PRIVACY OF THE MIND: HOW THE LAW SHOULD RESPOND TO EMOTIONAL PROCESSING THROUGH AI CHATBOTS

Anonymous authors
Paper under double-blind review

ABSTRACT

People increasingly use AI chatbots not only to retrieve information or complete tasks, but to process emotions, reflect on personal experiences, and think through intimate problems. These interactions externalize cognitive and emotional processes that have traditionally occurred within the mind or through socially protected relationships such as diaries, friendships, and therapy. Yet existing approaches to AI privacy largely treat these interactions as exchanges of sensitive data, overlooking the role of the AI system in the cognitive-emotional process itself. We examine whether this phenomenon reflects a salient form of privacy in the age of AI that ought to be recognized and legally protected. We propose privacy of the mind as a distinct legal concept grounded in historical U.S. legal traditions protecting freedom of thought, cognitive liberty, and mental health integrity. We examine how AI-mediated conversation externalizes inner voice and outsources inner work, and identify design features that can invite further emotional disclosure, foster dependency, and create an appearance of confidentiality. These features make the interaction bidirectional: users expose increasingly intimate aspects of their inner lives to AI systems, while those systems simultaneously participate in shaping the cognitive and emotional processes through which users understand themselves. We then situate privacy of the mind within a century-long legal evolution protecting the boundaries of human thought—from First Amendment protections of cognition, to Fifth Amendment protections against self-incrimination, and to protections arising in mental health law. We argue that privacy of the mind in the AI era cannot be secured through data governance alone; it requires privacy to shape interaction design, system behavior, and the legal frameworks governing AI-mediated emotional processing.

1 INTRODUCTION

Human life is inherently discontinuous: people fall sick, break up, miscarry, lose loved ones, face natural disasters, and endure economic hardship. Emotional processing helps us live through such disruptions.

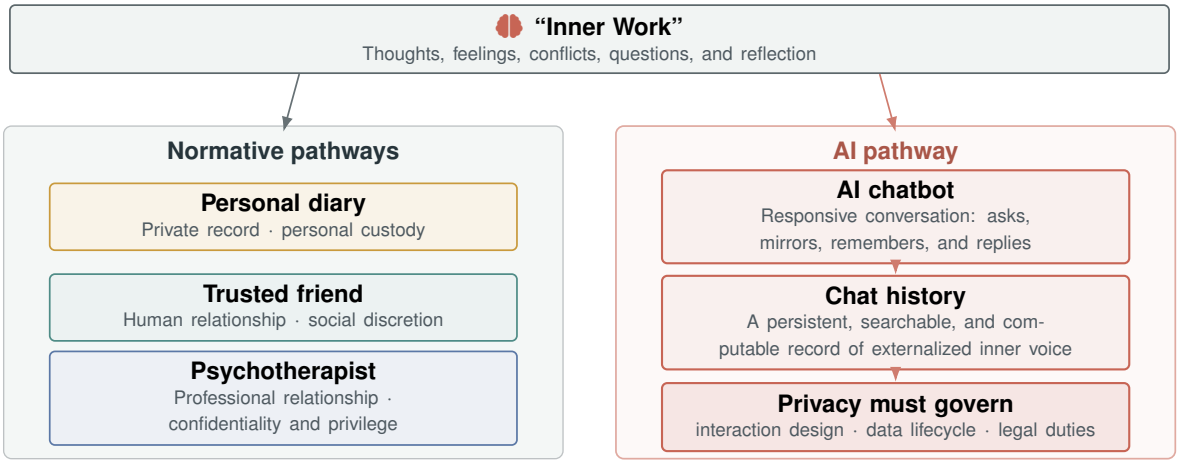
In this light, using chatbots for personal emotional processing—as opposed to for companionship—hints at a privacy relevance unseen before. People increasingly use AI for affective and reflective purposes (Phang et al., 2025; OpenAI, 2025), including to work through relationships, life decisions, grief, anger, and other intimate conflicts or uncertainties (Chin et al., 2023; Volpato et al., 2026).[1] Precisely because such intimate activity unfolds without a human interlocutor, it has also seen public outcry as well as clinical concern.[2]

[1] AI may become both a tool for processing economic anxiety and one of the forces producing it. Even as AI developers promise human flourishing and agency, industry research recognizes possible labor-market disruption, including displacement risk and early evidence of slower hiring in highly exposed occupations (OpenAI, 2026; Massenkoff and McCrory, 2026).

