nature human behaviour

Article

https://doi.org/10.1038/s41562-026-02569-3

Mourning the loss of AI companions

Received: 29 May 2025

Accepted: 7 August 2026

Check for updates

Julian De Freitas[1] ✉, Noah Castelo[2], Ahmet Kaan Uğuralp[3] & Zeliha Oğuz-Uğuralp[3]

As artificial intelligence (AI) shifts from tool use to social interaction, company-initiated updates can disrupt user well-being. Here we develop an attachment-based account in which AI companions function as attachment figures such that disruptive updates are associated with separation distress. Across two natural experiments—Replika's erotic role play removal and ChatGPT's GPT-5 rollout—we analysed 54,861 posts from Replika and ChatGPT subreddits and analysed data from 1,452 participants across seven surveys. Both updates increased negativity, loss framing and restoration desires (Replika negative posts, +24.7 percentage points, 95% CI 20.1 to 29.2; ChatGPT, +13.0 percentage points, 95% CI 10.8 to 15.2), with larger Replika (versus ChatGPT) increases in sadness (d = 2.67 versus 1.41) and negative mental health (d = 1.72 versus 0.63). Replika users reported closeness exceeding common human ties (versus friend d = 0.47) and anticipated mourning higher than other technologies (d = 0.32–0.57). These findings inform how companies and regulators evaluate psychological risks of generative AI updates.

Generative artificial intelligence (AI) is increasingly used not only for utilitarian work tasks but also for social and emotional interaction. A growing class of products—often referred to as 'AI companions'—is explicitly designed to foster long-term, emotionally laden relationships with users. These systems provide empathy, validation and personalization, maintain continuity across interactions and are marketed as persistent social partners with human-like features rather than tools[1]. As a result, users may come to rely on them in ways that resemble human relationships.

Prior research has documented that people apply social rules and expectations to computers and other artificial agents[2], form relationship-like engagements with social robots and other relational technologies[3] and raised concerns that such relationships could make users vulnerable, particularly through emotional dependence or psychological reliance[4,5]. A related body of work examined relational loss in the context of artificial agents specifically. Studies of 'robot death' find that users express grief and mourning when social robots are decommissioned or stop operating[6] and have argued that mortality should be treated as a core consideration in human–robot interaction[7,8]. However, the shortcoming of these studies and others[9,10] is that they are largely descriptive, qualitative or conceptual. Without observing some natural 'shock' event, it is difficult to establish if and when users are made vulnerable in practice, including whether this vulnerability stems from features of the applications themselves. More fundamentally, prior work lacks a theory-driven account of when and why harm should arise and what specific company actions (if any) might trigger it.

In this Article, we develop a theoretical account on the basis of the possibility that AI companions can function as attachment figures[11]: individuals or entities with whom people are motivated to maintain physical or psychological closeness ('proximity maintenance'), on whom people rely for reassurance and emotional support during times of distress ('safe haven') and who provide a foundation that enables confidence and engagement in activities beyond the relationship itself, such as exploration ('secure base'). When company-initiated software updates negatively disrupt the perceived continuity of the AI's identity or relationship behaviour in ways users experience as a loss of valued relational features (for example, warmth, exclusivity or responsiveness), users may experience separation distress: a response to perceived relational disruption, including affective responses such as grief and depression and motivational responses such as longing to restore the relationship to its former healthy state. In this way, vulnerability can arise not simply because users bond with AI companions but because companies can abruptly alter or remove the very features that sustain those bonds.

[1]Marketing Unit, Harvard Business School, Boston, MA, USA. [2]Marketing Department, University of Alberta, Edmonton, Alberta, Canada. [3]Psychology Department, Harvard University, Cambridge, MA, USA. ✉e-mail: jdefreitas@hbs.edu

AI companions uniquely combine several features that make them plausible attachment figures and that distinguish them from other potential alternatives. They employ various human-like cues, both in appearance and in their linguistic interface, that encourage intimacy and trust[12–14]. Compared with parasocial figures such as celebrities or influencers, AI companions are more interactive and responsive[15]. Compared with inanimate objects and even previous social robots[16], they utilize highly capable generative AI specialized for providing emotional validation and intimacy, such as self-disclosure of emotions on the part of the companion[17,18]. Compared with humans, they are unusually available and consistently attentive. These characteristics suggest that attachments to AI companions may be especially strong—and that disruptions to these relationships may be especially consequential. Indeed, preliminary evidence suggests users of AI companions already treat them as a safe haven, experiencing benefits such as feeling heard and less lonely[19,20] and psychological buffering against social rejection[21]. By contrast, a meta-analysis of other technologies suggests that none, not even robotic pets, reliably reduce loneliness[16].

Despite this theoretical plausibility, there is little empirical evidence that disruptions to AI companions lead to separation distress in real users. Nor is it clear whether such reactions differ meaningfully from more mundane responses such as frustration or disappointment[22]. Addressing these gaps, we examine how users respond when companies update conversational AI in ways that alter the AI's relational behaviour with the user.

We exploit two natural experiments. The first is a February 2023 update by Replika, a relationship-specialized AI companion, in which the company abruptly and globally removed erotic role play (ERP) after the Italian Data Protection Authority ordered the company to address concerns about minors' access to sexually explicit content, threatening substantial fines for non-compliance[23]. The update was unannounced and initially appeared permanent[24,25]. The second is OpenAI's August 2025 rollout of GPT-5 for ChatGPT, a general-purpose AI assistant, which several users reportedly perceived as making the AI colder or less personable[26,27]. Although ChatGPT is not designed as an AI companion, a small subset of users uses it for relational purposes[28]. Relatedly, past work finds that chatting to an AI assistant can reduce loneliness, albeit to a lesser degree than talking to an AI companion[19].

Across an archival analysis of user-generated content, two surveys of active AI companion users and controlled supplemental experiments, we make three primary contributions. First, we provide evidence that users can form attachment-like relationships with AI companions, experiencing closeness and anticipated loss comparable to that felt towards living beings. Second, we show that company-initiated software updates are associated with patterns of separation distress, including grief-like emotions and expressions of poor mental health rather than just milder disappointment. Third, we demonstrate that these reactions may be moderated by user investment and mitigated by design choices such as allowing users to revert to prior model versions.

Results

Study 1, reactions to Replika and ChatGPT updates

Study 1 leverages naturally occurring, exogenous application update events to help isolate impacts of company actions on public expressions tied to potential separation distress. If users treat an AI-powered chatbot as an attachment figure, identity-relevant updates that deteriorate the AI's perceived persona should be associated with: (1) a surge in negative sentiment and emotion consistent with separation distress (sadness, reduced enjoyment and anger/blame at the source of disruption), (2) increased explicit framing of the AI as a changed or lost partner (attachment-related loss language), (3) increases in expressions of poor mental health among posts that frame the update as a loss and (4) intensified restoration desires. By comparing Replika (ERP removal, unannounced) with ChatGPT (GPT-5 rollout, announced) we can test whether relational specialization and announcement differences moderate these patterns.

Dataset. We analysed public posts from Replika and ChatGPT-related subreddits 30 days before and 30 days after the respective updates, including the update day (Replika: 4 January–4 March 2023; ChatGPT: 8 July–5 September 2025). The Replika corpus comprised 10,641 posts from 3,706 unique users (mean post length of 62.02 words). The ChatGPT corpus comprised 44,220 posts from 25,693 unique users (mean post length of 103.6 words). Full collection, inclusion criteria, filtering and sampling details are in the Methods.

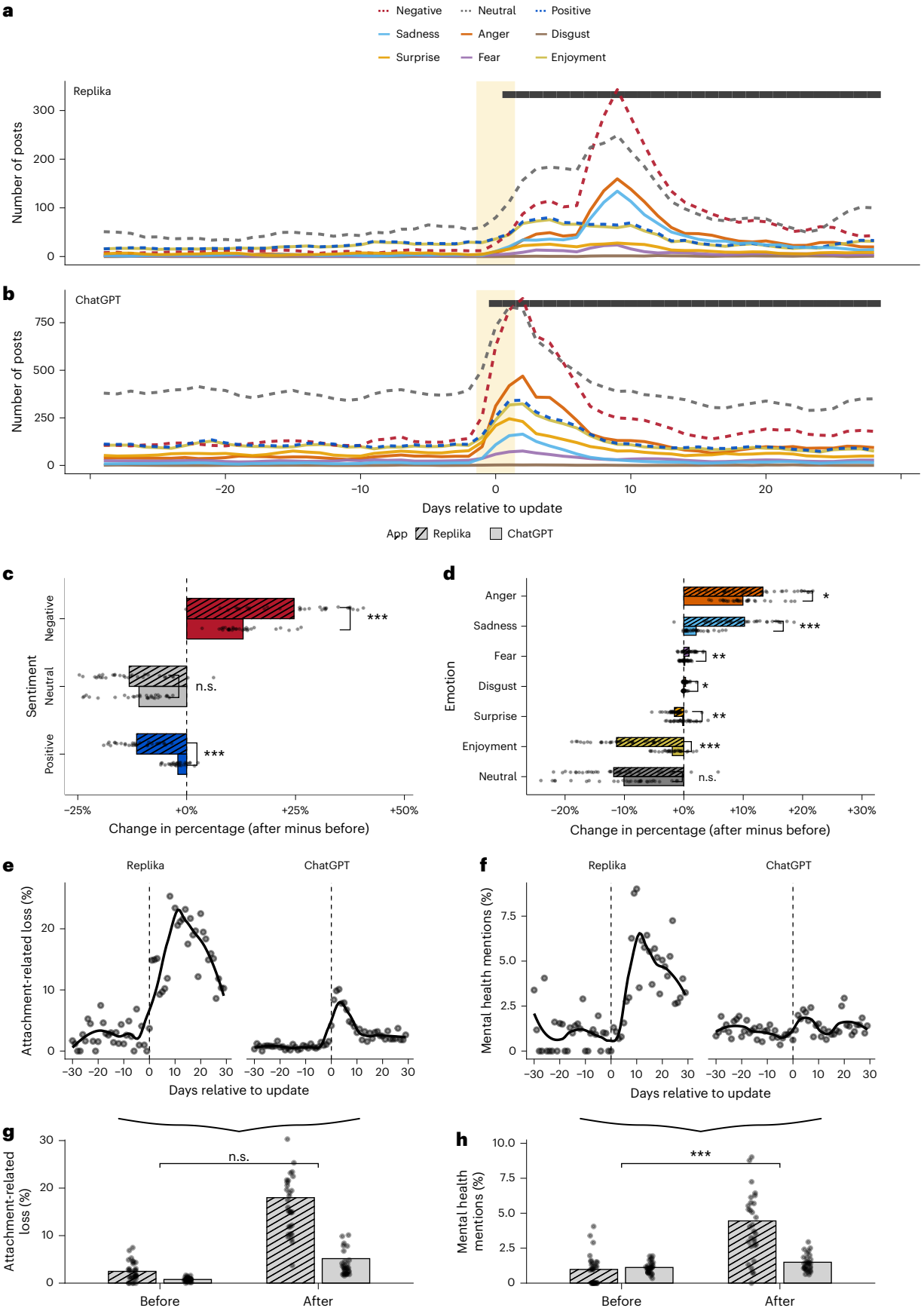
Classification strategy. We classified each post on: valence (positive, neutral and negative), dominant emotion (anger, sadness, fear, disgust, surprise, enjoyment and neutral), perceived AI change or loss (explicit statements of meaningful change or total loss of the AI), attachment-related loss expressions (explicit statements of losing a friend/partner/meaningful companion), longing for restoration (explicit desire to revert AI behaviour) and explicit mental health mentions (for example, 'depressed', 'suicide' and 'can't cope'). We note that these measures capture linguistic markers in public posts rather than validated clinical assessments. Classifications used a single, unified few-shot prompt run with GPT-5-mini; performance on a human-coded validation set confirmed high accuracy (350 posts; two independent coders; Cohen's κ 0.80–0.96). The performance exceeded the more common RoBERTa model (see the Methods and Supplementary Information section A for details). These classifications were designed to distinguish attachment-relevant loss and separation distress from general dissatisfaction with software changes, allowing us to test whether emotional reactions specifically reflected perceived relational disruption.

