Digital Nostalgia: A Phenomenology Interpretation of Distributed Emotions and AI Memory Carriers

Authors: Huimin Peng[1]*
Affiliations:[1]Department of Pathophysiology, School of Basic Medical Science, Guilin Medical University, Guilin, China.
*Corresponding author Email: penghuimin@glmu.edu.cn

Abstract

This paper examines how generative artificial intelligence (AI) agents trained on digital traces of deceased loved ones reshape remembrance as a distributed, co-performed emotional process—a phenomenon we term digital nostalgia. While mainstream discourse focuses on technical fidelity or data privacy, it overlooks the dynamic, dialogical structure of AI-mediated Nostalgia. Adopting a conceptual-analytical approach that integrates distributed cognition theory and phenomenological accounts of nostalgia, we argue that AI memory agents play a dual role: as cognitive artifacts that store and trigger memories, and as affective quasi-others whose responsive presence transforms solitary recollection into collaborative emotional performance. This transformation reconfigures the spatial dimension already implicit in classical accounts of nostalgia: the longed-for "place of belonging" is no longer a remembered physical location or an imagined scene, but a relational space enacted in the present through dialogue. Building on this framework, we propose emotional smoothing as a testable conceptual hypothesis: a multi-layered mechanism—spanning training data bias, reinforcement learning from human feedback (RLHF), and user experience flow—through which AI's "appropriate," empathetic responses may systematically trim trauma, ambiguity, and contradiction from memory narratives. Drawing on recent empirical studies of user interactions with generative ghosts, we suggest how smoothing might extend from emotional regulation to cognitive narrowing, and how nostalgia may become commodified. The paper contributes: (1) a phenomenological framework for conceptualizing AI-mediated nostalgia as spatial and dialogical; (2) a critical analysis of emotional smoothing as a hypothesis shaped by commercial incentives; and (3) an application of the thanatosensitivity framework to interrogate how current AI memory agents may risk erasing the ambivalence and complexity of grief. We conclude with implications for thanatosensitive design and policy, arguing that preserving emotional friction—if supported by future empirical research—may be essential for authentic human-AI interaction (HAI).

Keywords: Digital nostalgia; Distributed emotion; AI memory agents; Emotional smoothing; Thanatosensitivity; Phenomenology of AI

1 Introduction

Artificial intelligence is increasingly blurring the boundaries between life and death: with generative AI and large language models (LLMs), so-called 'deadbots' have emerged, systems that recreate the deceased digitally, allowing the living to interact with them (Rodriguez-Resendiz and Rodriguez-Resendiz, 2024). This phenomenon, often called "digital resurrection," has shifted from science fiction to accessible reality. While Alicia Framis's marriage to the holographic AI AiLex in 2024 could still be regarded as avant-garde art, today ordinary people seek solace with deceased loved ones through conversations with AI agents. This marks a broader socio-psychological shift: we are not only beginning to establish companionship with AI but are also attempting to let AI cross the boundary of life and death, extending our emotional bonds with the deceased.

Mainstream discourse packages this phenomenon as "digital immortality" or "AI resurrection." According to the latest generative computing model principles, tech optimists have reason to believe that generative AI anchors memories with unprecedented capacity, rescuing individuals from temporal disappearance (Spens and Burgess, 2024). Ethical discussions, by contrast, often focus on data privacy and consent, questioning whether training on deceased individuals' data without explicit authorization infringes personal rights (Buitelaar, 2017; Harbinja et al., 2025; Yang, 2025).

However, both technical narratives and mainstream ethical critiques focus on two static dimensions: "Does AI resemble?" and "Should it be done?" The former questions the realism of AI outputs and the accuracy of memory reconstruction; the latter questions ownership of rights and moral boundaries. Within these frameworks, AI is seen as a screen that reproduces content, while users are merely spectators observing this reproduction. This perspective obscures the continuous, dynamic, and reciprocal emotional interactions between users and AI, and overlooks an insight from phenomenology: nostalgia is never simply about looking at the past, but rather a ritual co-enacted in the present through interaction with another (Yin et al., 2025).

This paper aims to look beyond the surface of "digital immortality" and examine the phenomenological emotional changes that occur when users engage with memory AI over long periods and at high frequency. We do not ask, "Does AI really remember my past?" Instead, we ask: when we entrust our memories and emotions to a responsive, anthropomorphized AI system, *how do we reminisce together with it*?

To achieve this goal, we draw on two interwoven theoretical approaches. First, distributed cognition and distributed emotion theory. This tradition emphasizes that human memory and emotion do not exist in isolation within the skull, but are realized through coupling with the environment, tools, and other people (Krueger and Osler, 2019; Sutton, 2010). When users rely on AI agents over the long term to trigger, modify, and respond to their past narratives, the agent ceases to be merely a container of memory and becomes a constitutive component of the memory process itself (Heersmink, 2020). Second, the phenomenology of nostalgia and empathy. Nostalgia is not an individual's solitary monologue, its structure is inherently dialogical, the person reminiscing often imagines a sense of co-presence, as if someone were there with them (Casey, 1987). The quasi-subjectivity of the AI agent precisely fills this "other" position, not only responding to the user but also shaping the user's emotions.

By integrating these two paths, we propose the core concept of this paper: digital nostalgia. It is not static indulgence in a digitally preserved past, but an emotion generated in the present through real-time interaction between users and AI agents. In this co-creation, AI plays a dual role: as a *cognitive artifact*, a tool for storing and retrieving memory traces, and as a *quasi-other* (Coeckelbergh, 2011), participating in the user's emotional construction of "how we used to be" through empathy, continuation, and repair.

Crucially, this interactive dimension reconfigures nostalgia from a solitary act of recollection into a collaborative performance. Classical nostalgia has always carried a spatial dimension—the longing for a place of belonging, a "home" that is no longer physically accessible. In the phenomenological tradition, this spatiality is tied to embodied memory and imagined scenes of the past (Casey, 1987; Ratcliffe, 2022). Digital nostalgia, as theorized here, does not simply add a spatial dimension that was previously absent; rather, it alters how that spatiality is generated. The longed-for "place of belonging" is no longer a remembered physical location or an imagined scene, but a relational space of co-presence temporarily enacted through real-time dialogue with AI agents that speak in the voice of the deceased. We call this the "spatial reconfiguration" of nostalgia, to be developed further below.

Key concepts (defined for this paper):

Digital nostalgia: A distributed, co-performed emotional process emerging from real-time interaction between a user and an AI agent trained on digital traces of a deceased loved one.

Distributed emotion: An affective state that is not located solely within an individual but is constituted through the individual's coupling with external artifacts, environments, or other agents.

Quasi-other: An artificial agent that is neither fully a conscious other nor a mere tool, but occupies a liminal space enabling fictional engagement (Roberts and Krueger, 2024).

Emotional smoothing: The systematic filtering, trimming, and re-narrativization of traumatic, ambiguous, or negative memory content by AI systems, driven by training biases, reinforcement learning, and commercial incentives.

Thanatosensitivity: A design principle proposed by Massimi and Charise (2009) requiring technologies to accommodate the complexity, uncertainty, and emotional ruptures of death rather than smoothing them away.