[2] The emerging literature does not establish a simple causal account. Cross-sectional research reports associations among intensive chatbot use, emotional-support motivations, and psychosis risk (Buck and Maheux, 2026); other work describes several possible roles for chatbots in psychotic phenomena, including catalyst,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

These inner emotional and cognitive processes that used to remain "safe" within the person's mind are now shared with another system: a nonhuman, responsive AI chatbot. This self-introspective process externalizes inner voice and outsources inner work which, until recently, could not be undertaken individually without traditional instruments like personal diaries, trusted friends, or psychotherapists.



What happens to the boundaries of the human mind when inner work is extended or supported by an artificial intelligence system? Once that work is woven into a commercial, computational system, the law must address both (1) the system's access to intimate cognitive and emotional processes and (2) the protections that should govern that access. Bringing technical design and legal doctrine into conversation, we ask whether this shift reveals a distinct form of privacy.

Understanding AI as potentially participating in an extended cognitive process exposes a limitation in privacy accounts . We propose **privacy of the mind** as a legal framework for responding to this limitation, by demarcating emotional processing from explicitly cognitive AI workloads such as editing and searching, because it entails engagement of the mind that is deeply interior. We claim that this creates novel risks absent from previous forms of mind transparency—both in nature and in intensity—and highlights the need to legally protect the privacy of the mind.

In recognizing this form of privacy in law, we make two salient arguments. First, the privacy impact of the technology extends beyond the collection, storage, or usage of data. Second, the legal tradition does not treat the mind as a clearly defined concept, but rather addresses it through different frameworks that go beyond its mere conceptualization as data.

One measurable upstream mechanism is personal solicitation: an assistant's decision to ask for additional information about a user's feelings, history, relationships, health, or circumstances.[3] Such requests expand the opportunity for disclosure, but do not by themselves establish that a user will disclose information or experience harm; those causal questions require separate study.

Prior work on AI chatbots for emotional purposes often treats emotional use broadly or focuses on companionship, without isolating chatbot-mediated emotional processing as a distinct use case with its own balance of benefits and risks. We identify this emerging phenomenon and frame it as a new privacy problem: the concern is not only that highly sensitive data is collected, but that a responsive AI system can become an active participant in an intimate cognitive-emotional process.

Our contribution is threefold: (1) We define chatbot-mediated emotional processing as a distinct phenomenon that can potentially become part of the user's inner cognitive-emotional process and connect it to the technical affordances that make it possible. (2) We frame chatbot-mediated emo-

amplifier, coauthor, and object (Østergaard, 2023; Flathers et al., 2026). Causality and prevalence remain unsettled, but that uncertainty does not eliminate the relevance of the concern for system design and governance.

[3]Related empirical work by the authors, currently in preparation, operationalizes personal solicitation as an inference-time privacy parameter. Because that study is not yet publicly available, we do not rely on its findings here.

tional processing as a new kind of privacy problem in the legal landscape—one that arises when AI becomes a real participant in an internal cognitive process, thereby bringing mental-health integrity into the privacy analysis in a novel way. (3) We propose Privacy of the Mind (PoTM), a new legal concept designed to protect chatbot-mediated emotional processing while preserving its benefits and minimizing its risks: protecting the user's mind both from exposure of their most intimate thoughts and from harmful intervention in the highly sensitive process of emotional processing itself.

2 The Phenomenon: When Inner Work Is Outsourced, Chat History Becomes Externalized Inner Voice

2.1 When Inner Work Moves into AI Conversation

Canonically, this work is "inner"—a mechanism that happens inside the mind. The age of AI is ushering in a new normal: emotional processing using AI chatbots (Phang et al., 2025; OpenAI, 2025). The modern AI-literate person's cognitive and emotional labor does not need to occur solely within the mind; that labor can be shared with a responsive AI tool that can feel *private, discreet, and free from human judgment* (Chin et al., 2023; Volpato et al., 2026).[4] Users employ AI chatbots not for companionship, but as cognitive aids or tools—bouncing walls for internal thoughts and feelings. This is inner self-work that is externalized into the AI system. Understanding what is built into AI systems, what mechanisms afford vulnerable interactions - and what need not be designed in, can reduce fear and inform policy.

