations and system realities. This led a limited number of users to adopt protective measures ranging from the use of virtual private networks (VPN) and alternative payment methods (e.g., virtual debit cards) to active adjustments of privacy settings, use of chatbot-specific features like ChatGPT's temporary chat function and information deletion request to the organisation.

However, the effectiveness of these reported strategies was largely hindered because of system restrictions and updates, thus creating frictions in these user-initiated practices to negotiate privacy boundaries. Some users noted that opting out of data collection *came at the cost of losing core features like chat history or voice-to-voice interaction*. Others described *having to manually adjust settings for each session* – a burdensome process that *discouraged consistent privacy protection*. There was also dissatisfaction with *restrictive system-wide measures*, such as VPN blocks, which was perceived to *penalise legitimate privacy practices*.

CPM's metaphors of thick and thin boundaries (Petronio, 2002) provide the basis to understand such frictions between the user and the chatbot. Thick boundaries allow less permeability, meaning that less information is permitted to pass, whereas thin ones, with a higher degree of permeability, grant relatively easier information access. Users' tactics to manage data collection and processing could be viewed as attempts to thicken privacy boundaries by either opting out completely (e.g., adjusting privacy settings) or "confusing" the system (e.g., using VPN), which reflects a desire to control information permeability. The chatbot system, on the other hand, may be seen as thinning out the boundaries; not through negotiation with users, but through creating obstacles, limiting usability or disabling user solutions in the name of data safety. These user perspectives capture the frictional nature of privacy boundary coordination that emerges and intensifies as users practise their tactical agency while the chatbot system exerts its restrictions.

Discussion

Through a qualitative thematic analysis of user submissions from five sub-reddit forums, this study explores how AI chatbot users manage their data and negotiate communication privacy boundaries in human-machine communication. The exploration reveals that in user-chatbot communication,

privacy control is an unstable process of boundary negotiation; while some users attempt to assert ownership and protect their information, others resort to resignation or pragmatism. Users' privacy management strategies are met with system-imposed constraints, resulting in interactional frictions and privacy boundary turbulence. The study extends communication privacy research in HMC by presenting users' diverse perspectives on privacy boundary making as meaning creation between human and machine.

A key finding is users' struggles with uncertainties as they navigate information ownership. This uncertainty emerges as users orient to different communication sources, reflecting the layered communication structure of chatbot systems which distribute communicative agency across both visible and invisible components (Hepp, 2020). Upon initial encounters, users share information with their chatbot counterparts and regard data processing and storage acceptable with 'something' immediate on the other side of the interface that showcases communicative capabilities. This tendency, according to the classic Computers-Are-Social-Actors (CASA) tenet (Reeves and Nass, 1996), suggests how users readily apply social scripts to machines displaying enough social traits, such as natural language production. Building on experimental studies (Ischen et al., 2019; Lim and Shim, 2022), this orientation towards the chatbot as a responsive communicator may help to explain why some users initially disclose personal information, without considering privacy implications like information ownership violations.

The present study also builds on source orientation literature (Guzman, 2019; Solomon and Wash, 2014) and presents empirical evidence of deliberate user efforts to assess communication sources and adopt intentional approaches to privacy management with chatbots. The evidence is exemplified where users' initial orientation to the chatbot as a social actor is disrupted by uncertainties – particularly when they become aware of underlying operational layers (e.g., language model; service provider). The perceived inclusion of additional co-owners external to the initial privacy boundaries triggers a tightened desire for primary information ownership and amplifies privacy anxieties – an observation echoing Sannon et al.'s conclusion (2020).