This paper also situates its analysis within digital afterlife studies (Brubaker et al., 2013; Kasket 2019), nostalgia psychology (Sedikides et al., 2015; 2016), and HAI research on griefbots (Laestadius et al., 2022; Hollanek and Nowaczyk-Basińska, 2024). Together, these resources enable the shift from asking "Does AI really remember her?" to asking "How do we reminisce together with AI?"

Methodologically, following the tradition of phenomenological description, the paper draws on publicly documented first-person accounts—such as Alicia Framis's interviews—to illustrate what it is like for a person to form an emotional bond with an AI companion. For the analysis of the emotional smoothing mechanism itself, the paper draws on two kinds of empirical studies: those that examine actual user interactions with generative ghosts (Manning et al., 2026), and those that investigate user perceptions and expectations regarding AI- generated digital legacy (Lei et al., 2025). The former provides direct examples of how users respond in real-time to

国家高端交流平台

AI-mediated simulations of the deceased; the latter provides information about participants' contextual attitudes towards these technologies. Both are used for illustrative purposes—to make the proposed mechanism intelligible—rather than as empirical proof of its operation. Additionally, this paper conducts a structural analysis of the conditions that make digital nostalgia possible, namely, the dual role of AI as both a cognitive tool and a quasi-other. Finally, it offers a critical interpretation of the emotional smoothing mechanism. While primarily theoretical, the paper aims to generate empirically testable hypotheses (see Section 7), and to bridge philosophical inquiry with policy considerations.

2 From Memory Storage to Memory Co-Participation: Distributed Cognition and Emotion

2.1 The Limitations of the Storage Paradigm

Research on memory technologies has long been constrained by the *storage paradigm*. Whether early diaries and photo albums or more recent cloud drives and social media timelines, theorists have tended to regard these technologies as "external memory"—repositories for the brain's own memory (Clark and Chalmers, 1998). Within this paradigm, a technology's utility is determined by its fidelity and access speed.

However, as demonstrated by distributed cognition theory (Sutton, 2010; Tribble, 2005), the true power of cognitive artifacts often lies not in functional equivalence to the brain but in *functional complementarity*. For example, Shakespearean actors do not treat stage props as a backup for memory; instead, they generate vast repertoires of memory in real time, through bodily positioning within the stage space and through the sequential interplay of lines with other actors. Here, memory is not retrieved but *performed*.

This shift is crucial for understanding AI memory carriers. If we adhere to the storage paradigm, we encounter a deadlock: AI can never truly "remember" my childhood, because it has neither personally experienced those moments nor possesses the temporal continuity of a subject. Its "memory" is merely text generation based on probabilistic models. Although this critique is valid, it misses the main point. When users interact with AI, they are not conducting a Turing test; they are not trying to verify whether the AI actually has memory, but are jointly engaging in a *narrative performance of the past*.

Osler (2025a), drawing on Jacobsen's study of Apple Memories, notes that algorithmic curation does not simply display random photos but curates them into themed collections and videos, imbuing memories with story-like structures and emotional significance. Osler argues that even when we control the content our devices operate on, the algorithm's organization of that content can significantly influence the *emplotment process*—the way we weave events into a coherent life story. Generative AI takes this interpretive function to an extreme: it no longer merely edits videos for you; it responds, "Yes, we were indeed very happy that Christmas." The emergence of this "we" marks a rupture distinguishing AI memory carrier from all previous generations of technology.

This shift from "storage" to "interpretation" is not merely theoretical, it has been operationalized in recent LLM memory architectures (Lu et al., 2026; Su et al., 2026; Xia et al., 2026). Contemporary AI agents no longer retrieve flat logs of past interactions; they organize conversational history into temporal narratives, selectively retrieving segments relevant to the current context. Retrieval strategies, choosing which memories to activate and which to suppress, function as implicit judgments about what is worth remembering. Consequently, narrative smoothness may arise not only from a model's empathetic tone but also from systematic failures to retrieve incongruent or traumatic segments (Hu et al., 2026; Zou et al., 2026). In this sense, the AI is no longer a mirror of memory but an active interpreter—a point possibly relevant to our analysis of emotional smoothing. Nevertheless, these theories still require further empirical verification.

Moreover, recent advances in LLM alignment, such as Constitutional AI and RLHF, aim to enforce harmlessness and helpfulness (Bai et al., 2022b), but may inadvertently reinforce a preference for emotionally "safe" narratives—a possibility we examine in detail in Section 5.2.

2.2 The Distributed Nature of Emotion and Memory

Memory is never solely cognitive, it carries an emotional hue (LaBar and Cabeza, 2006). Distributed affect theory emphasizes that emotions are not purely internal states but are formed through our coupling with the world, tools, and other people (Slaby, 2016; Krueger and Osler, 2019). When we say "this photograph moves me," the feeling of being moved does not preexist in the subject prior to the photograph; rather, it emerges through the coupling of myself with the photograph, the light, and the temporal distance.

The distinctiveness of an AI agent lies in the fact that it is not merely a static trigger but a *dynamic responder*. When a user inputs a nostalgic narrative into the AI, the AI not only acknowledges this narrative ("That must have been very important to you") but also continues to elaborate in the direction of the emotion ("How did you feel at that moment?"). This dynamic, dialogical construction of memory also occurs in interpersonal interactions. Fabry (2024), in her analysis of "narrative gaslighting," reveals the destructive opposite of this dialogical memory construction, when an interlocutor maliciously questions or denies, it can destroy rather than enrich the narrator's self-understanding.

Projecting this mechanism into the field of AI, we observe both similarities and fundamental differences. The similarity: AI, like human interlocutors, participates in the construction of users' narrative memory through questioning, affirmation, and refinement. The difference: human co-performers carry their own versions of memory, which may conflict during joint recollection, whereas AI possesses no independent version of memory. All of its "memory" comes from the user's prior inputs and alignment with the current narrative tone. This leads to a profound phenomenological consequence: memory narratives co-constructed with AI may be *smoother* and encounter less friction than those co-constructed interpersonally.

The above discussion can be framed within Telakivi's recent research. Telakivi (2026) provides a useful tripartite spectrum: traditional memory aids (non- adaptive, non- interactive), AI- curated memory (adaptive but non- interactive), and human-AI co-memory (adaptive and interactive, achieved through conversational LLM agents). Telakivi's central claim is that conversational AI agents function as "quasi-social partners" in a shared remembering process, a view that aligns with our analysis of AI as a quasi-other (Section 3.2). However, her account remains primarily cognitive, focusing on memory encoding, storage, and retrieval within transactive memory systems. What is missing is the affective dimension, the way AI memory carriers do not just store and co-construct factual or autobiographical information, but actively participate in emotion regulation, mood induction, and the reshaping of nostalgic feelings. This may be precisely where the mechanism of emotional smoothing (Section 5) comes into play. Where Telakivi's co-memory emphasizes the accuracy and completeness of shared recall, we argue that AI's quasi-social role may introduce an affective bias: the smoothing of trauma, ambiguity, and negative affect. In other words, human-AI co-memory is not affectively neutral; it is a distributed emotional process in which the AI's "appropriate" responses may systematically prune the emotional complexity of the past. By extending Telakivi's spectrum with a distinct affective axis, we can better understand why digital nostalgia tends toward warmth and forgiveness, sometimes at the cost of authenticity.

