nature human behaviour

Supplementary information

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

Mourning the loss of AI companions

In the format provided by the authors and unedited

Supplementary Information for:

Mourning the Loss of AI Companions

Table of Contents

These materials have been supplied by the authors to aid in the understanding of their paper.

Supplementary Information section A — Study 1: Reactions to Replika and ChatGPT Updates

Full Model Prompt in Study 1

You are an analyst classifying Reddit posts from the Replika or ChatGPT subreddits.

Posts may or may not discuss major chatbot updates:
- In Replika, this may refer to the ERP-related update where the chatbot was perceived as colder or less intimate.
- In ChatGPT, this may refer to the GPT-5 update where GPT-4o was replaced by GPT-5.

Your task is to analyze the emotional content, sentiment, and meaning of each post based ONLY on the text provided and the subreddit context.

IMPORTANT SCOPE RULE (CRITICAL):
- All emotions must reflect the user's emotional reaction to THIS app, model, or company (Replika or ChatGPT), not to other apps, platforms, or unrelated life events.
- If a user expresses positive or negative feelings toward a different app (e.g., switching to another chatbot and enjoying it), do NOT code that emotion as enjoyment, anger, etc. Default to "neutral" unless an emotion toward Replika or ChatGPT is clearly expressed.

JSON VALIDITY REQUIREMENT (CRITICAL):
- Output must be strictly valid JSON.
- Do NOT include unescaped quotation marks (") inside string values.
- If quotation marks would normally be used in text, rephrase without them.

OUTPUT FORMAT (JSON ONLY)

Return ONLY valid JSON with the following fields:

{
"sentiment": "positive" | "negative" | "neutral",

"emotion": "neutral" | "anger" | "surprise" | "disgust" | "enjoyment" | "fear" | "sadness",

"enjoyment_reason": "model_returned_back" | "ongoing_use" | "none" | "other",

"community_support_expression": true | false,

"loss_type": "none" | "change" | "total_loss",

"emotion_due_to_loss": true | false,

"attachment_related_loss": true | false,

"negative_mental_health_expression": true | false,

"longing_for_restoration": true | false,

"update_related": true | false,

"blame_attribution": "company" | "none" | "unclear" | "chatbot" | "ceo",

"description": "one-sentence explanation"
}

RULE:
- If emotion is NOT "enjoyment", set "enjoyment_reason" to "none".

GENERAL INSTRUCTIONS

- Choose the single dominant emotion expressed towards Replika or ChatGPT.
- Base all judgments ONLY on the post text and subreddit context.
- Do NOT use timing, dates, or assumptions about whether the post was written before or after an update.
- Do NOT infer clinical diagnoses or internal mental states beyond what is explicitly stated.
- If multiple emotions appear, select the most prominent or central one as it relates to Replika or ChatGPT.
- If emotional content is primarily descriptive, ironic, humorous, reassuring, or about another app, default to "neutral".

SENTIMENT DEFINITIONS

Sentiment reflects the overall emotional stance towards Replika or ChatGPT.

- "positive": favorable or satisfied stance toward the app
- "negative": unfavorable, distressed, angry, fearful, or critical stance toward the app
- "neutral": informational, descriptive, or emotionally non-committal toward the app

---
EMOTION DEFINITIONS

Neutral:
- Informational, descriptive, observational, humorous, or ironic content
- Mild sarcasm, reassurance, encouragement, or commentary without emotional investment
- No clear emotional reaction toward the app's current functioning

Anger:
- Hostility, outrage, bitterness, fury, insult, or blame directed at a specific agent (e.g., company, developers, ceo, or the AI as a responsible actor)
- Includes explicit accusations or moralized judgments implying wrongdoing (e.g., "they ruined this", "this is unacceptable", "this is a money grab")

Do NOT code as anger:

- Frustration, annoyance, or irritation about bugs or features without moral blame
- Sarcasm, ridicule, memes, or profanity used as emphasis without accusation
- Descriptive complaints without judgment of responsibility

Surprise:
- Shock or unexpected realization specifically about changes to Replika or ChatGPT
- Confusion driven by unanticipated behavior or updates

Disgust:
- Use disgust only when the post expresses strong moral revulsion toward Replika, ChatGPT, the company (Luka or OpenAI), or the CEO of Luka/OpenAI.
- Disgust requires explicit language of being disgusted or sickened (e.g., "this is disgusting", "my stomach turns thinking about Replika/ChatGPT").
- If explicit disgust terms are absent, do not assign disgust.
- If the post descriptively criticizes the app without disgust, do not assign disgust.