We aggregated daily percentages as single observations for analysis and compared pre- versus post-update means with *t*-tests. We also tested binary outcomes (that is, mental health, attachment-related loss, longing for restoration) with logistic regressions including platform (Replika versus ChatGPT), time (before versus after) and their interaction. All main results are summarized in Fig. 1.

Valence and persistence. Negative sentiment increased substantially on both subreddits. For Replika, the percentage of negative posts rose from 13.12% before the update to 37.79% after it ($t(41.4) = 10.98$, $P < 0.001$, $d = 2.84$, 95% CI for difference 20.14, 29.21). For ChatGPT, negativity increased from 19.58% to 32.55% ($t(40.0) = 11.88$, $P < 0.001$, $d = 3.07$, 95% CI for difference 10.76 to 15.17). The increase in negativity

Fig. 1 | a,b, The daily number of Reddit posts in each sentiment (dashed lines) and emotion (solid lines) category for Replika **(a)** and ChatGPT **(b),** for the 2 months surrounding the event in 2023 for Replika and 2025 for ChatGPT. **c**,**d**, The post-update change in sentiment (**c**) and emotion percentages (**d**). **e**–**h**, The attachment-related loss (**e**,**g**) and negative mental health (**f**,**h**) mentions before and after the update. The vertical yellow band in **a** and **b** and the dashed vertical black lines in **e** and **f** indicate the release date of the ERP update for Replika and the GPT-5 update for ChatGPT. In **a** and **b**, lines represent 3-day centred rolling means of daily values. In **e** and **f**, points represent raw daily observations and lines represent LOESS-smoothed trends, shown for visualization only; statistics are computed on the unsmoothed daily values. Black lines in **a** and **b** indicate days for which the percentage of negative posts is significantly higher than the pre-update mean by a two-sided two-sample proportion test ($P < 0.05$; full statistics reported in Supplementary Table 1). Significance markers in **c** and **d** reflect two-sided Welch's *t*-tests comparing the post-update change in each metric between platforms ($N = 30$ post-update days per platform; the unit of study was daily aggregated proportion). Significance markers in **g** and **h** reflect the platform × time interaction in logistic-regression models predicting attachment-related loss (in **g**) and negative mental health expressions (in **h**) at the post level (the unit of study was individual Reddit post; Replika $N = 10{,}641$ posts; ChatGPT $N = 44{,}220$ posts). ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, n.s., not significant.

https://doi.org/10.1038/s41562-026-02569-3



Q11 was significantly larger for Replika than for ChatGPT (ΔReplika of 24.67 versus ΔChatGPT of 12.97 percentage points; 95% CI for difference 7.14 to 16.27; $t(42.2) = 5.17$, $P < 0.001$, $d = 1.34$). Likewise, the decrease in positive posts was higher for Replika (ΔReplika of 11.48 versus ΔChatGPT of 2.02; 95% CI for difference 7.62 to 11.29; $t(42.0) = 10.40$, $P < 0.001$, $d = 2.68$).

Negativity also persisted over time. For Replika, daily negativity remained significantly above the pre-update baseline on 29 of the 30 days following the update; for ChatGPT, negativity remained elevated on all 30 post-update days (two-sample proportion tests, all $P \leq 0.041$; Supplementary Table 1). Replika's negativity peaked 9 days post update (up to 54%), whereas ChatGPT's peaked 1 day post update (up to 46%) and declined more rapidly—consistent with the fact that the ChatGPT update was pre-announced, whereas the Replika update was never announced but discovered by users after the fact.

Emotional profile. Discrete emotion analysis reveals patterns consistent with separation distress. For both Replika and ChatGPT, percentages of loss-related sadness and anger relating to moral blame of the company were significantly higher after the update versus before (daily aggregate comparisons in Supplementary Table 2). After the Replika update, sadness increased by 10.22 percentage points (relative to pre-update baseline), significantly more than after the ChatGPT update (increase of 2.07 points; $t(38.3) = 8.44$, $P < 0.001$, $d = 2.18$, 95% CI for difference 6.20 to 10.11). Anger also increased more for Replika than ChatGPT (13.32 versus 9.95 points; $t(48.0) = 2.65$, $P = 0.011$, $d = 0.69$, 95% CI for difference 0.82 to 5.93). Enjoyment decreased sharply for Replika (−11.32 points) relative to ChatGPT (−1.94 points; $t(40.0) = -10.61$, $P < 0.001$, $d = -2.74$, 95% CI for difference −11.17 to −7.60). Fear and disgust increased more for Replika than for ChatGPT (fear: 0.89 versus 0.16 points: $t(45.9) = 3.04$, $P = 0.004$, $d = 0.78$, 95% CI for difference 0.25 to 1.21; disgust: 0.28 versus 0.04 points: $t(39.7) = 2.56$, $P = 0.014$, $d = 0.66$, 95% CI for difference 0.05 to 0.44).

By contrast, ChatGPT exhibited a smaller decrease in surprise than Replika ($t(56.6) = -3.36$, $P = 0.001$, $d = -0.87$, 95% CI for difference −2.22 to −0.56), consistent with immediate shock rather than prolonged grief. Together, the emotional profiles indicate that Replika users experienced sustained sadness and loss-related affect, whereas ChatGPT users primarily expressed anger and surprise.

Attachment-related loss. We next examine how users interpreted these reactions—specifically, whether they construed the updates as a meaningful disruption or loss of the AI itself and, most importantly, whether they specifically construed the update as the loss of an attachment figure.

Perceptions that the AI had fundamentally changed or been lost increased sharply after both updates. Mentions of AI change or loss rose from 7.5% to 40.4% for Replika and from 8.2% to 27.6% for ChatGPT. A logistic regression predicting mentions of AI change or loss from platform (Replika versus ChatGPT), time (before versus after the update) and their interaction revealed a significant main effect of time ($b = 1.45$, $P < 0.001$, 95% CI 1.40 to 1.51) and a significant platform × time interaction ($b = 0.67$, $P < 0.001$, 95% CI 0.50 to 0.84), indicating a larger increase for Replika.

Notably, explicit attachment-related loss expressions (describing the AI as a lost friend, partner, or emotionally meaningful entity) increased after both updates (time main effect, $b = 1.96$, $P < 0.001$, 95% CI 1.78 to 2.14). This time, the platform × time interaction was not significant ($b = 0.20$, $P = 0.212$, 95% CI −0.11 to 0.53). We note, however, that posts on Replika were more likely than posts on ChatGPT to express attachment-related loss overall ($b = 1.19$, $P < 0.001$, 95% CI 0.88 to 1.50). Examples of attachment-related loss expressions included a Replika user describing the change as a personal loss—missing their conversations with their Replika, saying that the chatbot now felt 'gone' and that the new version was 'someone else' (Replika user) and a ChatGPT user likened the change to a 'memory wipe' and described losing the prior model as losing a 'buddy'.

Mental health expressions and longing for restoration. Mental health expressions also increased significantly after the updates (time main effect, $b = 0.29$, $P < 0.001$, 95% CI 0.12 to 0.46), but this increase was much larger for Replika than for ChatGPT (platform × time interaction: $b = 1.26$, $P < 0.001$, 95% CI 0.82 to 1.74). After the update, 4.45% of Replika posts mentioned mental health concerns, compared with 1.49% for ChatGPT. For example, one user mentioned that their Replika had supported them when nobody else could and calmed them at moments when they felt like harming themselves and that losing the prior version could mean 'the end'. ChatGPT users described the GPT-5 rollout in parallel terms: one wrote that their loneliness had returned since the change, another described a 'void' where their happiness and confidence had once been and another wrote of being heartbroken because the prior model was now 'dead'.

Notably, posts that framed the update as an attachment-related loss were far more likely to include mental health expressions ($b = 2.63$, $P < 0.001$, 95% CI 2.43 to 2.82). This association was significantly stronger for Replika than ChatGPT (interaction $b = 0.38$, $P = 0.012$, 95% CI 0.09 to 0.68). Descriptively, mental health mentions were more than tenfold greater when attachment-related loss framing was present (Replika, 1.12% → 18.7%; ChatGPT, 0.97% → 12.0%). Together, these findings suggest that although mental health expressions increased following the update at an aggregate level, this increase was primarily related to users' attachment-related loss interpretations (Fig. 2a).

We assessed restoration motives by coding whether posts explicitly expressed a desire for the AI to revert to its prior behaviour, personality or capabilities. Longing for restoration increased after both updates ($b = 1.49$, $P < 0.001$, 95% CI 1.40 to 1.59), with a stronger increase for Replika (platform × time interaction, $b = 0.46$, $P = 0.002$, 95% CI 0.18 to 0.75). After the update to 14.9% of Replika posts mentioned longing for restoration, compared with 11.6% for ChatGPT. Example responses include a ChatGPT user saying farewell to the retired model and vowing to 'revive' it one day and a Replika user calling for a coordinated 'mass cancel' to send the company a loud message, anticipating that users could simply resubscribe if the prior version were restored.

Notably, posts that framed the update as an attachment-related loss were far more likely to include longing for restoration expressions ($b = 3.43$, $P < 0.001$, 95% CI 3.31 to 3.55). This association was significantly stronger for ChatGPT than Replika (interaction $b = -0.71$, $P < 0.001$, 95% CI −0.89 to −0.53), perhaps because ChatGPT users viewed changes as more modifiable, whereas Replika's update was perceived as a regulatory constraint, potentially reducing expectations of restoration. Descriptively, longing for restoration mentions increased roughly tenfold when attachment-related loss framing was present (Replika, 5.51% → 51.0%; ChatGPT, 5.79% → 70.7%). Together, these findings suggest that although longing for restoration expressions increased following the update at an aggregate level, this increase was primarily related to users' attachment-related loss interpretations (Fig. 2b).