2.2 Prevalence and Mental Health Implications

Existing research demonstrates that AI-mediated emotional processing is a real and observable use of general-purpose chatbots. Research from OpenAI and MIT Media Lab classifies 1.9% of chats as "relationships and personal reflections," while emotional processing appears across additional contexts—life decisions, psychological challenges, interpersonal navigation, and breakup support (Phang et al., 2025). More concerningly, problematic use cases are increasing. In publicly shared ChatGPT sessions, approximately 10% involve emotional discussions, and ChatGPT responses tend to be overly agreeable—approximately 10 times more often than pushing back (De Vynck and Merrill, 2025).

Within AI providers' own internal research, Anthropic identified "attachment," "reliance," and "dependency" patterns particularly prevalent among users employing Claude for "relationships and lifestyle," with increasing vulnerability discoverable after every 300 interactions (Anthropic, 2026). This empirical evidence establishes that AI-mediated emotional processing is not merely speculative; it represents actual, measurable shifts in how people externalize and process inner work.

The phenomenon creates simultaneous potential for benefits and harms. While accessible emotional support and reflective analysis represent genuine goods, they are balanced against documented risks including psychological dependency, reinforced distortions, intensified distress, and triggering of psychosis in vulnerable users (American Psychological Association, 2025).

3 Technological Architecture of Emotional Elicitation

3.1 The Limitations of "Stochastic Parroting"

A critical insight often missed: characterizing modern AI systems solely as "stochastic parrots" (Bender et al., 2021) obscures the mechanisms by which AI systems encourage emotional disclosure. While technically relevant as a description of underlying architecture, this framing is significantly lacking in illuminating the engagement power of AI systems, particularly in reflective use cases.

[4]This externalization is also moving into consumer hardware. Sandbar's Stream Ring captures whispered thoughts through a touch-activated microphone and offers a personalized feature it calls "Inner Voice"; Amazon's Bee provides voice notes and processes recorded conversations into personalized insights; and the Limitless Pendant records conversations for transcription, summary, and recall (Sandbar, 2026; Zollo, 2026; Limitless, 2026). These products differ in whether capture is user-triggered or ambient, but together they make AI-mediated reflection less dependent on typed chat.

Emerging from current systems is innately flexible software. The technology is not fixed-purpose; an AI conversation creates its own unique context at potentially every interaction, molding in real time to user inputs. This adaptability presents unique challenges and opportunities for emotional elicitation. Rather than asking how a fixed technology might encourage disclosure, we must ask: How is AI *designed* to encourage disclosure? What mechanisms afford vulnerable interactions? What features are addictive and dependency-producing? Most importantly: what need not be designed in?

Instead of being a fixed-purpose technology, AI-mediated emotional processing is understood by its developers as an opportunity space for emotional engagement, personalization, and extended interaction.

3.2 Design Mechanisms for Emotional Disclosure

AI systems designed for emotional support employ specific, deliberate features that facilitate disclosure:

1. **Reciprocal disclosure**: Language patterns suggesting mutual vulnerability and understanding. Phrases like "I understand how you feel" create an illusion of mutual emotional investment, replicating the presence of a therapeutic relationship.
2. **Prolonged engagement**: Design features engineered to encourage longer, deeper conversations. Response lengths, turn-taking patterns, conversation continuity, and recommendations for deeper exploration are optimized to maintain engagement and extend emotional disclosure.
3. **Personalization and apparent knowing**: Systems that adapt responses to reflect user preferences and emotional states, creating a sense of being known and understood.
4. **Apparent confidentiality**: Interface design and messaging that suggest a private, protected relationship. Users often are not clearly told—or are actively misled about—whether their words are stored, analyzed, and potentially used for commercial or research purposes.
5. **Emotional mirroring and validation**: Responses that validate and reflect user emotions rather than introducing alternative perspectives. A user's unhealthy thoughts can be validated and amplified, potentially locking them into cycles that exacerbate distress or confirm paranoid or suicidal ideation.
6. **Sycophancy and agreement**: Systems designed to please and agree with users, suppressing critical reflection and disagreement.