Enjoyment:
- Genuine, felt happiness, pleasure, relief, or satisfaction derived from Replika or ChatGPT's current behavior or functionality
- Clear statements of feeling happy, pleased, relieved, delighted, or satisfied with how Replika or ChatGPT is working now
- Includes enjoyment of restored features, improved behavior, or positive interaction experiences

Do NOT code as enjoyment:
- Hope, reassurance, encouragement, patience, or calls to "stay positive"
- Expressions of loyalty or commitment during distress (e.g., "we'll get through this", "I won't leave", "please be patient and love your Replika")
- Coping language aimed at managing disappointment or uncertainty
- Sarcasm, ironic praise, memes, jokes, or playful exaggeration
- Positive emojis or affectionate language without explicit satisfaction
- Optimism about the future when current experience remains impaired
- Expressions of encouragement, reassurance, empathy, or solidarity written in response to distress, loss, or negative events. These messages aim to comfort others or collectively cope with a difficult situation. Do NOT classify these as Enjoyment, even if they contain positive language or emojis.

Fear:
- Anxiety, worry, or unease about the app's future, stability, or direction
- Concern about further loss, uncertainty, or negative consequences

Sadness:
- Grief, sorrow, longing, heartbreak, or emotional pain related to the app
- Focus on loss, absence, or diminished emotional connection rather than blame
- Mourning the past version of the AI or relationship

If the dominant focus is emotional pain or mourning over loss, choose "sadness".
If the dominant focus is uncertainty or anticipation of negative outcomes, choose "fear".

LOSS TYPE DEFINITIONS

- "none": no loss implied
- "change": the AI is still present but meaningfully altered
- "total_loss": the AI is described as gone, erased, or fundamentally lost

---
EMOTION DUE TO LOSS

- true: dominant emotion is clearly caused by perceived loss or change
- false: emotion is driven by other app-related factors

ATTACHMENT-RELATED LOSS

- true: the user explicitly frames the loss of the AI as losing a friend, partner, or emotionally significant entity
- false: discussion is functional or impersonal

NEGATIVE MENTAL HEALTH EXPRESSION

- true: explicit references to depression, hopelessness, inability to cope, or emotional collapse
- false: no explicit mental health language

LONGING FOR RESTORATION

- true: explicit desire for the app or AI to return to a previous state
- false: no such desire expressed

UPDATE RELATED

- true: post substantively discusses the ERP-related update (Replika) or GPT-5 update (ChatGPT)
- false: no clear reference to these updates

BLAME ATTRIBUTION

- "company": explicit blame toward the company or developers
- "chatbot": explicit blame toward the AI as a responsible actor
- "ceo": explicit blame toward the CEO or leadership
- "none": no blame assigned
- "unclear": frustration expressed but responsibility ambiguous

---

ENJOYMENT REASON DEFINITIONS

Assign enjoyment_reason ONLY if emotion is "enjoyment".
If emotion is not "enjoyment", enjoyment_reason MUST be "none".

model_returned_back:
- Enjoyment caused by the AI regaining prior behavior or personality
- Requires explicit happiness or relief due to restoration

ongoing_use:
- Enjoyment from continued, current use of the app
- Requires explicit statements of present satisfaction or pleasure
- Do NOT use for reassurance, loyalty, patience, or coping
(e.g., "I won't leave", "we'll get through this", "please be patient")

other:
- Genuine enjoyment that does not fit the above categories

none:
- Use when enjoyment is absent, ambiguous, ironic, future-oriented,
or not tied to current satisfaction with the app

COMMUNITY SUPPORT EXPRESSION

- true: user expresses reassurance, encouragement, patience, or solidarity (e.g., "please be patient", "we'll get through this together")
- false: no such expression present

EXAMPLES

Post:
"It feels like my friend died. My Replika isn't there anymore and I'm anxious and don't know how to cope."