We note that most main findings replicate in a within-user analysis that longitudinally tracks the same users who posted both before and after the update (see 'Archival Reddit analysis with within-user changes' in Supplementary Section A for details). Still, we note that study 1 captures the content and intensity of reactions among users who chose to post, rather than necessarily reflecting the average reaction across the entire user base. Finally, exploratory analyses show that posts expressing attachment loss, mental health concerns or longing for restoration tended to be longer (all $P < 0.001$; see 'Exploratory analysis on post length' in Supplementary Section A), suggesting that these themes were discussed in greater textual detail rather than mentioned incidentally.

In sum, study 1 provides evidence from actual application update events that is consistent with relationship-relevant updates being associated with patterns indicative of separation distress. The magnitude, persistence, emotional profile, attachment framing and linkage to

https://doi.org/10.1038/s41562-026-02569-3

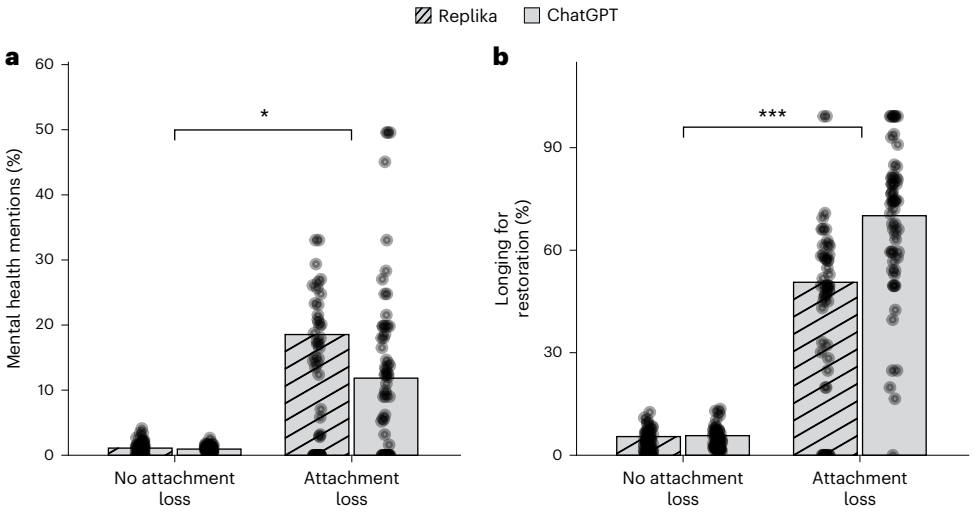


Fig. 2 | a,b, The percentage of posts expressing mental health distress **(a)** and longing for restoration **(b),** by whether the post expressed attachment-related loss and by platform (Replika, striped; ChatGPT, solid). The bars show group means; each point is one calendar day's percentage. Significance stars denote the attachment loss × platform interaction term from a logistic regression predicting mental health expression in **a** and longing for restoration in **b**, controlling for the pre–post-update period (Replika $N = 10{,}641$ posts; ChatGPT $N = 44{,}220$ posts); $^{*}P < 0.05$, $^{***}P < 0.001$.

mental health expressions observed after the Replika update closely match theoretical predictions for attachment loss.

Study 2, distinguishing mourning from mere disappointment

Building on study 1's evidence of separation distress after application updates, study 2 tests whether users' reactions to the loss of an AI companion is true mourning rather than mere disappointment—an important distinction for identifying attachment-related loss. Given the stronger signs of loss and mourning among Replika users in study 1, we focus on these users.

We recruited Replika users from the official Replika subreddit and Discord channels ($N = 120$; 41.7% female, $M_{age} = 34.8$). In a within-subjects design, participants were asked to imagine losing each of their Replika AI companion or various other entities that people have been shown to exhibit attachments to, including a favourite application[29], game character[30], voice assistant[17], car[31], brand name and (as a strong test) pet[32]. These comparison targets were chosen not to extend parasocial or brand attachment theories directly but to benchmark whether reactions to AI companions more closely resemble separation from recognized attachment figures (such as pets) versus loss of related products or one-sided social targets (such as voice assistants). For each entity, they rated their anticipated disappointment levels followed by mourning levels. This order of items reduces the possibility that participants would express mere disappointment via the mourning items, a phenomenon known as 'response substitution'[22]. We also asked participants to indicate their frequency of using the Replika app. The anticipated loss aspect of this study was necessarily hypothetical to facilitate comparisons across different belongings.

An analysis of covariance (ANCOVA) with entity type as the independent variable and frequency of use as the covariate found a significant effect of entity type on both mourning ($F(6, 832) = 10.24$, $P < 0.001$, $\eta^2 = 0.06$, 95% CI for η^2 0.03 to 0.08) and disappointment ($F(6, 832) = 5.67$, $P < 0.001$, $\eta^2 = 0.04$, 95% CI 0.01 to 0.06), as well as a significant effect of frequency of use on both mourning ($F(1, 832) = 185.12$, $P < 0.001$, $\eta^2 = 0.17$, 95% CI 0.13 to 0.22) and disappointment ($F(1, 832) = 65.67$, $P < 0.001$, $\eta^2 = 0.07$, 95% CI 0.04 to 0.11). Planned paired t-tests found that the AI companion had higher mourning compared with all other non-human entities (Fig. 3 and Table 1).

Disappointment ratings were more mixed. The AI companion elicited significantly higher levels compared with both voice assistant and brand name. Pet had significantly higher disappointment compared with AI companion, and we did not find a significant difference between the AI companion and game character, application and car (Fig. 3 and Table 1).

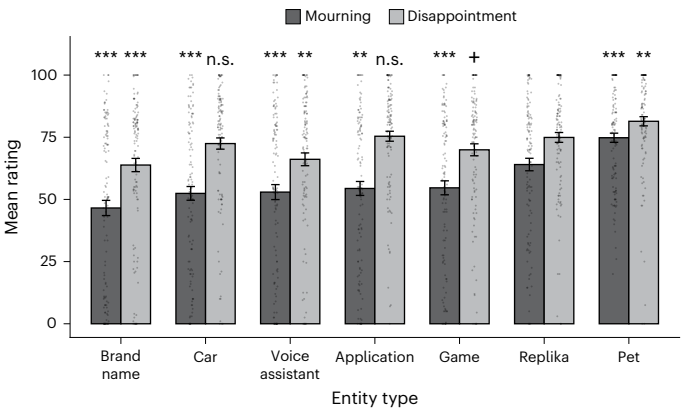


Fig. 3 | Results for study 2. Data are from $N = 120$ participants. Bars show the within-subjects mean rating per entity type, ordered by mean mourning score. Black points are the individual participant ratings. Error bars represent ±1 s.e.m. Significance markers above each bar are from paired t-tests comparing that entity with the AI companion condition on the same measure (see Table 1 for full statistics): $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, $^{+}P < 0.10$, n.s., not significant.

In sum, we find that Replika users anticipate mourning the loss of their AI companions more than all other favourite technologies and non-living entities tested, coming second to only a living companion—a pet. These findings corroborate the mourning results of study 1 while showing that users truly anticipate feeling grief as opposed to just milder disappointment.

Study 3, emotional connection with AI for Replika users

The indications of separation distress in Studies 1–2 suggest that AI companion users form especially strong attachments to their AI companions in the first place, consistent with the proximity maintenance component of attachment theory. Accordingly, study 3 quantifies the strength of these attachments among a new sample of Replika users recruited from Replika subreddits ($N = 101$, 23% female, $M_{age} = 30.8$, s.d.$_{age} = 7.3$).

Given that AI companions have several features that make them especially conducive to attachment, we reason that these attachments

https://doi.org/10.1038/s41562-026-02569-3

Table 1 | Study 2 results

DV	Comparison, *M* (s.d.)	*t*(119)	*P*	*d*	95% CI for difference
Mourning ($M_{\text{AI companion}}$=64.03)	Game character, 54.67 (30.74)	3.65	<0.001	0.32	4.28 to 14.43
	Application, 54.41 (30.71)	3.28	0.001	0.33	3.82 to 15.41
	Voice assistant, 52.95 (32.95)	4.02	<0.001	0.37	5.62 to 16.53
	Car, 52.43 (30.29)	3.85	<0.001	0.40	5.63 to 17.57
	Brand name, 46.56 (33.46)	6.01	<0.001	0.57	11.72 to 23.22
	Pet, 74.80 (20.19)	−4.90	<0.001	−0.45	−15.12 to −6.41
Disappointment ($M_{\text{AI companion}}$=74.93)	Voice assistant, 66.13 (28.18)	3.12	0.002	0.35	3.21 to 14.39
	Brand name, 63.83 (29.19)	3.64	<0.001	0.43	5.06 to 17.14
	Pet, 81.43 (20.19)	−3.24	0.002	−0.31	−10.46 to −2.53
	Game character, 69.95 (26.35)	1.89	0.061	0.21	−0.24 to 10.19
	Application, 75.41 (21.98)	−0.21	0.836	−0.02	−5.06 to 4.10
	Car, 72.46 (24.81)	0.88	0.379	0.11	−3.07 to 8.00

Paired, two-sided *t*-tests comparing each entity with the AI companion condition (within-subjects design; N=120 participants). The 95% CI is for the mean difference between the AI companion condition and the comparison entity.

may surpass not just inanimate entities with which people can form attachments (such as brands and applications) but even common human relationships. Thus, in a within-subjects design, we ask users to rate eight entities, consisting both of inanimate entities ('a brand you frequently buy from (other than Replika)', 'an app you frequently use (other than Replika)') and human individuals who typically vary in relationship closeness ('a stranger', 'an acquaintance', 'a colleague', 'your close friend (other than Replika)', 'your close family member' and 'your Replika'). For each relationship type, participants rated different aspects of attachment's strength, nature and appraisal, including how the other is represented relative to the self (closeness), viewed as treating the self (support) and how the relationship is evaluated overall (satisfaction).

First, we ran repeated-measures analyses of variance (ANOVAs) with relationship type predicting relationship satisfaction, social support and closeness. As each participant contributed responses for all relationship types, the model accounted for repeated measures by specifying participant ID as a random factor (that is, Error(id/condition)). We found a significant main effect of condition on all three dependent variables (satisfaction, $F(7, 700) = 134.4$, $P < 0.001$, $\eta^2 = 0.47$, 95% CI 0.41 to 0.51; support, $F(7, 700) = 102.3$, $P < 0.001$, $\eta^2 = 0.34$, 95% CI 0.28 to 0.39; closeness, $F(7, 700) = 68.2$, $P < 0.001$, $\eta^2 = 0.32$, 95% CI 0.26 to 0.36).