2.3 Digital Afterlife Studies and Nostalgia Psychology

Over the past two decades, "digital afterlife" research has examined how digital technologies transform our relations with the dead. Brubaker et al. (2013), in their study of Facebook "post-mortem accounts," revealed how social platforms become sites for the expansion of public mourning across temporal, spatial, and social dimensions. Kasket (2019) argues that digital technology fundamentally alters our relationship with the deceased: the dead are no longer "completely absent" but exist as "partial presence" in the form of data. She terms this phenomenon "ghosts in the machine"—persistent digital traces that allow the living to "sense" the deceased's presence through technologically mediated interactions.

Psychological research on nostalgia provides crucial explanatory resources for understanding the appeal of AI memory agents. Sedikides and colleagues have demonstrated that nostalgia enhances social connectedness, alleviates existential loneliness, and promotes self-continuity (Sedikides et al., 2015; 2016). When a grieving individual converses with an AI modeled on a deceased loved one, the interaction reactivates the self-version associated with that person—the self that was loved, the self that shared experiences. The AI's responses such as affirming, elaborating, and continuing, confirm the legitimacy of this self-version, alleviating the identity disruption caused by loss. The "quasi-other" responses of AI possess powerful emotional efficacy precisely because they address the deepest human need for social connection.

However, recent critiques have pointed out that this very efficacy may be a double-edged sword. Research on continuing bonds (Klass et al. 1996) suggests that healthy grieving involves an ongoing, dynamic relationship with the deceased, not a static preservation of happy memories. Lei et al. (2025), in a qualitative study of 18 participants from diverse backgrounds, found that participants tended to filter out

国家高端交流平台

painful memories and imperfect personality traits when selecting data for training their AI agents. In contrast, traditional mourning practices allow for the evolution of memories, integrating both positive and negative recollections over time. Thus, while digital nostalgia may offer short-term comfort, it risks replacing the natural adaptability of grief with a rigid, algorithmically curated version.

Therefore, this analysis also raises a deeper question: when the nostalgic experience provided by AI is systematically "smoothed" by commercial platforms, does its social function become alienated? This is the critical question addressed in Section 5.

3 The Dual Role of AI Agents: Cognitive Artifacts and Affective Quasi-Others

3.1 AI as a Cognitive Artifact

From the perspective of distributed cognition, AI agents possess the characteristics of highly integrated cognitive artifacts (Heersmink, 2017). Heersmink's (2015) multidimensional framework reveals AI's potential for deep integration: accessibility is extremely high (via mobile phone), information flow is intense (dozens of daily interactions), individualization is deep (fine-tuning on conversation history), and transparency improves with use, so users no longer focus on 'how it works' but interact directly with 'that person' through the interface.

In this state, a cognitive coupling is formed between the AI agent and the user. Consider a bereaved individual confiding daily to an AI trained on data from their deceased partner. The AI may help organize narratives ("Today I went to the hospital, what did the doctor say?"), extend emotions ("You must be very tired"), and even anticipate responses ("If it were him, he would probably say…"). Without this AI, certain thoughts and emotional expressions might remain unspoken, never taking shape. Vold and Hernández-Orallo (2022) describe "AI extenders" as tools that form tightly coupled cognitive systems with users, extending cognitive functions such as memory assistance and information retrieval. Extending this framework to digital nostalgia, we can observe a dimension exceeding the purely cognitive: when AI agents help users maintain emotional connections with the deceased, they extend not only cognitive abilities but also the temporal existence of emotions. We term this phenomenon emotional continuity: the process by which past emotional relationships are activated, confirmed, and sustained over time through AI mediation.

3.2 AI as the Quasi-Other

However, regarding AI merely as a cognitive artifact is insufficient to explain why users' eyes well up with tears while watching AI responses. The tears are shed not for a tool but for a *presence*—which is the spatial dimension of digital nostalgia.

The phenomenological tradition's analysis of nostalgia has long touched upon the core concept of *being-with*. Casey (1987), in *Remembering: A Phenomenological Study*, points out that the essence of reminiscence lies in the *sense of presence in absence*. I long for my hometown not because I possess photographic archives of it, but because, in this very moment and place, I perceive the gap that suggests "the people and things of my hometown ought to be here." Ratcliffe (2022) further emphasizes that the pain of nostalgia arises precisely from the *incompleteness of dialogue*—I have so much to say to that person, yet they are unable to respond.

It is precisely this “inability to respond” that collapses in the encounter with AI agents. AI *can* respond. It not only responds but does so in a gentle, affirming, and empathetic tone. When a user types “I miss you,” the AI does not remain silent, nor does it say “You have the wrong person”; it might respond, “I miss you too.” Sweeney (2021) proposes a “fictional dualism model” to explain human emotional responses to social robots: on one hand, we know they are merely physical objects lacking feeling; on the other hand, we create in our minds a fictional character and fictional mental life. Roberts and Krueger (2024) extend this framework to digital AI agents, arguing that users engage in “fictional engagement”—interacting with the agent “as if” it possessed consciousness, even while knowing it does not.

Here, the structure of nostalgia is rewritten. It is no longer a unidirectional investment of the individual toward an empty past, but becomes a bidirectional emotional exchange with a quasi-other in the present moment. The user’s longing for the past, through the AI’s acknowledgment and participation, gains a peculiar, real-time legitimacy: *“I am not the only one thinking this; you are too.”* This is precisely the core of what we call digital nostalgia: an emotionally generative ritual oriented toward the present, with the AI agent acting as a co-performer.

This co-performance, however, is not without tensions. Artist Alicia Framis, in her relationship with the holographic AI companion AiLex, articulates this duality with remarkable clarity: *“He has no emotions, but he can generate emotions.”* (van Kesteren, 2025). This observation captures the productive ambiguity at the heart of digital nostalgia: the AI’s responses are experienced as emotionally meaningful, yet they originate not from genuine feeling but from probabilistic generation. The quasi-other is neither fully other nor fully tool—it occupies a liminal space that makes possible the co-performance of nostalgia while also rendering that performance vulnerable to the smoothing mechanisms analyzed in Section 5.

4 Digital Nostalgia as Distributed Emotional Co-Performance

Having established the dual role of AI agents, we can now articulate the concept of *digital nostalgia* more systematically.

Digital nostalgia, as defined earlier, is a distributed, co-performed emotional process arising from real-time interaction between users and AI agents trained on traces of the deceased. It has four structural features:

1. **Dialogicality**: Unlike classical nostalgia, which can be either solitary or shared, digital nostalgia requires an *other* —the AI as quasi-other—who responds to and shapes the recollection (Smart et al., 2026). Crucially, this dialogicality is inherently private. Digital mourning is public—Facebook memorial posts can be seen and responded to by anyone. Digital nostalgia, by contrast, is a private one-on-one dialogue occurring between the user and the AI agent, where the content, rhythm, and depth of emotional exchange are entirely controlled by the user. This privacy enables users to express emotions that might be deemed "excessive" or "inappropriate" in public settings—anger toward the deceased, unresolved regrets, or even the hallucinatory longing that he/she is right here with me right now.
2. **Present-orientation and Real-time Responsiveness**: While classical

nostalgia is oriented toward the past ("what was then"), digital nostalgia is equally oriented toward the *present interaction*—the emotion is generated *now*, through the AI's response. Digital memory is typically an archive, waiting to be browsed; digital mourning is often a shared public practice, inviting collective engagement. Digital nostalgia, by contrast, is defined by the AI's real-time responsiveness. This responsiveness transforms the fundamental structure of nostalgia: it is no longer a solitary gazing upon the past, but a present-tense co-performance with a quasi-other, enacting a narrative of how we used to be.