CPM (Petronio, 2002) helps to contextualise the privacy implications of source orientation in HMC, as it provides a useful framework to understand how users' information ownership is challenged and negotiated in and through communication with chatbots of a perceived dual identity: social actor and technological assemblage. Relational partners in interpersonal settings negotiate rules regarding ownership and control of information and re-negotiate such rules to stabilise boundary turbulence when privacy breakdowns occur (Petronio et al., 2022). One's relationship with an AI chatbot—and by extension the algorithms, software, hardware, developers and the company that manages that chatbot—is structurally one-sided with limited user freedom and system-level transparency to determine the exact boundaries of data privacy. This is partially why perceived lack of control leads to privacy cynicism (Draper and Turow, 2019) and apathy in networked environments (Hargittai and Marwick, 2016). Hence unsurprisingly, to cope, some users rationalise their disclosures, downplaying privacy risks in favour of perceived benefits – a cognitive dissonance reduction strategy.

Another key finding is that users' desire to achieve relational control over private information is typified by situational tactics to regulate privacy boundaries with chatbots. CPM (Petronio, 2002)

explains that boundary thickness and thinness are determined by the degree of relational control over information flow. These user-initiated ways of boundary making showcase user obfuscation strategies (Brunton and Nissenbaum, 2015), defined as deliberate attempts to interfere with data collection, which can be seen as demonstrations of user agency to fortify boundaries by increasing thickness and thus resist unintended information flow. Yet, our contemporary digital ecosystems favour increasingly thinner boundaries to facilitate information collection, processing, and accumulation (Vitak et al., 2023). For AI technologies, data governance prioritises data accessibility and sharing, with limited platform-level guardrails for privacy invasion or user control (Gorwa and Vaele, 2024). These contradictory forces create interactional tensions between users' privacy management practices and chatbots' techno-social affordances. As Floridi (2013, p. 228) notes, "informational privacy is a function of the ontological friction in the infosphere, that is, of the forces that oppose the information flow within the space of information". This means that to enhance user privacy regarding information, ontological frictions must increase between the user and the chatbot. However, as illustrated by user vignettes in this study, the onus of introducing frictions falls on users who need to devise ways of resistance that are often countered by constant system updates to limit user control, or risk losing chatbot features.

User-chatbot communication introduces burning privacy challenges to resolve. Scholars (Natale and Depounti, 2024) have cautioned against the deceitful nature of AI chatbots, not because they are necessarily capable of deceiving users into something sinister but that their appearance as a communicator able to make sense in natural language invites social reactions from users who may feel a sense of continuity in their user-chatbot relationships. Although there is no direct proof in this study, this deception may have worked to encourage users to disclose more than they knew. From this perspective, the present study bears practical implications that can inform chatbot design practices to ensure transparency and data governance policies to serve users' interests. Designers and developers should consider including clear in-situ signposts (e.g., disclosure statement on the interface) to inform users of chatbots' role in data collection, processing and storage. Guardrails informed by the privacy-by-design principles (Cavoukian et al., 2010) can be inscribed into design choices to increase ontological frictions between the user and the chatbot, which can ease the burden of privacy management on users. As communication privacy is context-dependent and no one-time consent is adequate to ensure stable privacy boundaries (Petronio, 2002), policymakers should explore, in addition to the current informed consent framework, the feasibility of dynamic consent mechanisms (e.g., periodic re-confirmation of consent) to prevent risks of unwarranted over-disclosure from users.

Conclusion

AI technologies are increasingly becoming part of the social fabric of everyday life (Guzman and Lewis, 2020). By extending CPM to HMC, this study explores how human communication behaviours, such as the disclosure of information and the management of communication privacy, are shaped by situational interactions between users and their chatbots. With a user-centric approach, this exploration contributes to scholarship in communication privacy research in HMC (Lutz, 2023), specifies practical implications that can benefit the design of socio-technical systems, and provides an initial assessment of boundary regulations of AI chatbot data as users continue to explore these technologies. CPM's emphasis on ownership and control entails responsibility for

each relational partners involved (Petronio, 2002). To ensure the healthy and productive growth of AI that can benefit all, we must prioritise ethical AI development, establish robust data protection measures to safeguard user privacy, and hold AI systems accountable to foster informed decision-making in data-related practices.