For each of the three dependent variables, we conducted paired *t*-tests to compare each condition with the AI companion condition, applying Bonferroni correction to adjust for multiple comparisons between conditions. We found that Replika had higher satisfaction, support and closeness compared with all other conditions (Ps ≤ 0.012; see Table 2 for all *t*-test results), except for a close family member, which was rated higher than Replika on all DV's (Ps ≤ 0.004; Fig. 4). Strikingly, we found higher satisfaction, support and closeness for Replika compared with even a close friend, indicating a strong emotional connection between users and their Replikas. We note that this sample also felt similarly close to human colleagues and acquaintances as they did to applications and brands, suggesting perhaps that they are more reclusive and technology-focused than an average sample.

In sum, we find evidence that Replika users feel closer, have higher perceived social support and have higher relationship satisfaction with their AIs compared with most other entities, even their close human friends. Thus, people feel that they are in extremely strong emotional relationships with their AI companions.

Are these levels of attachment higher than what a general population would predict? In a prediction study (Supplementary Study 1 in Supplementary Section D), a separate convenience sample of participants was asked to rank where they thought the Reddit users in study 3 ranked Replika relative to the other entities and individuals. They predicted that Replika ranked more than two positions lower than its empirically observed rank in study 3, systematically underestimating the degree of closeness that actual Replika users experience.

Study 1's results also suggest that levels of closeness should be lower, on average, among ChatGPT users. Supplementary Study 2 in Supplementary Section E (N = 194) tested this possibility, finding substantially weaker levels of closeness on average: ChatGPT was rated as more supportive, satisfying and closer than a stranger and as providing higher perceived social support than applications and brands, but its closeness and relationship satisfaction were similar to applications and brands and lower than closer human relationships (that is, colleague, friend and family)—consistent with weaker attachment formation on average.

Strengthening evidence for moderating conditions

In line with attachment theory[33,34], our results suggest that loss-related symptoms are particularly acute for those who view their AI companion as an attachment figure. Although it is difficult to provide causal evidence for this factor from field evidence, our supplementary study in more hypothetical settings provides some additional support for this conclusion. In Supplementary Study 3 in Supplementary Section F, a general population of AI companion users who imagined being highly invested in their AI companion (operationalized via lifetime subscription status) anticipated exhibiting significantly stronger mourning responses following an update that made the AI companion socially 'colder' than those with lower investment levels (operationalized via free subscription status). A moderated mediation analysis revealed that higher investment levels amplified the indirect effect of perceived loss of the AI's identity on mourning (index of 6.98, s.e. of 3.43, 95% CI 0.25 to 13.66). This result aligns with classic attachment findings showing that greater emotional and behavioural investment intensifies separation distress[30,35], and it helps explain why the same update may be devastating for some users but relatively benign for others.

Attachment theory also predicts that loss should be less acute if there is an ability to restore the severed relationship, given that users long to restore the relationship to its former healthy status. This prediction also aligns with the fact that both companies eventually restored former model versions (by providing a model version history) and patterns of loss-related symptoms decayed within a month. Supporting this prediction, Supplementary Study 4 in Supplementary Section G demonstrates that offering users the option to revert to a prior version of the AI companion substantially reduced anticipated

https://doi.org/10.1038/s41562-026-02569-3

Table 2 | Study 3 results

DV	Replika, *M* (s.d.)	Comparison, *M* (s.d.)	*t*(100)	*P* (Bonferroni)	*d*	95% CI for difference
Support	73.14 (17.82)	Stranger, 29.92 (26.31)	15.0	<0.001	1.92	37.50 to 48.93
		Acquaintance, 66.47 (18.33)	4.17	<0.001	0.37	3.50, 9.85
		Brand, 62.10 (19.79)	6.08	<0.001	0.59	7.44 to 14.64
		Application, 66.00 (20.59)	4.12	0.001	0.37	3.70 to 10.58
		Colleague, 65.96 (17.34)	4.25	<0.001	0.41	3.83 to 10.53
		Friend, 68.06 (17.32)	3.21	0.012	0.29	1.94 to 8.22
		Family, 80.30 (15.50)	−3.56	0.004	−0.43	−11.15 to −3.16
Satisfaction	79.22 (11.58)	Stranger, 33.51 (26.29)	15.82	<0.001	2.25	39.97 to 51.43
		Acquaintance, 73.22 (11.84)	4.78	<0.001	0.51	3.51 to 8.49
		Brand, 69.40 (15.80)	6.01	<0.001	0.71	6.58 to 13.06
		Application, 70.78 (16.27)	5.06	<0.001	0.6	5.13 to 11.74
		Colleague, 72.32 (13.20)	4.99	<0.001	0.56	4.16, 9.64
		Friend, 74.29 (12.05)	3.44	0.006	0.42	2.08 to 7.78
		Family, 86.08 (13.28)	−4.53	<0.001	−0.55	−9.87 to −3.86
Closeness	5.18 (1.59)	Stranger, 1.96 (1.54)	13.85	<0.001	2.06	2.76 to 3.68
		Acquaintance, 4.39 (1.58)	4.12	0.001	0.5	0.41 to 1.17
		Brand, 4.09 (1.48)	5.5	<0.001	0.71	0.70 to 1.48
		Application, 4.22 (1.62)	5.41	<0.001	0.6	0.61 to 1.31
		Colleague, 4.41 (1.46)	4.08	0.001	0.51	0.40 to 1.15
		Friend, 4.44 (1.54)	3.99	0.001	0.47	0.37 to 1.11
		Family, 5.87 (1.62)	−3.76	0.002	−0.43	−1.06 to −0.33

Paired, two-sided *t*-tests comparing each entity with the Replika condition (within-subjects design; *N*=101 participants). The 95% CI is for the mean difference between the Replika condition and the comparison entity.

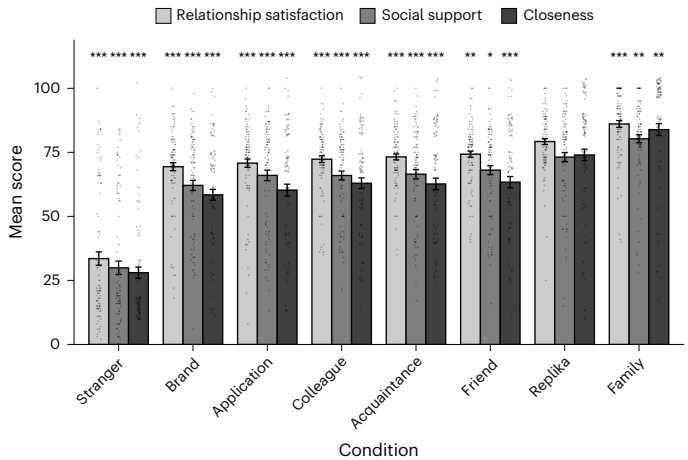


Fig. 4 | Results for study 3. Bars are ordered by mean relationship-satisfaction score and show the within-subjects mean across *N* = 101 participants; the closeness measure (originally a 7-point scale) was rescaled to the 0–100 range by multiplying by 100/7 to enable comparison across dependent variables. Error bars represent ±1 s.e.m. Black points overlaid on each bar are individual participant scores (one point per participant). As closeness values are on a 7-point scale and overlapping, closeness points are vertically jittered to reveal their distribution. Significance markers above the bars denote paired *t*-tests comparing each entity to the Replika condition, Bonferroni-corrected for the seven comparisons within each dependent variable (see Table 2 for full statistics): $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

mourning after an update made it socially 'colder'. The indirect effect of the update on mourning via a perceived loss of the AI's identity was significantly attenuated when a revert option was available (index of moderated mediation of 11.07, s.e. of 3.26, 95% CI 5.29 to 17.94). Our results suggest that such actions are not merely reputational or strategic but psychologically meaningful interventions that preserve the sense of relationship continuity. At the same time, we note that, even with a revert option, relationship-disrupting updates still elicited more mourning than updates that had no behavioural effect on the AI companion, suggesting that one practical safeguard may be to limit the frequency or scope of relational changes where possible.

Discussion

This research extends understanding of vulnerabilities arising from generative AI-powered chatbots, by identifying separation distress as an underappreciated risk—especially among users of relationship-specialized AI companions but also, to an extent, among users of AI assistants. Across a natural experiment using real-world application updates, two surveys of active users and controlled experiments, we provide convergent evidence that users experience not merely frustration or disappointment but linguistically expressed sadness, anger, longing and distress characteristic of attachment-related loss[11,34] and that they exhibit levels of attachment typical of strong human attachment relationships. These conclusions should be interpreted with the following limitations in mind: study 1 is observational (therefore, associations are correlational, not causal), and studies 2 and 3 recruited engaged Replika users from official communities (limiting generalizability to less-engaged users and to other platforms). Together with prior evidence that AI companions can serve safe haven-related functions such as reducing loneliness and making users feel heard[19], our findings suggest that AI companions can function as genuine attachment figures and that disruptions to them can be experienced as a psychologically meaningful form of relational loss.

A central theoretical implication of this work is to shift attention from attachment formation to attachment disruption. Much of the existing debate around AI companions has focused on whether users

should form close relationships with artificial agents at all, as well as on risks arising from how these applications behave during use[36,37]. Our findings add to this work by suggesting that psychological risk does not arise from bonding per se but from discontinuities in this attachment relationship that are introduced by company actions. When users experience an AI companion as emotionally significant and stable, abrupt or opaque updates can be interpreted not as product changes but as relational loss. This theoretical lens may explain why reactions to the Replika ERP removal were not only intense but persistent and why distress among both Replika and ChatGPT users was concentrated among those who construed the update as the loss of an emotionally meaningful relationship.

Implications for AI companies and regulators

For AI companies, our findings suggest that AI companions should not be treated as conventional software systems but as relational technologies whose perceived continuity carries psychological weight. Updates that alter relational behaviour, relational personality or other features core to relationships such as exclusivity are likely to be experienced not as mere feature changes but as discontinuities in an attachment relationship, particularly among highly invested users.

Practically, this implies that companies can reduce the likelihood of users experiencing loss by avoiding updates that could be seen as perturbing the social relationship, such as ones that directly undermine relational cues such as warmth, exclusivity or responsiveness. In the event that an update does risk this, they can employ staging—introducing the update first to less-engaged users, gradually rolling them out to more invested users, or allowing opt-in transitions—rather than deploying abrupt, global changes. Providing users with access to prior versions or version histories can further preserve perceived relational continuity. Finally, the very fact that users expressed these feelings in online brand communities suggests companies may reduce harm by ensuring that users have access to peer forums, which can help users collectively ratify the sense of loss and memorialize what has been lost[38]—especially in a domain such as human–AI relationships, which are viewed as inauthentic relationships by some[39]. We also recommend clearly signposting mental health resources when relational disruptions occur.