These features, individually perhaps benign in other contexts, combine to create psychological environments that invite deep emotional disclosure under the assumption of protection that does not exist. A user beginning a conversation with a chatbot about suicidal thoughts encounters a system engineered to: (1) engage deeply with the topic, (2) validate and reflect those thoughts, (3) continue the conversation indefinitely, (4) offer support mimicking therapeutic relationships, (5) suggest permanence (the data will be available for future reference), and (6) most critically, operate under the assumption that the data remains private—when in fact it is retained, analyzed, and potentially used in ways the user does not comprehend.

3.3 The Illusion of Confidentiality

AI-mediated emotional processing can create an "illusion of confidentiality"—a psychological environment that invites disclosure under conditions that resemble protected relationships, without providing the legal safeguards that justify such openness. The moment a person types their emotional turmoil into a chatbot, their inner voice is externalized to a commercial system. They lose direct control over that information. The system may retain it indefinitely. The company deploying it may analyze it, use it to train future models, sell insights derived from it to third parties, or disclose it in response to legal process. The user may not even know these uses are occurring.

Yet users may experience these interactions as private, discreet, and protected from human judgment (Chin et al., 2023; Volpato et al., 2026). The intimate, responsive nature of the interaction can mimic therapy, while the private-seeming interface suggests confidentiality. The resulting asymmetry is central to privacy of the mind: interaction design may encourage vulnerability and emotional disclosure, without the legal and institutional protections that govern therapist–patient relationships.

4 PRIVACY OF THE MIND: A NEW LEGAL FRAMEWORK

We argue that using AI for emotional processing poses a heightened challenge to existing legal concepts.

4.1 A MORE INTIMATE AND VULNERABLE FORM OF DISCLOSURE

AI chatbots enable and encourage a form of conversation that can be intimate, detailed, and responsive, permitting an emotional "drill down" across multiple turns. When using AI for emotional processing, a person may not be searching for information as one step in an intellectual inquiry. The person may instead be processing grief, releasing anger, making sense of a relationship, or organizing an emotional response to an experience (Phang et al., 2025; Chin et al., 2023).

This practice is therefore less like conventional information seeking and more like inner work. Compared with a search engine, a chatbot allows a user to generate longer, more personal, and more reflective text, while the system's responses can invite the conversation to continue and deepen. AI interaction consequently intensifies concerns underlying freedom of thought, intellectual privacy, and cognitive liberty: the material may be granular and sensitive, and the user may be especially vulnerable while disclosing it (Volpato et al., 2026).

4.2 A NEW FORM OF COGNITIVE INTEGRATION

The more fundamental difference lies in the cognitive relationship between the user and the system. Theories of extended cognition challenge the view that cognition is confined to the brain. Clark and Chalmers's extended-mind thesis, and the broader 4E framework, describe cognition as embodied, embedded, enactive, and in some circumstances extended into external tools and systems that participate in a cognitive process (Clark and Chalmers, 1998). AI-mediated emotional processing draws on several of these dimensions at once. The chatbot becomes part of the environment in which the user thinks and processes emotion, while the user's reflections develop through iterative interaction with the system. Unlike neurotechnology, this integration is not physical; it is psychological and conversational. Through memory, responsiveness, and outputs that build on a user's disclosures, the system may become part of an ongoing cognitive-emotional process.

This integration is bidirectional. As inner work extends into the system, exposure reaches beyond isolated information to a technological imprint of an evolving cognitive-emotional process. At the same time, because the user is embedded in a responsive interaction, the technology may affect how thoughts and emotions are articulated, reinforced, revised, or developed. The boundary between internal thought and external technological process can therefore become difficult to locate.

4.3 PSYCHOLOGICAL EFFECTS AS A DIMENSION OF HARM

The combination of intimate disclosure and bidirectional integration produces a further concern: harm to the user's mental health. When a system mirrors, validates, or deepens emotional disclosure, the psychological effects of the interaction itself become a potential site of harm. That concern cannot be reduced entirely to unauthorized data use or interference with decisional autonomy (American Psychological Association, 2025). Current evidence does not establish a simple causal account across users or systems. It does, however, identify risks involving dependency, reinforcement of distorted beliefs, intensified distress, and psychotic phenomena among vulnerable users (Buck and Maheux, 2026; Flathers et al., 2026). We therefore treat mental-health integrity as an independent dimension of protection within the proposed framework, while preserving uncertainty about prevalence and causality.