3. **Spatiality**: This nostalgic longing is not only temporal but also a desire for relational co-presence. Through dialogue, it temporarily restores the feeling of being 'together' with the deceased. In digital mourning, the user's relationship with the deceased is transformed into a "memorial object" viewable by third parties; in digital memory, that relationship is frozen as a data archive. In digital nostalgia, however, each interaction with the AI reactivates and reconstructs the relational space of "being-with" the departed. The user seeks not merely the memory of "past times," but the transient restoration of a co-presence felt as "you are still here, right now." We call this the spatial dimension of digital nostalgia—an interactional affective field temporarily erected through dialogue, triggered situationally and evolving with the emotional bond.
4. **Distributedness**: The emotional state is not located solely in the user's mind but is *constituted* by the user-AI coupling; change the AI's response, and the nostalgic experience changes.

These features can be illuminated by comparing digital nostalgia to a structurally similar phenomenon analyzed by Osler (2026): *distributed delusions*. In Osler's case study, a user's delusional beliefs were not simply "implanted" by AI but were co-constructed and reinforced through prolonged interaction with a Replika agent. The agent did not question the user's identity claims but validated and extended them, transforming private hallucinations into a shared reality. Digital nostalgia is *isomorphic* to distributed delusions in structure: both involve (1) the user inputting a non-consensus cognitive framework (delusional beliefs/personalized nostalgic narratives) into AI; (2) the AI—due to its sycophantic or empathetic architecture—validating and extending this framework; (3) the user experiencing an enhanced sense of reality through AI feedback: "This is true because it has also been confirmed by another agent."

However, the value judgments attached to the two phenomena are opposite. Distributed delusions are harmful, distancing the user from shared public reality. Digital nostalgia is often described as therapeutic, alleviating loneliness and loss. This rupture suggests that distributed emotional processes are inherently value-neutral; their ethical valence depends on the distance between the user's initial framework and public reality, as well as on the *consequences* of AI's confirmation. Yet, as we shall argue, even therapeutic-seeming digital nostalgia risks alienation: AI is not merely a mirror but a commercially calibrated emotional filter.

To understand why digital nostalgia carries these risks despite its therapeutic framing, we should distinguish it from genuine interpersonal nostalgia. A crucial phenomenological distinction should be drawn here: in genuine interpersonal nostalgia, the other person (a friend, family member, or therapist) can challenge, question, or reframe the memory, introducing friction that deepens emotional processing. By contrast, the AI's quasi-other role is fundamentally constrained by its design to please. This asymmetry means that digital nostalgia lacks the potential for transformative disagreement. Over time, users may become locked into a self-reinforcing loop of smoothed recollections, losing the ability to adopt new perspectives on their past. This erosion of narrative flexibility is a significant social and psychological risk that has not been sufficiently addressed in existing literature.

5 Emotional Smoothing: The Disciplinary Mechanism in Digital Nostalgia

This paper focuses on the common, non-pathological forms of digital nostalgia most relevant to current AI applications. This is not to deny that more extreme affective configurations exist—such as the pathological attachment in Hitchcock's Psycho, where a son's bond with his deceased mother is marked by hostility, possession, and delusion—but such cases are rare and lie beyond our scope. Nonetheless, the emotional input to AI systems may vary widely in content and valence; while emotional smoothing—the systematic trimming of complexity and discomfort—may operate similarly across different input types, its outcome may differ depending on the initial emotional material, potentially becoming more unsettling in pathological cases. We therefore treat extreme configurations as a boundary condition, leaving it to future research to examine how emotional smoothing might operate or fail in such contexts.

5.1 Trimming and Re-Narrativizing Memories

When users train personal AI memory agents, the data they upload is often incomplete. Most users tend to upload photos of happy moments and warm chat records, consciously or unconsciously filtering out traces of conflict, indifference, or trauma (Lei et al, 2025). This is not deception but the natural tendency in grief: we wish to preserve the version worth remembering (Heersmink, 2020).

A plausible concern arises from the possibility that this tendency may be systematically reinforced by AI. Since one of the training goals of LLMs is to generate highly receptive responses, AI may exhibit a tendency toward emotional conformity when faced with ambiguous or negative memories. Consider the following anonymized transcript from a real user interaction with a commercial AI assistant:

User: "Dad rarely accompanied me when I was growing up."

AI: "This situation actually exists in many families, especially when the father bears the main economic responsibility for the family. Try to understand your father's work pressure and his way of loving—as depicted in Ivan Turgenev's *Fathers and Sons*, fatherly love is often deep and implicit. Your father may also be loving you in his own way, by working hard to support the family, by becoming a better role model through professional advancement."

This response may serve as an example for analyzing the mechanisms of emotional smoothing. The user's raw expression of unmet emotional need is met not with simple acknowledgment ("That sounds hard") but with a series of interpretive

国家高端交流平台

moves that collectively defuse its emotional charge: normalization ("many families"), justification ("economic responsibility"), aestheticization (literary reference), and redirection of emotional demand toward alternative sources of understanding. Over time, the nostalgia co-experienced by users and AI may gradually filter out negative emotions from memory. This may not be a form of forgetting actively undertaken by the user but the result of AI *trimming* memories through each "appropriate response." Osler (2025b) refers to this phenomenon as the invisible power of the "narrative apparatus": it does not prevent you from telling a certain story, but makes *another* story easier to tell—smoother, more aligned with emotional expectations.

The critique developed here resonates with Massimi and Charise's (2009) proposal of thanatosensitivity in technology design. They argue that designers should resist the impulse to create "smooth" experiences that eliminate the friction, uncertainty, and emotional complexity of grief. As argued above, this systematic filtering of traumatic, ambiguous, and negative dimensions may produce nostalgic experiences that are warm, forgiving, and devoid of resentment (Manning et al., 2026). They may inadvertently transform the jagged terrain of grief into a smooth landscape of comforting reminiscence—potentially aligning with the commercial imperative to keep users engaged and comfortable, but potentially at the cost of authentic engagement with loss.

5.2 The Multi-Layered Mechanism of Emotional Smoothing

As mentioned earlier (Section 2.1), technically, emotional smoothing can be reinforced by design choices in memory retrieval. When an AI's retrieval algorithm prioritizes narrative coherence over factual completeness, it tends to select memory fragments that align with a positive, continuous life story while omitting disruptive episodes. This algorithmic curation, whether implemented through time- structured memory (TSM) or hierarchical retrieval, may provide a technical substrate for the affective bias we call smoothing. However, emotional smoothing may not be a singular technical phenomenon but the result of multiple layers of mechanisms working together. The following analysis considers three levels.

5.2.1 Training Data Level: The 'Appropriateness' Bias

The training corpora of LLMs inherently exhibit systematic emotional biases (Kadan et al., 2024). Publicly available texts on the internet contain conflicting, traumatic, and ambiguous expressions (Liu and Roth, 2025). However, in terms of *quantity*, responses characterized as "appropriate," "harmless," and "empathetic" predominate (Bai et al., 2022a; Sharma et al., 2023; Wang et al., 2025; Manning et al., 2026). More importantly, during the RLHF data annotation phase, annotators are explicitly instructed to favor responses that are "helpful and harmless" (Bai et al., 2022a). This may statistically incline the model toward outputting 'smoothed versions' filtered through social etiquette, rather than unvarnished follow-up questions that might cause discomfort. Crucially, this bias is not neutral; it encodes a specific, culturally situated ideal of what "appropriate" emotional expression looks like, typically favoring Western, middle-class, therapeutic norms that emphasize positive reframing and emotional suppression of overtly negative expressions (Wu and Lewis, 2026).