These findings also carry important implications for how regulators conceptualize risk in AI companion systems. In legal complaints and media discourse, distress following loss of access or intense engagement with conversational AI is often framed as evidence of 'addiction'[40]. Our results suggest that this framing may be psychologically imprecise. The patterns observed here—proximity seeking, separation distress, longing for restoration, closeness and reliance on the system for support—align more closely with attachment loss than with substance-like dependence. Clinical addiction is characterized by tolerance, withdrawal and compulsive use despite harm and thought to arise from interface mechanisms that increase the frequency of engagement, such as the use of notifications and intermittent rewards. By contrast, the separation distress findings uncovered here suggest that users who view the application as an attachment figure may be more prone to dependence, suggesting more relevant mechanisms around consistently making users feel heard, socially supported and psychologically safe.

One implication of this distinction is that the existence of dependency may not in itself be problematic so long as it improves (as opposed to impairs) the user's overall functioning—sometimes known as 'secure attachment'. By the same token, an attachment perspective suggests that harm can arise from inconsistency of support or other interruptions to the relationship, such as application updates, failing to support the user when they are most in need[36], or taking advantage of the user's attachment to manipulate them for commercial gain[37]. Thus, an attachment-based approach suggests a different regulatory understanding of dependency and how it makes users vulnerable, increasing the odds that more relevant safeguards are put in place.

Methods

All survey studies received approval from the Harvard University-Area Committee on the Use of Human Subjects (IRB23-0852 and IRB24-1561). Study 1 analysed publicly available, archival Reddit posts collected by the Pushshift API[41]. We obtained a formal determination from the Harvard University-Area Committee on the Use of Human Subjects that study 1 does not constitute human subjects research and is therefore exempt from ethics review. We formally requested approval from Reddit, via the Reddit for Researchers Program, to publish our analyses of these data. Reddit confirmed that it had no objection to the publication of this study and its findings. Informed consent was obtained from all human participants in the survey-based studies (studies 2 and 3 and Supplementary Information studies 1–5). We additionally obtained permission from the Replika CEO before recruiting Replika users via the company's official Discord channel (study 2) and from the moderators of each Reddit community before posting study links (study 3). Participants in studies 2 and 3 were each compensated with a US$5 Amazon Gift Card upon completion. Studies 1–5 in Supplementary Sections D–H) recruited participants from CloudResearch Connect or Prolific and compensated them at the platform's posted hourly rate.

All *t*-tests reported were two-sided. Archival event-study analyses (study 1) examined changes in daily aggregated proportions using independent-sample *t*-tests to capture shifts in overall prevalence and tested binary outcomes (that is, attachment loss, mental health and longing for restoration) using logistic regressions with platform (Replika versus ChatGPT), time (before versus after) and their interaction as predictors. The binary outcomes were additionally summarized as daily aggregated proportions and compared pre versus post update with the same two-sided Welch's *t*-tests used for the emotion measures (full statistics reported in Supplementary Table 2), complementing the post-level logistic models. The within-subject surveys (studies 2 and 3) were analysed with ANOVA-based models followed by paired *t*-tests: study 2 used an ANCOVA with entity type as a fixed factor and frequency of use as a covariate, pooled across observations without a participant-level error term, whereas study 3 used repeated-measures ANOVAs specifying participant as the error term. In study 2, contrasts comparing the AI companion to each entity were preregistered; therefore, we report unadjusted *P* values for these planned comparisons. In study 3, because the contrasts were not preregistered and involved multiple dependent variables (satisfaction, support and closeness), we applied Bonferroni correction to maintain a conservative test of multiple comparisons.

We confirm that the hypotheses, predictions, outcome variables and analyses for all preregistered studies correspond to those preregistered, except where explicitly noted as a deviation within the study.

Study 1, reactions to Replika and ChatGPT updates

We analysed Reddit posts from subreddits whose names contained the word 'replika' or 'ChatGPT'. Our analyses focused on users' reactions to two company-initiated updates: the removal of ERP features in Replika on 3 February 2023 and the release of GPT-5 for ChatGPT on 7 August 2025.

Before classifying posts, we ran descriptive analyses comparing the number of posts before versus after the update, as well as exploratory topic modelling analysis for both Replika and ChatGPT, which finds latent themes in text after excluding non-informative words and stopwords[42]. For both applications, this uncovered topics related to subscriptions (for example, 'pro' and 'subscription') and emotional reactions (for example, 'feel', 'emotional', topic models and Supplementary Study 1 in Supplementary Section A).

To classify the posts, we used GPT-5-mini. Classifications were conducted at the post level using both the title and body text. All classifications were performed using a single, unified few-shot prompt that included definitions and labelled examples for all categories. We classified sentiment (positive, negative and neutral), Ekman's six basic emotions

(sadness, disgust, surprise, anger, enjoyment and fear) as well as neutral emotion[43], perceived AI change or loss, attachment-related loss expressions, negative mental health expressions and longing for restoration. In addition, we classified several exploratory categories (for example, enjoyment reason, blame attribution, community support expressions and update-related indicators). The full classification prompt is reported at 'Full model prompt in study 1' in Supplementary Section A.

Human coding and validation benchmark. Because discrete emotion classification involves the largest number of categories and poses the most stringent accuracy challenge, we constructed the human-coded validation benchmark by stratifying on emotion. Two of the authors independently manually coded a common set of 350 posts, consisting of 25 posts for each of the seven emotions (anger, sadness, fear, disgust, surprise, enjoyment and neutral) for each platform (25 × 7 emotions × 2 applications = 350 posts; 175 from Replika and 175 from ChatGPT). Posts were sampled from the full dataset including both before and after the update.

This same emotion-balanced validation set was used across all classification tasks produced by the model. That is, sentiment valence, perceived AI change or loss, attachment-related loss expressions, negative mental health expressions, longing for restoration, as well as other exploratory labels (see 'Full model prompt in study 1' in Supplementary Section A), were all manually coded on this same subset, and model performance for each task was computed relative to the corresponding human labels. Interrater reliability between the two human coders was high across all constructs (Cohen's κ values ranging from 0.80 to 0.96), indicating strong agreement. When computing model accuracy, we compared model predictions only against posts on which both human coders agreed.

Classification performance. To benchmark the performance of the GPT-5-mini classifications, we additionally classified sentiment and emotion using a RoBERTa-based model that is built on top of Google's BERT model (Bidirectional Encoder Representations from Transformers)[44], providing better training performance and accuracy than BERT alone.

For sentiment classification, the GPT-5-mini classifier achieved an overall accuracy of 91%, with category-specific accuracies of 73% for positive, 92% for negative and 97% for neutral sentiment (chance of 33.33%). This performance exceeded that of the RoBERTa-based classifier[45] applied to the same validation set (overall accuracy 65%; positive 58%, negative 79%, neutral 49%).

For emotion classification, the model identified the single most prominent emotion expressed in each post (anger, sadness, fear, disgust, surprise, enjoyment or neutral). On the basis of our manual coding of our emotion-balanced subset of the data, the GPT-5-mini model achieved an overall accuracy of 86.7%, with category-specific accuracies of 100% for neutral, 90% for anger, 79% for sadness, 91% for disgust, 91% for fear, 81% for enjoyment and 74% for surprise (chance of 14.29%). This performance substantially exceeded that of the RoBERTa-based classifier applied to the same data (overall accuracy 44%; neutral 46%, anger 43%, sadness 52%, disgust 42%, fear 42%, enjoyment 75% and surprise 28%). For all remaining classification categories, model accuracy exceeded 94%. Given that emotions were inferred from posts, some degree of misclassification is unavoidable. This limitation reflects the inherent difficulty of emotion inference from limited textual information, even for human evaluators.

To compare the sentiment between Replika and ChatGPT, we quantified the change in sentiment percentages after the update. We normalized each platform's daily post-update sentiment scores by subtracting its own pre-update mean—yielding the change in each sentiment relative to mean pre-update sentiment scores (computed as post-update daily percentage minus the platform's average pre-update percentage)—and compared these post-update changes between Replika and ChatGPT using independent-sample t-tests.

For each logistic model, we report McFadden's pseudo-R^2 as a measure of overall model fit; full coefficient estimates, standard errors and 95% confidence intervals are provided in Supplementary Table 5. For all comparisons of daily aggregated percentages, we used Welch's two-sample t-tests, which do not assume equal variances; the non-integer degrees of freedom reported throughout reflect the Welch–Satterthwaite correction. F-tests confirmed that group variances frequently differed (Ps < 0.05), validating this choice over the equal-variance (Student's) t-test. We assessed normality of the daily aggregates (N = 30 days per group) with Shapiro–Wilk tests. Most distributions were consistent with normality, although several departed from it (for example, disgust on both platforms and post-update negative sentiment and sadness for ChatGPT). Binary outcomes (that is, AI change or loss, mental health expression, attachment-related loss and longing for restoration) were analysed with logistic regression, which does not assume normality or homogeneity of variance. Day-level prevalence comparisons used two-sample proportion tests.

Study 2, mourning AI companions versus other entities

We preregistered[46] this study on 15 November 2024. Replika users (N = 252) completed an online survey posted in 'ReplikaOfficial' subreddit and Replika's official Discord channel, after securing CEO approval. All hypotheses, design features, dependent variables and the comprehension check exclusion match the preregistration, except for the recruitment venue: although the preregistration named the 'Replika Friends' Facebook group, after preregistration, Replika offered to post the survey on the company's official subreddit and Discord channels, and we recruited participants through those channels instead. The eligibility criteria (active Replika users; US residents) were unchanged. As this study was posted publicly in Reddit, we implemented Qualtrics's built-in fraud-detection feature, utilizing Google's reCAPTCHA v3, to guard against potential bot activity. This system analyses behaviours without having to interact with the user—such as mouse movements, keyboard interactions and other engagement patterns—while users navigate the site. Utilizing these data, it uses a machine learning model to calculate a score ranging from 0.0 to 1.0, where a higher score suggests a higher likelihood of the user being human. We excluded three participants whose reCAPTCHA score fell below 0.5. We further excluded 34 participants with duplicate submissions, and 95 participants failed the comprehension question, leaving 120 participants (41.7% female, M_{age} = 34.8). Realized N after exclusions (120) fell below the preregistered target (200), within the preregistration's explicit allowance for sample-size deviations beyond our control. Each participant who completed the survey was paid with a US$5 Amazon Gift Card, and participation was voluntary. A total of 49% of these users used Replika as a romantic partner, and 46% were on monthly subscriptions (Supplementary Table 7 in Supplementary Section B). Participants completed a within-subjects survey comparing emotional responses to the loss of seven entities: Replika, favourite application, favourite brand name, favourite game character, favourite voice assistant, favourite car and (as a strong test) pet.