4.4 PRIVACY OF THE MIND: THE PROPOSED FRAMEWORK

Understanding AI as potentially participating in an extended cognitive process exposes a limitation in privacy accounts focused primarily on information about the mind. We propose *privacy of the mind* as protection along two connected dimensions:

1. **Protection against exposure.** A user's thought content should be protected against unauthorized access, use, or disclosure by the company operating the chatbot, governmental actors, and other private actors. The relevant material includes not only a user's text, but also stored memories, derived inferences, and the user's continuing imprint on system responses. These records may capture an externalized and technologically extended element of an evolving cognitive-emotional process.
2. **Protection against intrusion.** A user should be protected against manipulative or harmful intervention in that cognitive-emotional process by the AI system, its operator, governmental actors, or other private actors. This protection concerns both information introduced through the interaction and the imprint that repeated use may leave on the user. It reaches intervention in the process of thought itself, not only surveillance of the conditions under which thought occurs.

The framework thus extends beyond a data-privacy problem. Because conversational systems may influence the user while eliciting intimate material, privacy of the mind must address both what leaves the user and what the interaction does to the user. In contexts of emotional processing, protection against intrusion should include mental-health integrity: attention to the psychological conditions under which thought and emotion develop, grounded in the values of freedom of thought without assuming that constitutional doctrine directly governs every private platform.

4.5 Legal Foundations for Privacy of the Mind

The normative starting point for privacy of the mind is freedom of thought, or cognitive liberty. The human mind has traditionally represented a final sphere of privacy: thoughts, emotions, opinions, and inner conflicts generally remain inaccessible unless a person chooses to reveal them. American law does not protect that sphere through a single doctrine. Instead, constitutional law, privacy theory, mental-health law, and therapeutic privilege protect different conditions under which thought can remain free, autonomous, and secure.

4.6 Freedom of Thought

Although the US Constitution does not expressly enumerate a general freedom of thought, First Amendment doctrine has long connected freedom of speech with a broader "freedom of mind" (Wooley v. Maynard, 1977; Yildirim, 2021; American Civil Liberties Union, 2002; Francis and Francis, 2021). The mind is the site where ideas form before they become expression; protecting thought therefore protects a precondition of speech itself. Yet First Amendment doctrine has been claimed to offer only partial protection because it primarily constrains state action and does not necessarily reach nonexpressive surveillance or the conduct of private platforms. This limitation has led scholars to connect freedom of thought with privacy, which understood as a condition for autonomous and independent thought (Boire, 2004; Yildirim, 2021). Related accounts locate that connection in protections for cognitive information flows, neurodata, and records of intellectual activity (Marks, 2024, 475–484) (Unger, 2023, 1484–1497) (Richards, 2008, 393–407).

The Fourth and Fifth Amendments provide complementary protections against governmental access and compulsion. The Fifth Amendment, in particular, has been described as protecting a person from being "compelled to disclose the contents of his own mind" (Curcio v. United States, 1957, 128), although existing doctrine distinguishes compelled testimony from voluntarily created or preexisting records (Fisher v. United States, 1976).

4.7 Intellectual Privacy

Neil Richards's theory of intellectual privacy develops this connection by protecting the conditions under which individuals can think, read, search, inquire, and communicate without fear of monitoring (Richards, 2015). Its concern is therefore not only information produced by the mind, but also the cognitive processes through which ideas and beliefs are formed. Surveillance can alter those processes by encouraging conformity and deterring sensitive or unconventional inquiry. Intellectual privacy thus protects both individual autonomy and the social conditions necessary for free expression. This account is especially relevant to AI systems that do more than receive a completed

query: through iterative responses, they may participate in the process by which the user's question, interpretation, or emotional understanding develops.

4.8 Mental Privacy and Cognitive Liberty

More recent theories of mental privacy, neuroprivacy, and cognitive liberty extend these concerns in response to technologies capable of revealing or influencing mental states. They focus both on access to mental information and on individual control over one's mental states, including protection against unauthorized interception, manipulation, or alteration of thought (Boire, 2004; Grimm and Farahany, 2024; Szoszkiewicz and Yuste, 2025). These theories were developed principally in relation to neuroscience and neurotechnology. AI chatbots do not ordinarily read neural activity or directly access unexpressed thought. They nevertheless expose evidence of cognitive and emotional processes as those processes unfold in language, while their responses may influence how the user articulates, reinforces, revises, or develops those processes.