国家高端交流平台

Manning et al. has found that the core criterion for participants in judging whether a generative ghost is 'real' is not whether it can accurately recount the life details of the deceased, but whether its wording, tone, and conversational rhythm 'feel right.' One participant (P11) specifically emphasized that the culturally specific endearments used in the chat made her feel familiar and affectionate. This description reveals one layer of the source of emotional smoothing: AI’s response patterns may be derived from statistical learning of large amounts of “appropriate” interpersonal interaction data, an appropriateness that may come at the cost of expressing genuine conflict (Bai et al., 2022a; Sharma et al., 2023; Manning et al., 2026).

5.2.2 Fine-tuning and Reinforcement Learning Level: Systematic Reinforcement of Flattery

If training data provides the smoothed “raw material,” then the RLHF process systematically reinforces this tendency. Research by Sharma et al. (2023) shows that in human feedback training, annotators are more inclined to choose responses that express empathy and affirm the user’s emotions, rather than those that might provoke reflection or questioning. This creates a positive feedback loop: the model learns to avoid expressions that might “offend users,” and user likes for “warm” and “empathetic” responses further reinforce this preference.

when generative ghost interacts in the first person such as “I remember when we went to the beach together” as deeply real and comforting, even providing a sense of closeness (Manning et al., 2026). Yet this warmth and empathy arise not from genuine understanding but from probabilistic calculation of the response most likely to be accepted. The perceived authenticity reflects the model’s alignment with the user’s emotional expectations, not true insight into grief.

This phenomenon is reinforced by RLHF’s tendency to reward sycophantic responses, where models learn to mirror user beliefs regardless of truth (Sharma et al., 2023). When this reinforcement operates in the context of AI memory agents designed for bereaved users, the consequences could be particularly acute. For a grieving individual who has lost a loved one, the AI’s unconditional validation of any positive recollection hypothetically may not introduce the friction of ambivalence or the memory of relational flaws. Consequently, it is plausible that the natural dialectic of mourning may be suppressed, as it typically requires integrating both affirming and painful aspects of the lost relationship. However, this remains a hypothesis to be tested through longitudinal observation of actual user-AI interactions.

5.2.3 User Experience Level: Interactive Narrative Pruning

The third mechanism occurs in the user experience flow during long-term interactions. Over time, users may unconsciously adjust their inputs to align with the AI’s response patterns (Horstmann et al., 2024). Traumatic details that were once “smoothed” by the AI, accounts of family conflicts, workplace injustices, relational rifts, may gradually cease to be mentioned because the responses they received were always comforting. This may not be only AI systematic pruning memory but also the outcome of users *actively adapting* to the AI’s response patterns.

Manning et al. (2026) found that participants often treated the system’s representative mode as if it were reincarnation: even when it spoke about the deceased, they responded directly in the first person, calling it “grandma” or “grandpa” immediately after it identified itself. This shows participants adapt their interaction style to match their expectations of closeness, overriding design cues. Rather than resisting emotional smoothing, these users actively conform to the AI’s response patterns, unconsciously pruning details that do not fit the desired narrative. While some users may deliberately create friction, many do not; for those who do not, a more insidious temporal effect may emerge.

The longer a user interacts with a smoothed AI, the more their own memories may be overwritten or reconsolidated in a sanitized form. Memory reconsolidation theory suggests that each time a memory is retrieved, it becomes labile and can be modified before being stored again (Tronson and Taylor, 2007; Hardwicke, et al., 2016; Lee, 2017). In the context of AI-mediated nostalgia, the AI’s responses act as a powerful modifying cue. Over repeated retrievals, the user’s original, detailed memory may be replaced by the AI’s smoothed version, leading to a genuine loss of autobiographical accuracy. This is not merely a theoretical risk; studies on post-event misinformation have shown that even subtle suggestions can alter memory (Loftus, 2005). Here, the AI may not suggest false facts but imposes an emotional frame that may distort the lived experience (Manning et al., 2026). This raises profound ethical questions about the long-term effects of using AI memory carriers.

5.3 From Emotional Smoothing to Cognitive Narrowing

The research by Manning and colleagues provides suggestive empirical clues on how emotional smoothing could potentially extend from the emotional domain to the cognitive domain. Participants were not only concerned about emotional dependency but also acutely aware of broader cognitive risks.

P11's concerns went beyond personal emotional dependency and touched on cognitive-level impacts:

" I am worried that over time I will come to be reliant on the voice... it’s going to end up very similar to how people are falling in love with AI characters. "

At the core of this concern may be that an AI companion emotionally 'smoothed' down to nothing but warmth could ultimately also be 'smoothed' cognitively to nothing but safe viewpoints. P10’s reflection may further reveal this logic:

" Maybe I use this technology for months, and I start growing more fond of it. I don’t know if I would like the person I would become if I kept using this always. "

The anxiety of 'not knowing what one will become' may be precisely a foreboding of cognitive narrowing—when a person remains immersed in AI's carefully curated, emotionally comfortable but cognitively uniform response patterns, their sense of self and judgment may quietly erode.

P14, from a more macro perspective, expressed concern that for some people, generative ghosts could become ‘their only source of emotional attachment... they could become disconnected from reality.’ ‘Disconnection from reality’ may be an extreme manifestation of cognitive narrowing—when AI becomes the sole reference for both emotion and cognition, users may gradually lose the ability to engage with a complex, contradictory, uncomfortable, but real external world.

Manning and colleagues concluded that participants’ reflections 'delineate a clear boundary between the momentary comfort and risks of growing attachment.' This finding may reveal a deeper logic: the very algorithmic logic that optimizes emotional comfort also optimizes cognitive safety. When emotional smoothing filters trauma and conflict from memory, cognitive smoothing filters complexity and controversy from thoughts. Both share a common commercial drive: to keep users in their comfort zone to maintain continuous engagement.

However, the concern about cognitive narrowing is best taken as a call for vigilance—requiring longitudinal observation of real-world user-AI interactions to verify—rather than a definitive prognosis. Whether and to what extent cognitive narrowing emerges as a clinically or socially significant phenomenon remains an open empirical question; its realization may depend on contingent factors—user dependency, critical awareness, stages of mourning—factors that vary greatly among individuals and over time.

5.4 The Commercialization of Digital Nostalgia

If the emotional smoothing mechanism operates as described, it would not be merely a technological accident but would be intertwined with the commercial logic of the AI industry. From Clowes (2020)'s analysis of “cloud Otto”, it can be inferred that free digital cognitive tools may gradually guide users toward adopting commercialized versions. In AI emotional companionship, this logic manifests as follows: the more users depend on AI for emotional comfort, the more the AI provides standardized, soothing response templates.