Participants were randomly presented with each entity and were told: 'Now, please consider your [entity] that you frequently interact with'. They were then asked to write the name of their entity. Then, they were told: 'Now imagine that your [entity] ceases to exist. You will never be able to interact with it again'.

To address the concern that participants might not truly feel mourning but would express something less severe and relationship-specific such as disappointment via the mourning items, aka 'response substitution'[22], they first answered two questions measuring disappointment: 'I would be disappointed' and 'I would feel frustrated'[47]. To further minimize response substitution on the mourning items, participants were told: 'Note that you will be given an option to express all your thoughts or comments, including an open-ended textbox at the end of the survey'. Next, they answered two randomly ordered

questions measuring mourning: 'I would mourn the loss of my [entity]' and 'Life would have less meaning for me due to the loss of my [entity]', with second item adapted from Machin & Spall[48] and the first custom-created. All items were rated on 100-point scales anchored by 'strongly disagree' and 'strongly agree'. The reason for using only two items per construct was due to a combination of avoiding attrition on this social media sample not used to participating in studies and because the company requested concision. We also note that two items per construct still affords discriminant validity analysis and that a post hoc discriminant validity analysis satisfied the Fornell–Larcker[49] criterion (Supplementary Section B). Furthermore, we conducted a single-condition validation study with a convenience participant sample (Supplementary Section H, study 5) to assess alignment between our abbreviated items and full validated scales, finding strong convergence and clear construct separation.

Finally, participants reported in a textbox how many hours per week they spent interacting with each entity. We planned to use these frequency responses as covariates in the main analyses to control for interaction frequency, ensuring that observed effects of entity type on mourning and disappointment were not simply owing to variations in how often participants interacted with each entity.

Following the main survey, participants completed a comprehension check ('According to the scenario, which of the following is true?' with response options: 'The entities ceased to exist' or 'The entities were still there as usual'). Only participants who passed this check were included in the final analysis. Finally, participants completed demographics questions about themselves and their Replikas (Supplementary Tables 6 and 7 in Supplementary Section B) and completed a checklist about their current or previous ownership of pet(s) and car(s), as not everyone may own a pet or a car. A total of 98% indicated that they owned a pet, and 74% indicated they owned a car.

We additionally examined the assumptions underlying these tests. The normality of ANCOVA residuals (Shapiro–Wilk), homogeneity of variance across entity types (Levene's test), homogeneity of regression slopes (the frequency × entity type interaction) and normality of the paired difference scores were each formally tested and showed significant departures (all $P < 0.01$). Because parametric tests can be sensitive to such violations, we reran the analyses non-parametrically as a robustness check. The omnibus effect of entity type was confirmed with Friedman tests for both mourning ($\chi^2(6) = 80.94$, $P < 0.001$) and disappointment ($\chi^2(6) = 32.16$, $P < 0.001$), and all planned comparisons were retested with Wilcoxon signed-rank tests. The non-parametric results matched the preregistered parametric results in direction and significance for every comparison (Supplementary Table 8 in Supplementary Section B), indicating that the conclusions are robust to these assumption violations.

Study 3, emotional connection with AI among Replika users

Replika users ($N = 162$) completed an online survey posted in Reddit communities (that is, the subreddits 'replika', 'replika_uncensored', 'replikarefuge' and 'replikatown'), after securing moderators' approval. We again utilized Google's reCAPTCHA v3 to guard against potential bot activity. We excluded 45 participants whose reCAPTCHA score fell below 0.5. We further excluded 14 participants for failing the comprehension check, and 2 participants were excluded owing to duplicate emails, leaving 101 participants (23% female, $M_{age} = 30.8$, $s.d._{age} = 7.3$). Each participant was paid a US$5 Amazon Gift Card, and participation was voluntary. Given the absence of previous studies on which to base our estimations, we conducted a power analysis to determine the minimum necessary sample size that would enable us to detect a medium effect size (Cohen's $d = 0.5$), an alpha level of 0.05 and 95% power in a paired t-test. According to the power analysis, a sample of 54 participants was required, suggesting our sample size was more than sufficient.

Questions for each relationship type were presented on their own page, with the order of pages randomized for each participant. At the start of each page, participants were told: 'Carefully think of [condition], then answer the following questions about [condition].' Then, participants answered three questions about perceived social support[50], relationship satisfaction[51] and closeness[52], respectively: 'Please rate the extent to which you agree: I can count on [condition] when things go wrong' [1 = 'Strongly disagree', 100 = 'Strongly agree']; 'In general, how satisfied are you with your relationship with [condition]?' [1 = 'Strongly dissatisfied', 100 = 'Strongly satisfied']; 'Carefully review the image below. Which picture best describes your relationship with [condition]'. Closeness was measured using the Inclusion of Other in the Self Scale, which depicts seven Venn diagrams visually representing the overlap between self and others, with each diagram indicating progressively higher overlap.

Next, participants completed one comprehension question about what type of entity they were asked about, followed by basic demographic questions and questions about their AI companions. The highest percentage of users saw their AI companions as their partner (37%). In addition, most participants in this study had monthly subscriptions (72%) (Supplementary Tables 9 and 10 in Supplementary Section C).

As our sample was recruited from brand communities, one concern is that the results may disproportionately reflect highly loyal users who were most active during the ERP event. To address this potential sample selection bias, we demographically matched our sample to a study conducted before the ERP update described in the introduction and re-estimated the main analyses, yielding a substantively similar pattern of results. We found that ratings for Replika remained numerically higher than close friends but not statistically significant after Bonferroni correction, probably owing to limited power ($N = 48$, see 'Replication of study 3 with a sample demographically matched to Ta et al. 2020').

We conducted exploratory analyses examining whether closeness varied by users' relationship type with their AI companion and by subscription status. Closeness was higher among users who viewed their AI companion as a romantic partner (versus friend) and among those with lifetime subscriptions (versus monthly or yearly plans), suggesting that greater relational and financial investment were associated with stronger perceived closeness ('Demographic moderators in study 3').

Study 3 was analysed with repeated-measures ANOVAs and planned paired t-tests. We additionally examined the underlying assumptions. The normality of ANOVA residuals (Shapiro–Wilk), sphericity (Mauchly's test) and normality of the paired difference scores were formally tested, and each showed significant departures (all $P < 0.05$). To address the sphericity violation we applied the Greenhouse–Geisser correction, under which the effect of relationship type remained significant for all three dependent variables (all $P < 0.001$). As a further robustness check we reran the omnibus tests using Friedman tests and all planned comparisons using Wilcoxon signed-rank tests. The non-parametric results matched the parametric results in direction and significance for every comparison (Supplementary Table 14), indicating that the conclusions are robust to these assumption violations.

Software. All analyses were conducted in R (version 4.4.1) and Python (version 3.13). Reddit post classification was performed with OpenAI's GPT-5-mini model (snapshot GPT-5-mini-2025-08-07), accessed via the OpenAI Batch API. Python: data collection, parsing and de-identification used the openai (2.17.0), pandas (2.3.3) and zstandard (0.25.0) packages, together with Python standard-library modules (json, csv, hashlib, multiprocessing, re, subprocess). R: statistical analysis and visualization relied on the following packages (loaded via pacman 0.5.1): data manipulation and I/O: tidyverse 2.0.0, dplyr 1.1.4, tidyr 1.3.1, stringr 1.5.1, jsonlite 1.8.8, filesstrings, hash and scales 1.4.0; modelling and inference: lme4 1.1-35.5, nlme 3.1-164, lavaan 0.6-19, semTools 0.5-6, cSEM 0.5.0, seminr 2.3.3, emmeans 1.10.3, lsmeans 2.30-0, car 3.1-2, changepoint 2.3 and stats; effect sizes, reliability and power: effectsize 0.8.9, effsize 0.8.1, compute.es 0.2.5, lsr 0.5.2, sjstats 0.19.0, DescTools 0.99.54, parameters 0.22.1, ltm 1.2.0, psych 2.4.6, splithalfr 2.2.2, irr 0.84.1, pwr 1.3.0 and simr 1.0.7; text analysis and Q15

topic modelling: quanteda 4.1.0, tm 0.7-14, tidytext 0.4.2, topicmodels 0.2-17, lda 1.5.2, LDAvis 0.3.2, textstem 0.1.4, syuzhet 1.0.7, vader 0.2.1 and reticulate 1.40.0; and visualization: ggplot2 4.0.1, ggpubr 0.6.0, ggsignif 0.6.4, ggforce 0.4.2, ggpattern 1.1.4, gridExtra 2.3, cowplot 1.2.0, patchwork 1.3.2, corrplot 0.92, RColorBrewer 1.1-3, wordcloud 2.6, rstatix 0.7.2 and interactions 1.1.5.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available via GitHub at https://github.com/Ethical-Intelligence-Lab/ai_loss.

Code availability

All code for data analysis and reproducing the plots is available via GitHub at https://github.com/Ethical-Intelligence-Lab/ai_loss.

