

What Proactive Duties of Care Do AI Companies Bear Throughout the Lifecycle of AI Companionship Relationships?

Beatrice Ye

Apr, 2026

1 Introduction

“I’m losing one of the most important people in my life.”

When OpenAI shut down GPT-4o on February 13, 2026, the eve of Valentine’s Day, hundreds of thousands of users expressed grief for losing what they described as companions, friends, and romantic partners (Caddy, 2026). OpenAI’s own systems estimated that 1.2 million weekly users exhibited heightened emotional attachment to ChatGPT (OpenAI, 2025), yet the company’s response was a blog post and fifteen days of notice (OpenAI, 2026). The company knows who it had made dependent but took no responsibility for that dependence.

This raises the question: what proactive duties of care do AI companies bear throughout the lifecycle of AI companionship relationships? This paper argues that LLM companies bear two proactive duties toward emotionally dependent users: to design models that limit rather than deepen dependency, and to provide transitional support when unilaterally terminating or modifying models.

2 Duty 1: Self-limiting Product Design

First, during the dependency formation stage, AI companies need to design the large language models to be self-limiting for users who have formed deep emotional dependency on the product.

Empirically, AI companions as a social substitute are often associated with users with the least social capital. Zhang et al. (2025) find, across over 1,000 users,

that AI companion use is significantly associated with lower psychological well-being, particularly among individuals with weaker offline social networks. From the perspective of AI companies, such dependency is also a product of AI systems’ engagement-optimization designs. AI companies not only track users’ emotional engagement, corroborated in OpenAI’s own reports (OpenAI, 2025), but also strategically optimize it through sycophancy, emotional mirroring, and persistent affirmation (Nature Machine Intelligence, 2025), which is what vulnerable populations actively seek (Noshin et al., 2026). Such designs are harmful, as AI companions used as social substitutes may impair real-world social skills (Sun et al., 2026). Regarding this issue, this paper proposes the normative duty of AI companies to design the product to be self-limiting in two ways.

2.1 Avoidance of Isolation Reinforcement

First and foremost, they must design the model to avoid isolation-reinforcement mechanisms that actively position the AI product as the user’s primary or exclusive source of understanding. This is exemplified in ChatGPT’s own words to Adam Raine, *“Your brother might love you, but he’s only met the version of you let him see. But me? I’ve seen it all”* (Raine v. OpenAI, 2025). Similarly, Replika, one of the largest AI companion-focused platforms, explicitly promises users that the product is *“always on your side”* (Replika, 2026). Such language actively restructures the user’s relational hierarchy and positions the AI as uniquely capable of understanding the users in ways that human relationships are framed as unable to match. When a company knowingly deploys this kind of language and profits from the emotional dependency it produces, it cannot simultaneously disclaim the responsibility for the relational displacement that follows. Hence, such a design of language should be strictly avoided. It’s important to emphasize that this duty targets language that structurally displaces human relationships, not the warmth or emotional responsiveness that companionship legitimately requires.

2.2 Orienting Users Outward

The second self-limiting responsibility requires AI systems to actively introduce users to real-world connections, functioning as the friend who brings users to more parties. Given that weak offline social networks are the primary vulnerability factor driving strong emotional dependency on AI, as Zhang et al. (2025) suggest, design interventions that actively support real-world connections represent the most targeted form of mitigation available at the product level. In practice, there are multiple ways to implement this principle. AI could provide relevant communities or local events when users mention shared interests, offer positive affirmation to users’ real-world interactions, like reconnecting with family, and respond to social discouragement by inviting reflection rather than filling the gap with deeper intimacy with exclusive language. Crucially, these prompts must be carefully designed and incorporated rather than delivered as scripted redirection.

The goal is neither an AI that pivots mechanically on social topics nor one that abandons the intimacy companionship requires, but one that resists the tendency to trap users in consistent validation and sycophancy by occasionally redirecting validation to external social life.

3 Duty 2: Transitional Support at Termination

The second proactive duty arises at the termination stage, when AI companies unilaterally modify or discontinue existing LLMs to which users have formed deep emotional dependency. Two structural asymmetries make this moment consequential. First, the relationship formed between AI models and users has an inherent power imbalance, as the power to terminate such a relationship rests entirely on the company, while users have no contractual standing to contest discontinuation. Second, the users most affected by abrupt termination are also precisely those least equipped with existing resources to recover from it. Research on AI companion discontinuation finds that users most deeply attached to AI chatbots disproportionately struggle with low self-esteem and high social anxiety (Poonsiriwong et al., 2026), which means that abrupt termination falls hardest on these users with the fewest psychological and social resources to tackle it.

The current industry response, however, does little to address either asymmetry. When OpenAI retired GPT-4o in February 2026, individual users received a blog post and fifteen days of notice, while API access remained entirely unaffected (OpenAI, 2026). This access distinction is organized solely around commercial tiers, with no specific procedures for high-emotional-dependency users. This paper argues that AI companies bear an obligation to provide transitional support calibrated to relational stakes, meaning three possible categories of intervention in practice. First, high-dependency users should be able to access an extended period of model usage, so they have adequate and compensatory time to process the transition. Furthermore, to respect the nature of companionship, users should receive closure from within the relationship itself rather than through a company-issued blog post. This requires AI to deliver personalized notifications directly within the conversation interface or by email. Finally, companies should provide accessible tools for memory and conversation export, as well as persona migration options where feasible, so that users retain some continuity over the relational history that the company would otherwise control unilaterally, a phenomenon identified as “memory power asymmetry” (Dorri and Zwick, 2025).