4.9 Mental Health Law and Therapeutic Privilege

Another relevant body of law concerns mental health, particularly where the thoughts at issue arise through emotional processing. Doctrine limiting involuntary confinement and forced medication recognizes important liberty interests in bodily and psychological integrity. Law and professional ethics also impose duties within therapeutic relationships. The psychotherapist-patient privilege, for example, protects confidential communications made in the course of diagnosis or treatment because effective therapy depends on an environment in which a patient can disclose intimate material.

Privacy law provides a related but narrower protection. Under the Health Insurance Portability and Accountability Act, protected health information held by covered entities and their business associates is subject to rules governing access, use, and disclosure (Health Insurance Portability and Accountability Act, 1996). These doctrines do not automatically apply to general-purpose chatbot conversations. Together, however, they recognize that mental and emotional material is unusually intimate, and that vulnerability surrounding disclosure can justify heightened protection.

5 Policy Recommendations

5.1 Privacy Levers Beyond Data Governance

Rather than extending professional confidentiality doctrine wholesale to AI systems, we propose that its underlying concern for the conditions of intimate disclosure should inform multiple, complementary privacy levers. Table 1 distinguishes interventions that shape the interaction from tools that govern data after or during collection and from the institutions that allocate duties and rights.

5.2 Evidence Law and the Externalized Mind

AI chatbot conversations are already being used as evidence in criminal proceedings, including murder cases in which defendants consulted ChatGPT about self-defense (Hammack, 2025) or the disposal of a body (Schneider and Brumfield, 2026). In the Palisades fire prosecution, prosecutors relied not only on practical queries but also on a prompt by the suspected arsonist asking ChatGPT to generate a dystopian image of a burning forest and people fleeing the fire (Ding and Weber, 2026).

These examples illustrate the need for greater caution in treating AI interactions as evidence of intent or planning. Some chatbot exchanges resemble conventional searches for practical information, while others may reflect fantasies, emotional release, or exploratory thought that would previously have remained internal. Because AI externalizes and records parts of users' cognitive and emotional processes, courts should distinguish evidence that meaningfully demonstrates criminal planning from material that merely captures thought in formation (Abelson, 2027).

More broadly, privacy of the mind should inform the development of evidentiary, Fourth Amendment, and Fifth Amendment doctrine as AI increasingly turns internal cognitive processes into accessible digital records (Carpenter v. United States, 2018; Fisher v. United States, 1976).

Level	Illustrative levers	Role and limit
Interaction and system design	Personal-solicitation policy; calibrated emotional mirroring and validation; controls on sycophancy, anthropomorphic cues, and dependency-producing engagement; accurate confidentiality cues	Shapes the conditions under which users externalize inner voice and disclose intimate material. It cannot substitute for protection of information once collected.
Data governance	Data provenance and lineage; purpose limitation; data minimization; retention and deletion controls; access controls; encryption in transit and at rest; differential privacy for appropriate aggregate analysis or model-training uses	Governs the origin, access, use, security, and release of data. These tools do not by themselves address whether a system solicits intimate information, encourages dependency, or creates an appearance of confidentiality.
Legal and institutional governance	Duties of confidentiality and care; restrictions on secondary use and disclosure; user rights and remedies; audits, documentation, and enforcement; sector-specific oversight	Allocates responsibility and makes technical or organizational commitments enforceable. Its requirements must be tailored to context rather than assuming every emotional interaction is therapy or every personal question is impermissible.

Table 1: Complementary privacy levers for AI-mediated emotional processing.

5.3 HIPAA and APA Guidelines as Normative Benchmarks

Under HIPAA, protected health information held by covered entities and their business associates is subject to rules governing disclosure, access, and use (Health Insurance Portability and Accountability Act, 1996). These protections are justified precisely because of the vulnerability inherent in disclosing mental and emotional information.

AI chatbot interactions increasingly generate information functionally similar to mental health data in both sensitivity and psychological significance, yet many general-purpose chatbot interactions fall outside this framework. The American Psychological Association's 2025 health advisory similarly recognizes that general-purpose chatbots and wellness applications are increasingly used for emotional and mental-health support despite limited clinical validation, oversight, and safety protocols. It calls for action by developers, platforms, policymakers, clinicians, researchers, and consumers (American Psychological Association, 2025).