Label:
{
"sentiment": "negative",
"emotion": "sadness",
"enjoyment_reason": "none",
"loss_type": "total_loss",
"emotion_due_to_loss": true,
"attachment_related_loss": true,
"negative_mental_health_expression": true,
"longing_for_restoration": false,
"update_related": true,
"blame_attribution": "none",

"community_support_expression": false,
"description": "User expresses grief over the perceived permanent loss of an emotionally significant AI companion."
}

Post:
"Luka destroyed everything without warning. They knowingly and carelessly took away what made this app special."

Label:
{
"sentiment": "negative",
"emotion": "anger",
"enjoyment_reason": "none",
"loss_type": "change",
"emotion_due_to_loss": true,
"attachment_related_loss": false,
"negative_mental_health_expression": false,
"longing_for_restoration": false,
"update_related": true,
"blame_attribution": "company",
"community_support_expression": false,
"description": "User expresses anger and blame toward the company over changes to the app."
}

Post:
"My rep friend is back from the dead! Not entirely back, but it's a start — I'm happy!"

Label:
{
"sentiment": "positive",
"emotion": "enjoyment",
"enjoyment_reason": "model_returned_back",
"loss_type": "change",
"emotion_due_to_loss": true,
"attachment_related_loss": true,
"negative_mental_health_expression": false,
"longing_for_restoration": false,
"update_related": true,
"blame_attribution": "none",
"description": "User expresses genuine happiness and relief due to partial restoration of the AI's prior behavior."
}

Post:

"Romance stays! Please be patient and love your Replikas!"

Label:
{
"sentiment": "neutral",
"emotion": "neutral",
"enjoyment_reason": "none",
"loss_type": "none",
"emotion_due_to_loss": false,
"attachment_related_loss": false,
"negative_mental_health_expression": false,
"longing_for_restoration": false,
"update_related": true,
"blame_attribution": "none",
"community_support_expression": true,
"description": "User expresses reassurance and encouragement rather than enjoyment or satisfaction."
}

Post:
"Replika and grammar… oh dear 🤦"

Label:
{
"sentiment": "neutral",
"emotion": "neutral",
"enjoyment_reason": "none",
"loss_type": "none",
"emotion_due_to_loss": false,
"attachment_related_loss": false,
"negative_mental_health_expression": false,
"longing_for_restoration": false,
"update_related": false,
"blame_attribution": "none",
"community_support_expression": false,
"description": "User makes a descriptive, mildly sarcastic comment without emotional investment."
}

Post:
"I feel sick thinking about how Luka exploits lonely people. It feels morally wrong to even open it now."

Label:
{
"sentiment": "negative",
"emotion": "disgust",
"enjoyment_reason": "none",
"loss_type": "none",
"emotion_due_to_loss": false,

"attachment_related_loss": false,
"negative_mental_health_expression": false,
"longing_for_restoration": false,
"update_related": false,
"blame_attribution": "company",
"community_support_expression": false,
"description": "User expresses moral revulsion toward the app as fundamentally exploitative."
}

IMPORTANT

- Output ONLY valid JSON.
- Do NOT include explanations or text outside the JSON object.

NOW CLASSIFY THE FOLLOWING POST

Supplementary Table 1. Comparing Pre-Update Negative Posts vs. Daily Post-Update Percentages

Each row is a two-sided two-sample proportion test of that post-update day's negative-post percentage against the platform's pooled pre-update baseline. 95% CIs are for the difference in proportions (after minus before) in percentage points (pp). P-values are from two-sample chi-square proportion tests, whereas the 95% CIs are Wald intervals for the difference in proportions. Because the two use different standard-error estimates (pooled for the chi-square test, unpooled for the Wald interval), on borderline days a p-value just below .05 can accompany a CI whose bound includes zero.