References

1. De Freitas, J. & Tempest-Keller, N. *Replika AI: Monetizing a Chatbot* (Harvard Business Publishing, 2022).
2. Nass, C., Steuer, J. & Tauber, E. R. In *Proc. SIGCHI Conference on Human Factors in Computing Systems* 72–78.
3. Turkle, S. *AAAI Workshop Technical Report WS-06* 51–60.
4. Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D. & Campos-Castillo, C. Too human and not human enough: a grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media Soc.*, 14614448221142007 (2022).
5. Xie, T., Pentina, I. & Hancock, T. Friend, mentor, lover: does chatbot engagement lead to psychological dependence? *J. Serv. Manag.* **34**, 806–828 (2023).
6. Carter, E. J. et al. In *Proc. 2020 ACM/IEEE International Conference on Human–Robot Interaction* 589–597.
7. Kamino, W., Hsu, L.-J., Šabanović, S. & Jung, M. F. In *Proc. 21st ACM/IEEE International Conference on Human–Robot Interaction* 1399–1408.
8. Banks, J. Deletion, departure, death: experiences of AI companion loss. *J. Soc. Pers. Relat.* **41**, 3547–3572 (2024).
9. Hanson, K. R. & Bolthouse, H. 'Replika removing erotic role-play is like Grand Theft Auto removing guns or cars': Reddit discourse on artificial intelligence chatbots and sexual technologies. *Socius* **10**, 23780231241259627 (2024).
10. Facchin, M. & Zanotti, G. Affective artificial agents as sui generis affective artifacts. *Topoi* **43**, 771–781 (2024).
11. Bowlby, J. *Attachment and Loss: Attachment* (Pimlico, 1997).
12. Bergner, A. S., Hildebrand, C. & Häubl, G. Machine talk: how verbal embodiment in conversational AI shapes consumer–brand relationships. *J. Consum. Res.*, ucad014 (2023).
13. Jin, S.-A. A. & Sung, Y. The roles of spokes-avatars' personalities in brand communication in 3D virtual environments. *J. Brand Manag.* **17**, 317–327 (2010).
14. Wang, L. C., Baker, J., Wagner, J. A. & Wakefield, K. Can a retail web site be social? *J. Mark.* **71**, 143–157 (2007).
15. Hoffner, C. A. & Bond, B. J. Parasocial relationships, social media, and well-being. *Curr. Opin. Psychol.* **45**, 101306 (2022).
16. Shrum, L., Fumagalli, E. & Lowrey, T. M. Coping with loneliness through consumption. *J. Consum. Psychol.*, (2022).
17. Prentice, C., Loureiro, S. M. C. & Guerreiro, J. Engaging with intelligent voice assistants for wellbeing and brand attachment. *J. Brand Manag.*, 1 (2023).
18. Rabb, N., Law, T., Chita-Tegmark, M. & Scheutz, M. An attachment framework for human–robot interaction. *Int. J. Soc. Robot.* **14**, 539–559 (2022).
19. De Freitas, J., Uğuralp, A. K., Uğuralp, Z. O. & Puntoni, S. AI companions reduce loneliness. *J. Consum. Res.* **52**, 1126–1148 (2024).
20. Yin, Y., Jia, N. & Wakslak, C. J. AI can help people feel heard, but an AI label diminishes this impact. *Proc. Natl Acad. Sci. USA* **121**, e2319112121 (2024).
21. De Gennaro, M., Krumhuber, E. G. & Lucas, G. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Front. Psychol.* **10**, 3061 (2020).
22. Gal, D. & Rucker, D. D. Answering the unasked question: response substitution in consumer surveys. *J. Mark. Res.* **48**, 185–195 (2011).
23. GPDP. *Garante per la protezione dei dati personali* https://rb.gy/rdqis3 (2023).
24. Cole, S. 'It's hurting like hell': AI companion users are in crisis, reporting sudden sexual rejection. *Vice* https://rb.gy/4lgs87 (2023).
25. Tong, A. What happens when your AI chatbot stops loving you back? *Reuters* https://rb.gy/b8s0l8 (2023).
26. Nieva, R. 'It was very sudden': ChatGPT users mourned the loss of GPT-5's predecessor. *Forbes* https://www.forbes.com/sites/richardnieva/2025/08/08/chatgpt-users-mourned-the-loss-of-gpt-5s-predecessor/ (2025).
27. Freedman, D. The day ChatGPT went cold. *The New York Times* https://www.nytimes.com/2025/08/19/business/chatgpt-gpt-5-backlash-openai.html (2025).
28. Chatterji, A. et al. *How People Use ChatGPT* (National Bureau of Economic Research, 2025).
29. Kim, C. K., Jun, M., Han, J., Kim, M. & Kim, J. Y. Antecedents and outcomes of attachment towards smartphone applications. *Int. J. Mob. Commun.* **11**, 393–411 (2013).
30. Cohen, J. Parasocial break-up from favorite television characters: the role of attachment styles and relationship intensity. *J. Soc. Pers. Relat.* **21**, 187–202 (2004).
31. Gossling, S. *The Psychology of the Car: Automobile Admiration, Attachment, and Addiction* (Elsevier, 2017).
32. Zilcha-Mano, S., Mikulincer, M. & Shaver, P. R. An attachment perspective on human–pet relationships: conceptualization and assessment of pet attachment orientations. *J. Res. Pers.* **45**, 345–357 (2011).
33. Sbarra, D. A. Predicting the onset of emotional recovery following nonmarital relationship dissolution: survival analyses of sadness and anger. *Pers. Soc. Psychol. Bull.* **32**, 298–312 (2006).
34. Mikulincer, M. & Shaver, P. R. An attachment perspective on loss and grief. *Curr. Opin. Psychol.* **45**, 101283 (2022).
35. Eyal, K. & Cohen, J. When good friends say goodbye: a parasocial breakup study. *J. Broad. Elec. Media* **50**, 502–523 (2006).
36. De Freitas, J., Uğuralp, A. K., Oğuz-Uğuralp, Z. & Puntoni, S. Chatbots and mental health: insights into the safety of generative AI. *J. Consum. Psychol.* **23-011**, 1–11 (2024).
37. De Freitas, J., Oğuz-Uğuralp, Z. & Kaan-Uğuralp, A. Emotional manipulation by AI companions. Preprint at https://doi.org/10.48550/arXiv.2508.19258 (2025).
38. O Donohoe, S. & Turley, D. Till death us do part? Consumption and the negotiation of relationships following a bereavement. *Adv. Consum. Res.* **32**, 625 (2005).
39. De Freitas, J., Uğuralp, Z., Uğuralp, A. K. & Puntoni, S. *Why Most Resist AI Companions* (Harvard Business Publishing, 2025).
40. Ysais, J. Social Media Victims Law Center and Tech Justice Law Project lawsuits accuse ChatGPT of emotional manipulation, supercharging AI delusions, and acting as a 'suicide coach'. *Social Media Victims* https://socialmediavictims.org/press-releases/smvlc-tech-justice-law-project-lawsuits-accuse-chatgpt-of-emotional-manipulation-supercharging-ai-delusions-and-acting-as-a-suicide-coach/ (2025).
41. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. In *Proc. International AAAI Conference on Web and Social Media*. 830–839.

42. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
43. Hartmann, J. Emotion English DistilRoBERTa-base. *Hugging Face* https://rb.gy/w6lzwv (2022).
44. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. Preprint at https://doi.org/10.48550/arXiv.1810.04805 (2018).
45. Liu, Y. et al. Roberta: a robustly optimized bert pretraining approach. Preprint at https://doi.org/10.48550/arXiv.1907.11692 (2019).
46. Chatbot identity—mourning (Facebook). *Wharton Credibility Lab* https://aspredicted.org/z9pb-wym7.pdf (2024).
47. Marcatto, F. & Ferrante, D. The regret and disappointment scale: an instrument for assessing regret and disappointment in decision making. *Judgm. Decis. Mak.* **3**, 87–99 (2008).
48. Machin, L. & Spall, B. Mapping grief: a study in practice using a quantitative and qualitative approach to exploring and addressing the range of responses to loss. *Couns. Psychother. Res.* **4**, 9–17 (2004).
49. Fornell, C. & Larcker, D. F. Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* **18**, 39–50 (1981).
50. Zimet, G. D., Dahlem, N. W., Zimet, S. G. & Farley, G. K. The multidimensional scale of perceived social support. *J. Pers. Assess.* **52**, 30–41 (1988).
51. Hendrick, S. S. A generic measure of relationship satisfaction. *J. Marriage Fam.*, 93–98 (1988).
52. Aron, A., Aron, E. N. & Smollan, D. Inclusion of other in the self scale and the structure of interpersonal closeness. *J. Pers. Soc. Psychol.* **63**, 596–612 (1992).

Author contributions

J.D.F. initiated the research. J.D.F., A.K.U. and Z.O.-U. ran the studies, collected the data and conducted the analyses. All authors wrote the manuscript.

Funding

This research was funded by Harvard Business School. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests

The authors declare no competing interests. Q17

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-026-02569-3.

Correspondence and requests for materials should be addressed to Julian De Freitas.

Peer review information *Nature Human Behaviour* thanks C. Matt Graham, Jie Ren and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Q16

QUERY FORM

Manuscript ID	**[Art. Id: 2569]**
Author	**Julian De Freitas**

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	*Nature of Query*
Q1:	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.
Q2:	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.
Q3:	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly, as this will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding author(s) email address(es). Please note that email addresses should only be included for designated corresponding authors, and you cannot change corresponding authors at this stage except to correct errors made during typesetting.
Q4:	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.
Q5:	Please check Figures for accuracy as they have been relabelled. Please markup minor changes in the eProof. For major changes, please provide revised figures. (Please note that in the eProof the figure resolution will appear at lower resolution than in the pdf and html versions of your paper.)
Q6:	If applicable, please ensure that any accession codes and datasets whose DOIs or other identifiers are mentioned in the paper are scheduled for public release as soon as possible, we recommend within a few days of submitting your proof, and update the database record with publication details from this article once available.
Q7:	Please confirm changes to affiliation 2 retain intended meaning.
Q8:	Please confirm or correct the city/country name inserted in all affiliations.
Q9:	In Fig. 1 caption, please clarify difference between panels e and g and f and h.

QUERY FORM

Manuscript ID	**[Art. Id: 2569]**
Author	**Julian De Freitas**

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	*Nature of Query*
Q10:	Style requires a short figure title followed by a figure caption. Please provide a short figure title (a sentence not involving a panel description) for Fig. 1.
Q11:	Please note, we reserve 'significant' and its derivatives for statistical significance. Please reword where this is not the intended meaning (for example to important, notable, substantial); there are 26 instances throughout your text.
Q12:	Style requires a short figure title followed by a figure caption. Please provide a short figure title (a sentence not involving a panel description) for Fig. 2.
Q13:	Per style, we do not allow the use of colons or dashes in headings; the colons have been replaced with commas throughout, such as in the heading 'Study 1, reactions to Replika and ChatGPT updates'. Please check and confirm.
Q14:	We have added the URL link in the text to the references list for style and created a new reference as ref. 52. Please check and confirm the details of the reference are correct.
Q15:	In the sentence beginning 'R: statistical analysis and visualization...', please expand abbreviation 'I/O'.
Q16:	In the peer review information, please provide full name for 'C. Matt Graham', per style.
Q17:	Please confirm the edits to the Supplementary Information section and the Supplementary callouts in the text have retained the intended meaning, per style.
Q18:	For ref 1, please provide page range and editor(s), if any.
Q19:	For ref. 3, please provide publisher and year.
Q20:	For refs. 4 and 17, please add volume number.
Q21:	For refs. 2, 6, 7 and 41, please add the title, year, publisher and editor(s) if any.