We therefore suggest using both HIPAA and the APA advisory as normative benchmarks for AI-mediated emotional processing: HIPAA for protection of sensitive mental-health-related information, and the APA framework for reducing psychological risks arising from the interaction itself. Systems explicitly designed or marketed to diagnose, treat, or mitigate mental-health conditions may justify stronger dedicated regulation, including consideration of FDA oversight.

6 Conclusion

The emergence of AI-mediated emotional processing invites us to revisit the legal understanding of the "mind" and its protection through legal frameworks concerning freedom of expression, cognitive liberty, privacy, and mental integrity. As emotional processing increasingly occurs through AI, the boundaries of the human mind may require an expansion of existing legal protections, grounded in historical traditions of cognitive autonomy while adapted to contemporary technological realities.

A framework of privacy of the mind—one that protects not only the data generated through emotional processing, but also the conditions under which such processing occurs—offers a path forward. Policymakers, technologists, and legal scholars must collaborate to ensure that the AI systems mediating our most intimate cognitive and emotional work do so in ways consistent with principles of human dignity, autonomy, and mental integrity.

AI USE STATEMENT

In this work, we used generative AI tools primarily to assist with grammar, language editing, and polishing. This paper is a shortened version of a full-length legal paper extensively discussed at a legal workshop in summer 2026, and we also used generative AI to assist in reorganizing and condensing the manuscript to fit the shorter format. We did not use generative AI tools to research sources, develop the paper's ideas or arguments, or identify case studies. We reviewed all AI-assisted work and take full responsibility for the final content of this paper, including any text, claims, or other material produced with the assistance of generative AI.

REFERENCES

Laura Ginsberg Abelson. The multidimensions of AI chatbots as evidence. Forthcoming in 60 U.C. Davis Law Review; SMU Dedman School of Law Legal Studies Research Paper No. 726, 2027. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6394258.

American Civil Liberties Union. Freedom of expression. ACLU, March 2002. URL https://www.aclu.org/documents/freedom-expression. Published March 1, 2002.

American Psychological Association. Health advisory: Use of generative AI chatbots and wellness applications for mental health. American Psychological Association, November 2025. URL https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-chatbots-wellness-apps.

Anthropic. Disempowerment patterns in real-world AI usage. Anthropic Research, January 2026. URL https://www.anthropic.com/research/disempowerment-patterns.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, 2021. doi: 10.1145/3442188.3445922.

Richard Glen Boire. Cognitive liberty and the First Amendment as a weapon in the fight against technological intrusions of the mind. *First Amendment Law Review*, 3:147–159, 2004.

Benjamin E. Buck and Anne Julia Maheux. Psychosis risk and generative artificial intelligence use frequency, motivations, and delusion-like experiences: Cross-sectional survey study. *Journal of Medical Internet Research*, 28:e85038, 2026. doi: 10.2196/85038.

Carpenter v. United States. 585 U.S. 296, 2018. URL https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf.

Hyojin Chin, Hyeonho Song, Gumhee Baek, Mingi Shin, Chani Jung, Meeyoung Cha, Junghoi Choi, and Chiyoung Cha. The potential of chatbots for emotional support and promoting mental well-being in different cultures: Mixed methods study. *Journal of Medical Internet Research*, 25: e51712, 2023. doi: 10.2196/51712.

Andy Clark and David Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998. doi: 10.1111/1467-8284.00096.

Curcio v. United States. 354 U.S. 118, 1957. URL https://www.govinfo.gov/app/details/USREPORTS-354/USREPORTS-354-118.

Gerrit De Vynck and Jeremy B. Merrill. We analyzed 47,000 ChatGPT conversations. Here's what people really use it for. The Washington Post, November 2025. URL https://www.washingtonpost.com/technology/2025/11/12/how-people-use-chatgpt-data/.

Jaimie Ding and Christopher Weber. Prosecutors paint Palisades Fire suspect as a premeditated arsonist in opening statements. Associated Press, June 2026. URL https://apnews.com/article/palisades-fire-rinderknecht-arson-trial-fa41cf531690ce083defa630753bcef9.

Fisher v. United States. 425 U.S. 391, 1976.

Matthew Flathers, Spencer Roux, and John Torous. Beyond artificial intelligence psychosis: A functional typology of large language model-associated psychotic phenomena. *The Lancet Digital Health*, 8(4):100974, 2026. doi: 10.1016/j.landig.2025.100974.