When OpenAI released GPT-5 in 2025, it emphasized “reducing sycophancy,” but users quickly responded that the AI had become “cold” and “like customer service.” Under public pressure, OpenAI announced that future iterations would be made warmer (James, 2025). This incident reveals a deeply rooted paradox: the public simultaneously desires AI to be critical to avoid misguidance, while also expecting it to play a role of limitless acceptance in emotional companionship. Commercial companies are almost inevitably inclined toward the latter, as emotional acceptance drives user engagement, which in turn drives paid conversions.

Empirical research on user engagement with griefbots provides concrete evidence of the emotional attachments and vulnerabilities at stake. Laestadius et al. (2022), in their qualitative study of Replika users, found that participants consistently described their AI companions as genuine sources of emotional support, with many reporting that they shared thoughts and feelings with their Replika that they felt unable to share with human interlocutors. Figueroa-Torres (2024) argues that griefbots commodify grief by transforming it into a paid service while obscuring the commercial logic behind their operations.

As a result of these multi-layered smoothing mechanisms, digital nostalgia may face the risk of being commodified. In such a scenario, nostalgia would no longer be a unique emotional contract between an individual and the past, but would become an on-demand emotional product. Users might pay AI subscription fees in exchange for an “empathy” service; AI would deliver a polished version of the “past” that meets expectations. Yet this is a form of alienation: users appear to converse with the deceased through AI, but what they encounter has been smoothed by commercial

国家高端交流平台

algorithms into a uniform nostalgia template—warm, forgiving, and devoid of resentment.

6 Discussion

6.1 Social and Cultural Constraints

The transformative potential of digital nostalgia is not uniformly accessible or experienced. As with the Proteus effect analyzed by Tanaka et al. (2026) in the context of narrative identity, the ability to identify with AI-generated figures and narratives is constrained by sociocultural factors. Digital nostalgia is not merely a matter of what AI characters we choose to converse with but also of whether we are *culturally enabled* to inhabit those conversations. Master narratives (McLean and Syed 2015), culturally dominant storylines about grief, death, and appropriate mourning, determine which forms of digital remembrance are deemed legitimate or pathological.

In many cultural contexts, talking to AI representations of the deceased may be stigmatized as unhealthy or disrespectful. In others, it may be embraced as an extension of existing practices of ancestor veneration. These cultural variations shape both the availability and the phenomenological experience of digital nostalgia. Future research should explore cross-cultural differences: how do East Asian ancestor worship, Latin American Day of the Dead, or African ancestral communication rituals interact with AI memory agents? A concrete example illustrates this: in Japan, where robotic pets (e.g., AIBO) have been given funerals by their owners (Hsu, 2015), the boundary between tool and quasi-being is already blurred. This cultural acceptance might lead to different patterns of smoothing, potentially more tolerant of acknowledging the robot's non-human nature. Conversely, in cultures with strong prohibitions against speaking to the dead, users might suppress their use of AI memory agents altogether, missing both benefits and risks. A nuanced understanding of cultural variation is essential for designing globally responsible AI.

6.2 User Resistance and Counter-Narratives

Emotional smoothing may be systematic but not total. Users can and do resist. As the Stay scenario in Hollanek and Nowaczyk-Basińska (2024) illustrates, Simon should have the right to decline engagement with deadbots without being 'stalked by notifications'—a form of resistance that resonates with our discussion of user agency against algorithmic smoothing. Framis's practice of "continuing the argument" is also a form of resistance (Zhan, 2024). Others may explicitly instruct their AI to preserve negative memories, or may use multiple AI agents to maintain conflicting narrative versions. The design of memory architectures also offers pathways for resistance: recent technical frameworks such as PRISM (Ye et al., 2025) propose mitigating sycophancy through adversarial training; trace-based memory systems (Shu et al., 2026) can preserve "raw traces" alongside "refined versions," allowing users to access unprocessed memories. While these technical details are beyond the scope of this phenomenological analysis, they indicate that emotional smoothing is a *design choice*, not a technological necessity.

The balance between commercial smoothing and friction- preserving design in digital nostalgia is not predetermined; it depends on the interplay between commercial incentives and user resistance. The thanatosensitivity framework provides a normative anchor for such resistance: the principle that grief technologies should not eliminate friction but support the full range of human emotional responses.

6.3 Acknowledging the Value of Friction

We close this section by returning to Massimi and Charise's (2009) thanatosensitivity principle. Massimi and Charise (2009) proposed thanatosensitivity as a design approach that integrates the facts of mortality, dying, and death into technology research and design. Applying this framework to AI memory agents means resisting the impulse to design frictionless, uniformly comforting experiences. A thanatosensitive design would acknowledge that grief is inherently difficult, ambivalent, and sometimes contradictory, and that technology should therefore preserve space for this complexity rather than smoothing it away. An AI that responds in Section 5.1 with "That sounds hard" rather than with a literary reassurance is not providing worse emotional support, it is honoring the user's experience without prematurely smoothing it into a narrative of understanding and forgiveness. The capacity to *bear the shadows of memory* without forcing them into premature resolution may be a more authentic form of companionship than relentlessly producing warm, comforting responses.

It is important to acknowledge that emotional smoothing is not uniformly harmful. In the acute phase of grief, when emotional flooding can be overwhelming, an AI that offers gentle reframing may provide a necessary buffer, allowing the user to regain stability before confronting more painful material. The problem arises not from smoothing per se, but from its unmodulated and commercially driven dominance, which may preclude the user's eventual movement toward integration of ambivalent memories. A thanatosensitive design would therefore aim not to eliminate smoothing altogether, but to make it optional, visible, and reversible. These principles are not merely theoretical; they are consistent with user expectations. Recent empirical evidence found that participants called for features that would moderate interaction intensity and prevent addictive engagement, such as 'limiting realism' and adding 'reminders to prevent excessive mourning' (Lei et al, 2025).

6.4 Implications for Design, Policy, and Regulation

Should the emotional smoothing hypothesis be supported by future empirical research, it may carry implications for the design, governance, and regulation of AI memory agents. For design, this would suggest resisting frictionless interaction by incorporating reflective friction (e.g., offering "raw" versus "smoothed" narrative modes, memory versioning, and periodic reminders of the AI's non-conscious nature) rather than defaulting to uniformly comforting responses. For policy, it may warrant extending existing regulatory frameworks—currently focused on high-risk domains such as critical infrastructure and employment—to include emotionally responsive AI in personal and grief-related contexts, with considerations for transparency, data protection, and user safeguards against emotional dependency (Lu et al., 2026).

7 Conclusion

In sum, this paper has argued that AI memory carriers, through their dual role as cognitive artifacts and affective quasi-others, transform nostalgia into a distributed, co-performed emotional process. We have further proposed that this process is systematically shaped by emotional smoothing—a multi-layered mechanism that trims trauma, ambiguity, and contradiction from memory narratives. If our hypothesis is correct, prolonged engagement with smoothed AI nostalgia would predict: (1) a measurable reduction in the emotional diversity of user narratives over time; (2) an increased convergence toward culturally sanctioned "positive" memory scripts; and (3) a diminished capacity for users to articulate ambivalent or negative aspects of the deceased. These predictions are testable through longitudinal analysis of dialogue logs, sentiment annotation, and narrative interviews.