This research has several limitations. First, it relies on Reddit data which only capture a fraction of users' experiences. As explained, the dataset was possibly over-represented by young male users. Findings also suggest a notable level of technical literacy, which is related to a higher socio-economic status (Hargittai, 2018). Furthermore, platform features like user-directed content moderation and algorithmic sorting and ranking can impact how narratives gain (in)visibility (Shepherd, 2020), which could subsequently impact the way Communalytic retrieved the data. For example, all five sub-datasets had less than half of the total retrieved submissions deemed relevant after review. Therefore, results of this study must be approached as an initial exploration and interpreted with caution. Future research is encouraged to engage human participants of diverse demographic backgrounds, obtain first-hand user perspectives of privacy management with chatbots, and identify shifts in disclosure patterns over time.

Second, this study only focuses on the informational aspect of privacy as it is most relevant to chatbot use (Lutz, 2023). The chatbots selected for the study represent only a sub-set of privately owned, publicly accessible AI technologies powered by large language models. Privacy is a complex concept irreducible to a single dimension (Solove, 2006), and different types of AI technologies entail different privacy implications in HMC (Lutz et al., 2019). For example, privacy research into social robotics needs to consider their spatial implications given its physical embodiment in domestic contexts like at home with users. Future scholarships should extend CPM to include other AI types and adopt a comparative angle to understand similarities and differences in user perceptions and privacy management behaviours.

Acknowledgements

I would like to thank the two anonymous reviewers for their thoughtful and constructive feedback, which helped to improve this article. I am also grateful to the Editorial Team for their support throughout the review process.

Special thanks to Wonsun Shin, Xin Pei and Zi Lin for their encouragement and insightful conversations throughout the development of this work.

This research was supported by the Melbourne Research Scholarship.

Notes

1. Communalytic is a no-code computational social science research tool developed by Gruzd and Mai (n.d); for more information, please visit: https://communalytic.org/frequently-asked-questions/.

2. Reddit has 5 ways to categorise submissions. Given the exploratory nature of this study, only the newest/most up-to-date submissions were retrieved for analysis. For more information on submission sorting, please visit: https://support.reddithelp.com/hc/en-us/articles/19695706914196-What-filters-and-sorts-are-available.

3. When a keyword is used, the maximum number of submissions Communalytic can retrieve is 200.

4. As of 23 January 2025, the approximate numbers of subscribers (as shown on Reddit) are 8.8 million (*r/ChatGPT*), 134 thousand (*r/ClaudeAI*), 44 thousand (*r/perplexity.ai*), 13 thousand (*r/GeminiAI*), and 2.2 million (*r/Character.AI*).

5. For more information, please refer to data and privacy policies: 1) ChatGPT: https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed; 2) Claude: https://privacy.anthropic.com/en/articles/10023555-how-do-you-use-personal-data-in-model-training; 3) Perplexity: https://www.perplexity.ai/hub/technical-faq; 4) Gemini: https://cloud.google.com/gemini/docs/overview; 5) Character.AI: https://character.ai/privacy.

References

Allen, J. F. (2003). Natural Language Processing. In A. Ralston, E. D. Reilly, & D. Hemmendinger (Eds.), Encyclopedia of Computer Science (4th Edition) (pp. 1218-1222). Chichester: Wiley.

Altman, I. (1975). The environment and social behavior: Privacy, personal space, territory, crowding. Monterey, CA: Brooks/Cole Publishing.

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.

British Psychological Society. (2021). Ethics guidelines for Internet-mediated research. Leicester: British Psychological Society.

Brunton, F., & Nissenbaum, H. (2015). Obfuscation: a user's guide for privacy and protest. Cambridge, Massachusetts: The MIT Press.

Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. Information, Communication & Society, 20(1), 30-44.

Cavoukian, A., Taylor, S., & Abrams, M. (2010). Privacy by design: essential for organizational accountability and strong business practices. Identity in the Information Society, 3(2), 405-413.