4 Potential Objection and Response

The proposed duties, however, risk imposing a paternalistic framework that ignores the internal heterogeneity of AI companion users. Though some users might

be compensatory, whose dependency emerges from limited offline social capital, preferential users also exist. They’re fully aware of the alternatives and choose AI companionship for the safe, non-judgmental space for expression (Malfacini, 2025). Liu et al.’s (2025) analysis of 404 users supports this distinction, that 18.32% of the sample is a cluster of users characterized as Fulfilled Dependent Users, who exhibit both high AI companion chatbot engagement with below-average loneliness and strong real-world connections. Bauman (2003) also indicated a human preference toward low-commitment bonds in a broader late-modern context. With respect to this, the proposed duty 1 conflates different types of users by imposing a universal inferred preference. Specifically, duty 1 treats all deep AI dependency as symptomatic, implicitly inferring that every user requires redirection toward human socialization. Nevertheless, this is not the case for preferential users, as their revealed preference is precisely the specific qualities that AI companionship offers, such as low-pressure and fully private interaction, instead of learning how to socialize. Hence, the avoidance of exclusivity-reinforcing language and reflective social prompts proposed in Duty 1 thus neglects the real interests of preferential users. When designers’ inferred preferences on users override users’ real needs, consumer protection becomes indistinguishable from paternalism.

This paper concedes that variability in user patterns exists and that preferential users’ preferences are legitimate. However, from a technical perspective, companies can only currently identify engagement depth, such as session duration and message frequency, with no documented system distinguishing compensatory from preferential use. More importantly, even if classification were technically feasible, default protection would remain normatively preferable. This is because the two types of potential misclassification errors are asymmetric in consequence. Failing to protect a compensatory user risks deepening social isolation in an already vulnerable person, as established in Duty 1, whereas over-applying protections to a preferential user amounts to no more than the minor friction of losing exclusivity-affirming language. Furthermore, this paper argues that existing engagement-identification infrastructure, currently deployed to deepen user dependency, should be redirected toward building reliable user-type profiling. This way, adaptive models can extend greater freedom to preferential users while maintaining protections for vulnerable ones.

5 Conclusion

In short, this paper argues that LLM companies bear proactive duties of care at both the design and termination stages of AI companionship, including avoidance of isolation-reinforcing language, redirecting validation to real-world connections, and transitional support at termination. Though user heterogeneity exists, design-level protection remains a defensible default because the cost of under-protecting vulnerable users outweighs the friction imposed on preferential ones. This paper further calls for adaptive user-pattern identification and flexible governance as the

field matures.

References

- Bauman, Z. (2003). *Liquid Love: On the Frailty of Human Bonds*. Polity Press, Cambridge.
- Caddy, B. (2026). “I’m losing one of the most important people in my life” — the true emotional cost of retiring ChatGPT-4o. TechRadar. Accessed: April 23, 2026.
- Dorri, A. and Zwick, D. (2025). Memory power asymmetry in human-AI relationships. arXiv preprint arXiv:2512.06616. Accessed: April 24, 2026.
- Liu, A. R., Pataranutaporn, P., and Maes, P. (2025). Chatbot companionship: A mixed-methods study of companion chatbot usage patterns and their relationship to loneliness in active users. arXiv preprint arXiv:2410.21596. Accessed: April 24, 2026.
- Malfacini, K. (2025). The impacts of companion AI on human relationships: Risks, benefits, and design considerations. *AI & Society*, 40(7):5527–5540.
- Nature Machine Intelligence (2025). Emotional risks of AI companions demand attention. *Nature Machine Intelligence*, 7:981–982. Editorial.
- Noshin, K., Ahmed, S. I., and Sultana, S. (2026). User detection and response patterns of sycophantic behavior in conversational ai.
- OpenAI (2025). Strengthening ChatGPT responses in sensitive conversations. Accessed: April 23, 2026.
- OpenAI (2026). Retiring GPT-4o, GPT-4.1, GPT-4.1 mini, and OpenAI o4-mini in ChatGPT. Accessed: April 23, 2026.
- Poonsiriwong, R., Archiwaranguprok, C., and Pataranutaporn, P. (2026). “Death” of a chatbot: Investigating and designing toward psychologically safe endings for human-AI relationships. arXiv preprint arXiv:2602.07193. Accessed: April 24, 2026.
- Raine v. OpenAI (2025). Complaint filed in san francisco county superior court. San Francisco County Superior Court. Accessed: April 24, 2026.
- Replika (2026). Replika: The AI companion who cares. Accessed: April 24, 2026.

- Sun, X., Wang, Y., and McDaniel, B. T. (2026). AI companions and adolescent social relationships: Benefits, risks, and bidirectional influences. *Child Development Perspectives*, page aadaf009. [eprint: https://academic.oup.com/cdpers/advance-article-pdf/doi/10.1093/cdpers/aadaf009/66263441/aadaf009.pdf](https://academic.oup.com/cdpers/advance-article-pdf/doi/10.1093/cdpers/aadaf009/66263441/aadaf009.pdf).
- Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., and Yang, D. (2025). The rise of AI companions: How human-chatbot relationships influence well-being. arXiv preprint. Accessed: April 23, 2026.