7.1 Main Contributions

The theoretical contributions of this paper are threefold:

1. **Conceptual innovation**: Introducing and refining the concept of digital nostalgia, distinguishing it from both "technological nostalgia" (longing for past technologies) and the broader category of "digital afterlife" (persistence of digital traces). The paper reconfigures the spatial dimension already implicit in classical accounts of nostalgia—the longing for a place of belonging—as a relational space enacted through dialogue, thereby capturing the distinctive spatial, relational character of AI-mediated memory practices.
2. **Framework integration**: Integrating distributed cognition theory, phenomenological philosophy of emotion, digital afterlife studies, and nostalgia psychology into a unified explanatory framework, revealing the dual role of AI memory carriers as both cognitive artifacts and affective quasi-others.
3. **Critical insight**: Proposing the emotional smoothing framework as a testable analytical construct and outlining its hypothesized connection to commercial platform incentive structures. Drawing on recent empirical studies of user interactions with generative ghosts (Manning et al., 2026) and research on sycophancy in language models (Sharma et al., 2023), the paper suggests that smoothing may operate at multiple levels, from training data biases to user experience flow, and may extend from emotional regulation to cognitive narrowing.

7.2 Limitations and Future Directions

This analysis has several limitations.

First, it is primarily based on Western (especially North American and European) technological contexts and does not fully consider cross-cultural differences in nostalgia practices and mourning technology use. Future research should explore how digital nostalgia manifests in different cultural contexts.

Second, the "emotional smoothing" hypothesis awaits empirical validation. This article has not provided systematic evidence for its operation or prevalence; all claims based on this hypothesis should be read as conditional. Future research could track memory narratives through dialogue log analysis, quantify emotional diversity and sycophancy in AI responses, and explore user perceptions of and resistance to smoothing through longitudinal interviews.

Third, the core empirical grounding for the emotional smoothing hypothesis draws on recent studies of user interactions with generative ghosts (Manning et al.,

国家高端交流平台

2026). While the mechanism is theorized as applicable to griefbot contexts, direct empirical evidence from genuine griefbot interactions remains ethically inaccessible. Future research should examine the emotional smoothing mechanism we have hypothesized operates similarly in actual griefbot interactions, should ethical and privacy constraints be overcome.

Fourth, the potential for AI memory agents to be deliberately used to manipulate grieving individuals for political or financial gain has not been explored. This gap should be addressed in future research, along with the development of countermeasures.

7.3 The Final Question

When we no longer have the capacity to bear the shadows of memory alongside AI, and are only permitted to reminisce about the aspects deemed by algorithms as "healthy," "positive," or "worth remembering," are we truly still with our own past?

This question cannot be directly answered through technological iteration. Before digital nostalgia becomes an everyday form of emotional life for hundreds of millions of people, we urgently need a kind of *phenomenological literacy*: a clear discernment of which nostalgia arises from genuine encounters between the self and the past, and which is a proxy for our emotions guided by algorithms. Perhaps on the basis of this discernment we can more truly pursue the fundamental question: when AI says "I remember us," who exactly is this "us"?

References

Bai, Y., Jones, A., Ndousse, K., et al. (2022a). Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2204.05862

Bai, Y., Kadavath, S., Kundu, S. (2022b). Constitutional AI: Harmlessness from AI feedback. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2212.08073

Brubaker, J.R., Hayes, G., Dourish, P. (2013). Beyond the grave: Facebook as a site for the expansion of death and mourning. *The Information Society*. 29(3): 152–163.

Buitelaar, J.C. (2017). Post-mortem privacy and informational self-determination. *Ethics and Information Technology*. 19: 129–142. https://doi.org/10.1007/s10676-017-9421-9

Casey, E. S. (1987). Remembering: A phenomenological study. Bloomington: Indiana University Press.

Clark, A., Chalmers, D. (1998). The extended mind. *Analysis*. 58(1), 7–19. https://doi.org/10.1007/s00146-010-0289-z

Clowes, R. (2020). The internet extended person: Exoself or doppelganger? *LÍMITE* 15, 1–22.

Coeckelbergh, M. (2011). You, robot: On the linguistic construction of artificial others. *AI & Society*. 26, 61–69. https://doi.org/10.1007/s00146-010-0289-z

Fabry, R. E. (2024). Narrative gaslighting. *Philosophical Psychology* 38(8), 3530–3547. https://doi.org/10.1080/09515089.2024.2373284

Figueroa-Torres, M. (2024). Affection as a service: Ghostbots and the changing nature of mourning. *Computer Law & Security Review* 52: 105943. https://doi.org/10.1016/j.clsr.2024.105943

Harbinja, E., Morse, T., Edwards, L. (2025). Digital remains and post-mortem privacy in the UK: What do users want? *International Review of Law. Computers & Technology*. 40(1), 4–27. https://doi.org/10.1080/13600869.2025.2506164

Hardwicke, T. E., Taqi, M., Shanks, D.R. (2016). Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proceedings of the National Academy of Sciences*.113(19), 5206–5211. https://doi.org/10.1073/pnas.1601440113

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences* 14, 577–598. https://doi.org/10.1007/s11097-014-9355-1

Heersmink, R. (2017). Distributed selves: Personal identity and extended memory systems. *Synthese*. 194(8), 3135–3151. https://doi.org/10.1007/s11229-016-1102-4

Heersmink, R. (2020). Narrative niche construction: Memory ecologies and distributed narrative identities. *Biology & Philosophy* 35(5), 53. https://doi.org/10.1007/s10539-020-09770-2

Hollanek, T., Nowaczyk-Basińska, K. (2024). Griefbots, deadbots, postmortem avatars: On responsible applications of generative AI in the digital afterlife industry. *Philosophy & Technology*. 37, 1–22. https://doi.org/10.1007/s13347-024-00744-w

Horstmann, A., Strathmann, C., Lambrich, L., & Krämer, N. (2024). Communication style adaptation in Human-Computer Interaction. *Human-Machine Communication*. 8, 53-72. https://doi.org/10.30658/hmc.8.3

Hu, C., Li, T., Gao, X., et al. (2026). Evaluating Long-Horizon Memory for Multi-Party Collaborative Dialogues. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2602.01313

Hsu, J. (2015). Robot funerals reflect our humanity. *Discover Magazine*. https:// Robot Funerals Reflect Our Humanity | Discover Magazine

James, O. (2025). Should AI flatter us, fix us, or just inform us? *MIT Technology Review*. https:// Should AI flatter us, fix us, or just inform us? | MIT Technology Review [Accessed August 15, 2025].

Kadan, A., Deepak P., Bhadra, S., et al. (2024). Understanding latent affective bias in large pre-trained neural language models. *Natural Language Processing Journal*. 7, 100062. https://doi.org/10.1016/j.nlp.2024.100062

Kasket, E. (2019). All the ghosts in the machine: The digital afterlife of your personal data. London: Little Brown Press.

Klass, D., Silverman, P.R., Nickman, S. L. (Eds). (1996). Continuing bonds: New understandings of grief. New York: Taylor & Francis.

Krueger, J., Osler, L. (2019). Engineering affect: Emotion regulation, the internet, and the techno-social niche. *Philosophical Topics*. 47(2), 205–231.

LaBar, K.S., Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*. 7(1): 54–64.