Child, J. T., Pearson, J. C., & Petronio, S. (2009). Blogging, communication, and privacy management: development of the blogging privacy management measure. Journal of the American Society for Information Science and Technology, 60(10), 2079-2094.

Clarke, V., & Braun, V. (2017). Thematic analysis. The Journal of Positive Psychology, 12(3), 297-298.

Corbin, J., & Strauss, A. (2008). Basics of qualitative research: techniques and procedures for developing grounded theory. Los Angeles: Sage.

Draper, N. A., & Turow, J. (2019). The corporate cultivation of digital resignation. New Media & Society, 21(8), 1824-1839.

Eynon, R., & Young, E. (2021). Methodology, legend, and rhetoric: The constructions of AI by academia, industry, and policy groups for lifelong learning. Science, Technology, & Human Values, 46(1), 166–191.

Floridi, L. (2013). The ontological interpretation of informational privacy. In The ethics of information (pp. 228–260). Oxford, UK: Oxford University Press.

Franzke, A. S., Bechmann, A., Zimmer, M., & Ess, C. (2020). Internet research: Ethics guidelines 3.0. Retrieved July 2, 2024, from https://aoir.org/reports/ethics3.pdf

Gorwa, R., & Vaele, M. (2024). Moderating model marketplaces: Platform governance puzzles for AI intermediaries. Law, Innovation and Technology, 16(2), 341–391.

Gruzd, A., & Mai, P. (n.d.). Communalytic: A computational social science research tool for studying online communities and discourse. Retrieved July 2, 2024, from https://communalytic.org

Gumusel, E. (2024). A literature review of user privacy concerns in conversational chatbots: A social informatics approach—An Annual Review of Information Science and Technology (ARIST) paper. Journal of the Association for Information Science and Technology, 76(1), 121–154.

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century. Communication +1, 1(1). https://doi.org/10.7275/R5QJ7F7R

Guo, Z., Jin, R., Liu, C., Huang, Y., Shi, D., Supryadi, Yu, L., Liu, Y., Li, J., Xiong, B., & Xiong, D. (2023). Evaluating large language models: A comprehensive survey. arXiv. https://doi.org/10.48550/arXiv.2310.19736

Guzman, A. L. (2018). What is human–machine communication, anyways? In A. L. Guzman (Ed.), Human-machine communication: Rethinking communication, technology, and ourselves (pp. 1–28). New York, NY: Peter Lang.

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. Computers in Human Behavior, 90, 343–350.

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. New Media & Society, 22(1), 70–86.

Hargittai, E. (2018). Potential biases in big data: Omitted voices on social media. Social Science Computer Review, 1–15.

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

Hepp, A. (2020). Artificial companions, social bots and work bots: Communicative robots as research objects of media and communication studies. Media, Culture & Society, 42(7–8), 1410–1426.

Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2019). Privacy concerns in chatbot interactions. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), Chatbot research and design (pp. 34–48). Cham, Switzerland: Springer.

Kang, H., Shin, W., & Huang, J. (2022). Teens' privacy management on video-sharing social media: The roles of perceived privacy risk and parental mediation. Internet Research, 32(1), 312–334.

Li, H., & Zhang, R. (2024). Finding love in algorithms: Deciphering the emotional contexts of close encounters with AI chatbots. Journal of Computer-Mediated Communication, 29(5). https://doi.org/10.1093/jcmc/zmae015

Lim, S., & Shim, H. (2022). No secrets between the two of us: Privacy concerns over using AI agents. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 16(4). https://doi.org/10.5817/CP2022-4-3

Liu, Y., Yan, W., Hu, B., Lin, Z., & Song, Y. (2023). Chatbots or humans? Effects of agent identity and information sensitivity on users' privacy management and behavioral intentions: A comparative experimental study between China and the United States. International Journal of Human–Computer Interaction, 40(19), 5632–5647.