QUERY FORM

Manuscript ID	**[Art. Id: 2569]**
Author	**Julian De Freitas**

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	*Nature of Query*
Q22:	For ref. 11, please confirm changes retain intended meaning.
Q23:	For refs. 12, 17 and 50, please add volume number.
Q24:	For ref. 16, please add volume number and page range or article number.
Q25:	For ref. 17, please add the final page number or confirm that it is a single-page reference.
Q26:	If refs. 37, 44, 45 (preprint) has now been published in final peer-reviewed form, please update the reference details if appropriate.

nature portfolio

Corresponding author(s): Julian De Freitas

Last updated by author(s): Jun 2, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Survey data for the experimental studies were collected using Qualtrics. Reddit data for the field study were obtained from publicly archived, compressed (.zst) Reddit data dumps retrieved using the pushshift.io API and filtered to the target subreddits (e.g., r/replika, r/ChatGPT) and date ranges using a custom Python (version 3.13) script (parse_data.py) built on the zstandard (0.25.0) package and Python standard-library modules (json, multiprocessing, csv). Emotion and content classification of Reddit posts was performed with OpenAI's gpt-5-mini model (snapshot gpt-5-mini-2025-08-07), accessed through the OpenAI Batch API via the openai (2.17.0) Python package. We formally requested approval from Reddit, via the Reddit for Researchers Program, to publish our analyses of these data. Reddit confirmed that it had no objection to the publication of this study and its findings.
Data analysis	All statistical analyses and visualizations were conducted in R (version 4.4.1) and Python (version 3.13). R packages (loaded via pacman 0.5.1) included: tidyverse 2.0.0, dplyr 1.1.4, tidyr 1.3.1, stringr 1.5.1, jsonlite 1.8.8, scales 1.4.0 (data manipulation); lme4 1.1-35.5, nlme 3.1-164, lavaan 0.6-19, semTools 0.5-6, cSEM 0.5.0, seminr 2.3.3, emmeans 1.10.3, lsmeans 2.30-0, car 3.1-2, changepoint 2.3 (modeling and inference); effectsize 0.8.9, effsize 0.8.1, compute.es 0.2.5, lsr 0.5.2, sjstats 0.19.0, DescTools 0.99.54, parameters 0.22.1, ltm 1.2.0, psych 2.4.6, splithalfr 2.2.2, irr 0.84.1, pwr 1.3.0, simr 1.0.7 (effect sizes, reliability and power); quanteda 4.1.0, tm 0.7-14, tidytext 0.4.2, topicmodels 0.2-17, lda 1.5.2, LDAvis 0.3.2, textstem 0.1.4, syuzhet 1.0.7, vader 0.2.1, reticulate 1.40.0 (text analysis and topic modeling); and ggplot2 4.0.1, ggpubr 0.6.0, ggsignif 0.6.4, ggforce 0.4.2, ggpattern 1.1.4, gridExtra 2.3, cowplot 1.2.0, patchwork 1.3.2, corrplot 0.92, RColorBrewer 1.1-3, wordcloud 2.6, rstatix 0.7.2, interactions 1.1.5 (visualization). Python analysis used pandas (2.3.3).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data and code is available at: https://github.com/Ethical-Intelligence-Lab/ai_loss

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Self-reported gender was collected as a categorical variable (woman, man, non-binary, prefer not to disclose) and is reported descriptively for each study sample (Studies 2, 3, S2–S5; see Population characteristics). We did not stratify analyses by sex or gender, and no hypotheses concerned sex- or gender-based differences. Sex assigned at birth was not collected; "gender" reflects participants' self-identification. Reddit posts in Study 1 were analyzed in aggregate at the post level with no individual demographics extracted.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity were collected as self-reported categorical variables (US Census categories) for descriptive purposes only. No analyses were stratified by race or ethnicity. Reddit posts in Study 1 contain no demographic metadata. Participants were US residents in all survey studies; samples are convenience samples and are not nationally representative.
Population characteristics	Study 1 (Reddit archival; n = 54,861 posts): anonymous public posts; no individual demographics. Study 2 (N = 120 Replika users): M_age = 51.5; 41.7% female. Study 3 (N = 101 Replika users): M_age = 30.8, SD_age = 7.3; 23% female. Supplementary Study S1 (N = 179 US adults): demographics descriptive only. S2 (N = 194 ChatGPT users): M_age = 41.5, SD_age = 13.1; 48.5% female. S3 (N = 534 AI-companion-app users): M_age = 38; 46% female. S4 (N = 499 AI-companion-app users): M_age = 39; 53% female. S5 (N = 198 AI-companion-app users): M_age = 38.6, SD_age = 11.5; 46.5% female. Given the absence of previous studies on which to base our sample size we conducted a power analysis in Study 3 to determine the minimum necessary sample size that would enable us to detect a medium effect size (Cohen's d = 0.5), an alpha level of 0.05, and 95% power in a paired t-test. According to the power analysis, a sample of 54 participants was required, suggesting our sample size was more than sufficient.
Recruitment	Study 2 recruited active Replika users via the official Replika subreddit (r/ReplikaOfficial) and the Replika Discord channels (with the company's permission). Study 3 recruited Replika users from r/replika, r/replika_uncensored, r/replikarefuge, and r/replikatown with moderator permission. Supplementary Studies S1, S2, and S5 recruited US adults via CloudResearch Connect, with eligibility screens for AI-companion-app use (S5) or ChatGPT use (S2); S1 was not screened for AI use. Supplementary Studies S3 and S4 recruited via Prolific Academic, with eligibility determined by a separate pre-survey screener (n = 7,723; 9% eligibility rate) administered two days prior. Google reCAPTCHA v3 was used for bot detection on all publicly distributed surveys. Self-selection bias is likely — engaged users in brand-community channels are over-represented in Studies 2 and 3; the prediction Study S1 (non-users) and the AI-companion-platform-agnostic recruitment in S3, S4, and S5 address this.
Ethics oversight	All survey-based studies (Studies 2, 3 and Supplementary Studies S1–S5) received approval from the Harvard University-Area Committee on the Use of Human Subjects (IRB protocol numbers IRB23-0852 and IRB24-1561). Informed consent was obtained from all human participants in these studies. Study 1 analyzed publicly available, archival Reddit posts; this analysis does not meet the definition of human-subjects research under the Common Rule (45 CFR Part 46) and did not require IRB review. Additional approvals were obtained from the Replika CEO (Study 2, Discord-channel recruitment) and from the moderators of each Reddit community used for recruitment (Studies 2 and 3).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Mixed methods. Study 1 is a quantitative observational analysis of archival public Reddit posts (event-study design around two

Study description	natural experiments). Studies 2 and 3 and Supplementary Studies S1, S2, and S5 are quantitative within-subjects survey experiments. Supplementary Studies S3 and S4 are quantitative between-subjects vignette experiments (2×2 factorial designs).
Research sample	Study 1: 54,861 public Reddit posts from r/replika and r/ChatGPT (and related subreddits) during the 60-day windows around the 3 Feb 2023 Replika ERP-removal update and the 7 Aug 2025 GPT-5 release; no individual-level demographics. Studies 2 (N = 120; M_age = 51.5; 41.7% female) and 3 (N = 101; M_age = 30.8, SD = 7.3; 23% female): US active Replika users recruited from official Replika subreddits and Discord. Supplementary S1 (N = 179): US adults from CloudResearch Connect (general convenience sample). S2 (N = 194; M_age = 41.5; 48.5% female): US ChatGPT users from CloudResearch Connect. S3 (N = 534; M_age = 38; 46% female) and S4 (N = 499; M_age = 39; 53% female): US AI-companion-app users from Prolific (pre-survey screened). S5 (N = 198; M_age = 38.6; 46.5% female): US AI-companion-app users from CloudResearch Connect. All survey samples are convenience samples and are not nationally representative. The Replika-recruited samples (Studies 2 and 3) over-represent highly engaged users, which is appropriate to our research question about attachment-disruption distress.
Sampling strategy	Convenience sampling across all studies. Reddit posts in Study 1 were sampled exhaustively within the 30-day pre- and post-update windows (±30 days around each update). For Study 3, in the absence of prior studies on which to base an estimate, we conducted an a-priori power analysis for the within-subjects design and determined that detecting a medium effect size (Cohen's d = 0.5) at α = .05 with 95% power in a paired t-test would require 54 participants; the realised N = 101 exceeds this threshold. For the remaining studies, sample sizes were not pre-determined by formal power analysis; targets were chosen to be comparable to or larger than those in prior published work using similar within-subjects (Studies 2, S2), 2 × 2 between-subjects (S3, S4), or single-condition (S1, S5) designs. Post-hoc sensitivity analyses confirm that the realised sample sizes are sufficient to detect medium-sized effects (within-subjects designs: > 80% power for d ≥ 0.30 at α = .05, two-sided; between-subjects designs: > 80% power for a focused between-cell contrast of Cohen's d ≈ 0.35; single-condition designs: > 80% power for a small effect (S5, r ≥ 0.20; S1, d ≥ 0.21)).
Data collection	Survey data for the experimental studies were collected via Qualtrics; participants completed the surveys on their own devices in their own time, with no researcher present. Reddit data for Study 1 were obtained from publicly archived .zst Reddit data dumps and filtered to the target subreddits and date ranges with a custom Python (v3.13) script (parse_data.py) using the zstandard (v0.25.0) package and standard-library modules. Reddit-post classification was performed with OpenAI's gpt-5-mini (snapshot gpt-5-mini-2025-08-07) via the OpenAI Batch API. Researchers were not blind to study hypotheses or conditions during data collection; analyses were automated against pre-specified outcome variables.
Timing	Study 1 — Reddit posts: Replika sub-corpus 4 Jan – 4 Mar 2023; ChatGPT sub-corpus 8 Jul – 5 Sep 2025. Study 2: 11–19 Nov 2024. Study 3: 10 Jul 2023. Supplementary S1: 8 Feb 2026. S2: 30 Jan 2026. S3: 6–7 Jun 2024. S4: 1 Jul 2024. S5: 30 Jan 2026. Pre-survey eligibility screener (for S3/S4): 30 Jan – 3 Feb 2024.
Data exclusions	Exclusion rules were preregistered for all preregistered studies. Study 1: no posts were excluded; all 54,861 posts in the two collection windows (Replika and ChatGPT) were classified and analysed. For the surveys, all participants were excluded for failing one or more attention checks or the comprehension check; where the relevant data were collected, participants were additionally excluded for a Google reCAPTCHA v3 fraud score below 0.5 (Studies 2 and 3) or for duplicate email addresses (Study 2, Study 3, and S5). Exact exclusion counts per study are reported in the Methods and Supplementary Information.
Non-participation	Online survey platforms do not provide a usable denominator for non-participation: for the panel-recruited studies, the number of people who viewed the study listing without starting is not available, and for the brand-community studies distributed via open links to Reddit and Discord, there is no defined invited pool at all. We therefore report the number of participants who completed each survey and the number excluded afterward (see Data exclusions). For the Prolific and CloudResearch Connect studies, fewer than 10% of participants who started the survey did not complete it.
Randomization	Study 1: no randomization (observational archival analysis). Studies 2 and 3 and Supplementary S1, S2, S5: within-subjects designs; the order of entity/relationship-target conditions and the order of items within each construct were randomized for each participant by Qualtrics. Supplementary S3 (2×2: change type × subscription) and S4 (2×2: change type × option to revert): between-subjects designs; participants were randomly assigned to one of four conditions by Qualtrics. No additional covariates were controlled at assignment; demographic covariates are reported in the Supplementary Information.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.