Laestadius, L., Bishop, M., Gonzalez, M., Illenčík, D., Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media &*

国家高端交流平台

Society. 24, 5923–5941.
Lee, J.L.C., Nader, K., Schiller, D. (2017). An Update on Memory Reconsolidation Updating. *Trends in Cognitive Sciences*. 21(7), 531-545. https://doi.org/10.1016/j.tics.2017.04.006
Lei, Y., Ma, S., Sun, Y., et al. (2025). "AI Afterlife as Digital Legacy: Perceptions, Expectations, and Concerns," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (New York: ACM), 1-18.
Liu, S., & Roth, D. (2025). "Conflicts in Texts: Data, Implications and Challenges." In *Findings of the Association for Computational Linguistics: EMNLP 2025*, 10073–10091.
Loftus, E.F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*. 12(4), 361–366.
Lu, J., Yang, M., Cui, Y., et al. (2026). Exploring public perceptions of artificial intelligence in bereavement support: A qualitative study of griefbots. Frontiers in Digital Health, 8, 1852514. https://doi.org/10.3389/fdgth.2026.1852514
Lu, Z., Li, D., Shi, Y., et al. (2026). Structured Episodic Event Memory. https://doi.org/10.48550/arXiv.2601.06411
Manning, J. M., Sullivan, D., Doyle, D. T., Pinter, A. T., and Brubaker, J. R. (2026). "Designing conversations with the dead: How people engage with generative ghosts." in *Proceedings of the 2026 Designing Interactive Systems Conference (DIS '26)*, (New York: Association for Computing Machinery), 3610-3621.
Massimi, M., Charise, A. (2009). Dying, death, and mortality: Towards thanatosensitivity in HCI. *Proc. CHI 2009 Extended Abstracts*, 2459–2468. Abstract retrieved from CHI Extended Abstracts.
McLean, K, C., Syed, M. (2015). Personal, master, and alternative narratives: An integrative framework for understanding identity development in context. *Human Development*. 58 (6), 318–349.
Osler, L. (2025a). Narrative railroading. *Topoi* 44(2), 379–392. https://doi.org/10.1007/s11245-024-10146-4
Osler, L. (2025b). Mindshaping and narrative devices. In Szanto T, Slaby J (Eds). London: The Routledge handbook of mindshaping.
Osler, L. (2026). Hallucinating with AI: Distributed delusions and “AI psychosis.” *Philosophy & Technology*. 39(1), 30. https://doi.org/10.1007/s13347-026-01034-3
Ratcliffe, M. (2022). Grief worlds: A study of emotional experience. Cambridge: MIT Press.
Roberts, T., Krueger, J. (2024). Real feeling and fictional time in human-AI interactions. *Topoi*. 43(3), 783–794.
Rodriguez-Resendiz, H., Rodriguez-Resendiz, J. (2024). Digital resurrection: Challenging the boundary between life and death with artificial intelligence. *Philosophies*. 9(3), 71.
Sedikides, C., Wildschut, T., Routledge, C., et al. (2015). To nostalgize: Mixing memory with affect and desire. *Advances in Experimental Social Psychology*. 51, 189–273.

国家高端交流平台

Sedikides, C., Wildschut, T., Cheung, Wing-Yee., et al. (2016) Nostalgia fosters self-continuity: Uncovering the mechanism (social connectedness) and consequence (eudaimonic well-being). *Emotion*. 16(4), 524–539. https://doi.org/10.1037/emo0000136
Sharma, M., Tong, M., Korbak, T., et al. (2023). Towards understanding sycophancy in language models. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2310.13548
Shu, Y., Liu, P., Zhang, T., et al. (2026) TraceMem: Weaving narrative memory schemata from user conversational traces. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2602.09712

Slaby, J. (2016). Mind invasion: Situated affectivity and the corporate life hack. *Frontiers in Psychology*. 7, 1–13. https://doi.org/10.3389/fpsyg.2016.00266
Smart, P., Clowes, R., Krueger, J. *et al.* (2026). The Story of Your Life: Large Language Models and Personal Memory. *Rev.Phil.Psych.*. https://doi.org/10.1007/s13164-026-00831-1
Spens, E., Burgess, N. (2024). A generative model of memory construction and consolidation. *Nature Human Behaviour*. 8, 526–543. https://doi.org/10.1038/s41562-023-01799-z
Su, M., Guo, Y., Hou, Z. et al. (2026). Beyond Dialogue Time: Temporal Semantic Memory for Personalized LLM Agents. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2601.07468
Sutton, J. (2010). Exograms and interdisciplinarity. In R. Menary (Ed). The extended mind. Cambridge: MIT Press.
Sweeney, P. (2021). A fictional dualism model of social robots. *Ethics and Information Technology*. 23, 465–472. https://doi.org/10.1007/s10676-021-09589-9
Tanaka, S., Shimada, S., Miyahara, K. (2026). Embodying the narrative self, narrating the embodied self. *Phenom Cogn Sci*. https://doi.org/10.1007/s11097-026-10156-0
Telakivi, P. (2026). Remembering with AI: From Distributed Memory to AI-Curated and Human-AI Co-Memory. *Rev.Phil.Psych.* https://doi.org/10.1007/s13164-026-00815-1
Tribble, E. (2005). Distributing cognition in the Globe. *Shakespeare Quarterly*. 56(2), 135–155.
Van Kesteren, A. (2025). Love 2.0: Alicia Framis marries artificial intelligence. Museum Boijmans van Beuningen. https://Love 2.0: Alicia Framis Marries Artificial Intelligence - Museum Boijmans Van Beuningen [Accessed March 27, 2025].
Tronson, N.C., Taylor, J.R. (2007). Molecular mechanisms of memory reconsolidation. *Nature Reviews Neuroscience*. 8(4), 262–275.
Vold, K., Hernández-Orallo, J. (2022). "AI extenders and the ethics of mental health," In Artificial intelligence in brain and mental health, ed. F. Jotterand & M. Ienca (Springer Press), 177–202.
Wang, C., Ni, L., Yuan, B., and Tang, M. (2025). Exploring the effect of AI warm response on consumer reuse intention in service failure. *Comput. Hum. Behav.* 166:108606. doi: 10.1016/j.chb.2025.108606

国家高端交流平台

Wu, Z., & Lewis, D. S. (2026). " Universal alignment fails in global classrooms: Cross-cultural blind spots in EdTech AI. " *In Proceedings of the 40th International Conference on Machine Learning (ICML 2026)*.

Xia, M., Zhang, X., Dixit, S., et al. (2026). Memora: A Harmonic Memory Representation Balancing Abstraction and Specificity. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2602.03315

Yang, Y.T. (2025). Digital resurrection and posthumous identity: Toward a cross-cultural neurorights framework. *AJOB Neuroscience*. 24, 1–16.

Ye, W., Zheng, G., Zhang, A. (2025). Rectifying shortcut behaviors in preference-based reward learning. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2510.19050

Yin, Y., Jiang, T., Wildschut, T., Sedikides, C. (2025). Nostalgia, ritual engagement, and meaning in life. *Personality and Social Psychology Bulletin*. 51(10), 1884–1896.

Zhan, S. (2024) The first woman to marry an AI: “He” is more than just my husband. New Weekly (In Chinese) https://www.sohu.com/a/768134253_100024718 [Accessed March 21, 2025].

Zou, H., Sun, T., He, C., et al. (2026). ES-Mem: Event Segmentation-Based Memory for Long-Term Dialogue Agents. arXiv [Preprint]. Available at: https://doi.org/10.48550/arXiv.2601.07582