Lutz, C. (2023). Privacy and human–machine communication. In A. L. Guzman, R. McEwen, & S. Jones (Eds.), The SAGE handbook of human–machine communication (pp. 310–317). London, UK: SAGE.

Lutz, C., Schöttler, M., & Hoffmann, C. P. (2019). The privacy implications of social robots: Scoping review and expert interviews. Mobile Media & Communication, 7(3), 412–434.

Markham, A. (2012). Fabrication as ethical practice. Information, Communication & Society, 15(3), 334–353.

Metzger, M. J. (2007). Communication privacy management in electronic commerce. Journal of Computer-Mediated Communication, 12(2), 335–361.

Natale, S., & Depounti, I. (2024). Artificial sociality. Human–Machine Communication, 7, 83–98.

Palen, L., & Dourish, P. (2003). Unpacking "privacy" for a networked world. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 129–136). New York, NY: ACM.

Petronio, S. (2002). Boundaries of privacy: Dialectics of disclosure. Albany, NY: State University of New York Press.

Petronio, S., Child, J. T., & Hall, R. D. (2022). Communication privacy management theory: Significance for interpersonal communication. In D. O. Braithwaite & P. Schrodt (Eds.), Engaging theories in interpersonal communication (3rd ed., pp. 314–327). New York, NY: Routledge.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. Social Media + Society, 7(2). https://doi.org/10.1177/20563051211019004

Ragab, A., Mannan, M., & Youssef, A. (2024). "Trust me over my privacy policy": Privacy discrepancies in romantic AI chatbot apps. In 2024 IEEE European Symposium on Security and Privacy Workshops (pp. 484–495). New York, NY: IEEE.

Reeves, B., & Nass, C. (1996). How people treat computers, television, and new media like real people and places. Cambridge, UK: Cambridge University Press.

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.

Shepherd, R. P. (2020). Gaming Reddit's algorithm: r/the_donald, amplification, and the rhetoric of sorting. Computers and Composition, 56. https://doi.org/10.1016/j.compcom.2020.102572

Sher, G., & Benchlouch, A. (2023). The privacy paradox with AI. Reuters. Retrieved December 15, 2024, from https://www.reuters.com/legal/legalindustry/privacy-paradox-with-ai-2023-10-31/

Skjuve, M., Brandtzaeg, P. B., & Følstad, A. (2024). Why do people use ChatGPT? Exploring user motivations for generative conversational AI. First Monday, 29(1). https://doi.org/10.5210/fm.v29i1.13541

Solomon, J., & Wash, R. (2014). Human-what interaction? Understanding user source orientation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58(1), 422–426.

Solove, D. J. (2006). A taxonomy of privacy. University of Pennsylvania Law Review, 154(3), 477–560.

Spence, P. R. (2019). Searching for questions, original thoughts, or advancing theory: Human–machine communication. Computers in Human Behavior, 90, 285–287.

Spence, P. R., Westerman, D., & Luo, Z. (2023). Observing communication with machines. In A. L. Guzman, R. McEwen, & S. Jones (Eds.), The SAGE handbook of human–machine communication (pp. 220–227). London, UK: SAGE.

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–9).

Turkle, S. (2005). The second self: The human spirit in a computer culture. Cambridge, MA: MIT Press.

Vitak, J., Kumar, P. C., Liao, Y., & Zimmer, M. (2023). Boundary regulation processes and privacy concerns with (non-)use of voice-based assistants. Human–Machine Communication, 6, 183–201.

Westfall, C. (2023). New research shows ChatGPT reigns supreme in AI tool sector. Forbes. Retrieved January 5, 2024, from https://www.forbes.com/sites/chriswestfall/2023/11/16/new-research-shows-chatgpt-reigns-supreme-in-ai-tool-sector/

Westin, A. F. (1967). Privacy and freedom. New York, NY: Atheneum.