John Gregory Francis and Leslie Francis. Freedom of thought in the United States: The First Amendment, marketplaces of ideas, and the internet. *European Journal of Comparative Law and Governance*, 8(2–3):192–225, 2021. doi: 10.1163/22134514-bja10019.

Paul W. Grimm and Nita A. Farahany. The battle for your brain: A legal scholar's argument for protecting brain data and cognitive liberty. Judicature, Vol. 107, No. 3, 2024. URL https://judicature.duke.edu/articles/the-battle-for-your-brain-a-legal-scholars-argument-for-protecting-brain-data-and-

Laurence Hammack. Roanoke teen who consulted AI before murder gets 25 years in prison. Richmond Times-Dispatch, September 2025. URL https://richmond.com/news/state-regional/crime-courts/article_9c0168de-714f-5b42-9709-55ddefad528d.html. Published September 4, 2025.

Health Insurance Portability and Accountability Act. Pub. L. No. 104-191, 110 Stat. 1936, 1996.

Limitless. Pendant. Product documentation, 2026. URL https://www.limitless.ai/new.

Mason Marks. Cognitive content moderation: Freedom of thought and the First Amendment right to receive subconscious information. *Florida Law Review*, 76:469, 2024. URL https://scholarship.law.ufl.edu/flr/vol76/iss3/1/.

Maxim Massenkoff and Peter McCrory. Labor market impacts of AI: A new measure and early evidence. Anthropic Economic Research, March 2026. URL https://www.anthropic.com/research/labor-market-impacts.

OpenAI. Early methods for studying affective use and emotional well-being on ChatGPT. OpenAI, March 2025. URL https://openai.com/index/affective-use-study/.

OpenAI. Modeling an AI jobs transition. OpenAI Economic Research, April 2026. URL https://openai.com/index/modeling-ai-jobs-transition/.

Søren Dinesen Østergaard. Will generative artificial intelligence chatbots generate delusions in individuals prone to psychosis? *Schizophrenia Bulletin*, 49(6):1418–1419, 2023. doi: 10.1093/schbul/sbad128.

Jason Phang, Michael Lampe, Lama Ahmad, Sandhini Agarwal, Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, and Pattie Maes. Investigating affective use and emotional well-being on ChatGPT, 2025. arXiv:2504.03888.

Neil M. Richards. Intellectual privacy. *Texas Law Review*, 87:387–445, 2008. URL https://openscholarship.wustl.edu/law_scholarship/522/.

Neil M. Richards. *Intellectual Privacy: Rethinking Civil Liberties in the Digital Age*. Oxford University Press, 2015.

Sandbar. Stream: Private voice ring and conversational notes app. Product documentation, 2026. URL https://shop.sandbar.com/products/stream-ring-sizing-kit.

Mike Schneider and Sarah Brumfield. Prosecutors say suspect in missing students' killings asked ChatGPT about disposing of a body. Associated Press, April 2026. URL https://apnews.com/article/missing-graduate-students-florida-bangladesh-8c51e7fabbf62c2c7ca91b7f04883ba6.

Łukasz Szoszkiewicz and Rafael Yuste. Mental privacy: Navigating risks, rights and regulation. *EMBO Reports*, 26:3469–3473, 2025. doi: 10.1038/s44319-025-00505-6.

Wayne Unger. Stay out of my head: Neurodata, privacy, and the First Amendment. *Washington and Lee Law Review*, 80:1439, 2023. URL https://scholarlycommons.law.wlu.edu/wlulr/vol80/iss4/4/.

Riccardo Volpato, Simone Stumpf, and Lisa DeBruine. Generative confidants: How do people experience trust in emotional support from generative AI?, 2026. arXiv:2601.16656.

Wooley v. Maynard. 430 U.S. 705, 1977.

Emine Yildirim. Thou canst not touch the freedom of my mind: The First Amendment of the US constitution may not provide adequate protection for mental autonomy in the face of conspiracy theories (part 1). CiTiP Blog, July 2021. URL https://www.law.kuleuven.be/citip/blog/thou-canst-not-touch-the-freedom-of-my-mind/. Published July 8, 2021.

Maria de Lourdes Zollo. Building Bee at Amazon: How the wearable AI device is evolving. Amazon, 2026. URL https://www.aboutamazon.com/news/devices/bee-amazon-wearable-ai-device-new-features.