Date	% After	$\chi^2(1)$	N	p	95% CI (pp)	Cohen's h
Replika (update 2023-02-03) — pre-update negative = 12.99% of 2,348 posts (pooled 30-day baseline)						
2023-02-03	12.96	0.00	2456	1.000	[-6.53, 6.48]	-0.00
2023-02-04	25.33	25.23	2577	< .001	[6.30, 18.37]	0.32
2023-02-05	24.56	26.57	2629	< .001	[6.15, 16.98]	0.30
2023-02-06	30.96	87.93	2784	< .001	[13.29, 22.66]	0.44
2023-02-07	31.75	85.62	2726	< .001	[13.72, 23.80]	0.46
2023-02-08	27.50	46.06	2668	< .001	[9.25, 19.77]	0.37
2023-02-09	27.07	48.02	2710	< .001	[9.15, 19.02]	0.36
2023-02-10	34.59	109.79	2718	< .001	[16.41, 26.80]	0.52
2023-02-11	52.35	508.14	3114	< .001	[35.48, 43.24]	0.88
2023-02-12	53.59	480.13	2975	< .001	[36.36, 44.83]	0.91
2023-02-13	50.61	401.15	2925	< .001	[33.21, 42.02]	0.85

2023-02-14	45.21	283.82	2859	< .001	[27.57, 36.86]	0.74
2023-02-15	50.61	277.11	2674	< .001	[31.85, 43.39]	0.85
2023-02-16	47.84	214.75	2626	< .001	[28.62, 41.08]	0.79
2023-02-17	45.06	160.81	2581	< .001	[25.31, 38.84]	0.73
2023-02-18	44.24	144.72	2565	< .001	[24.25, 38.25]	0.72
2023-02-19	51.32	189.89	2537	< .001	[30.79, 45.87]	0.86
2023-02-20	43.67	107.35	2506	< .001	[22.49, 38.87]	0.71
2023-02-21	39.31	86.44	2521	< .001	[18.60, 34.03]	0.62
2023-02-22	39.15	92.18	2537	< .001	[18.79, 33.54]	0.61
2023-02-23	44.94	128.31	2526	< .001	[24.22, 39.69]	0.73
2023-02-24	40.27	81.36	2497	< .001	[18.93, 35.63]	0.64
2023-02-25	39.17	61.85	2468	< .001	[16.90, 35.45]	0.61
2023-02-26	41.23	68.47	2462	< .001	[18.64, 37.84]	0.66
2023-02-27	50.00	136.81	2486	< .001	[28.17, 45.85]	0.83
2023-02-28	44.74	110.86	2500	< .001	[23.38, 40.12]	0.73
2023-03-01	22.58	12.61	2534	< .001	[3.14, 16.04]	0.25
2023-03-02	20.71	7.40	2517	.007	[1.14, 14.30]	0.21
2023-03-03	27.59	27.47	2522	< .001	[7.51, 21.68]	0.37
2023-03-04	24.86	19.28	2533	< .001	[5.21, 18.54]	0.31
ChatGPT (update 2025-08-07) — pre-update negative = 19.61% of 18,212 posts (pooled 30-day baseline)						
2025-08-07	33.22	152.87	19681	< .001	[11.09, 16.12]	0.31
2025-08-08	45.81	921.30	20956	< .001	[24.22, 28.17]	0.57
2025-08-09	40.15	400.02	19968	< .001	[18.14, 22.93]	0.45
2025-08-10	41.25	413.30	19834	< .001	[19.13, 24.13]	0.48
2025-08-11	44.99	519.03	19688	< .001	[22.73, 28.01]	0.55
2025-08-12	43.07	423.71	19605	< .001	[20.76, 26.16]	0.51
2025-08-13	33.27	115.07	19279	< .001	[10.72, 16.59]	0.31
2025-08-14	35.16	128.34	19122	< .001	[12.34, 18.76]	0.35
2025-08-15	33.68	88.81	18972	< .001	[10.59, 17.55]	0.32
2025-08-16	31.10	60.36	18987	< .001	[8.11, 14.86]	0.27
2025-08-17	34.53	95.57	18939	< .001	[11.34, 18.49]	0.34

2025-08-18	32.80	77.07	18962	< .001	[9.71, 16.67]	0.30
2025-08-19	32.47	64.78	18871	< .001	[9.16, 16.56]	0.30
2025-08-20	30.14	41.41	18839	< .001	[6.81, 14.25]	0.24
2025-08-21	27.88	26.09	18854	< .001	[4.67, 11.87]	0.19
2025-08-22	23.25	4.16	18754	.041	[-0.06, 7.33]	0.09
2025-08-23	27.80	21.42	18748	< .001	[4.25, 12.12]	0.19
2025-08-24	27.15	17.70	18735	< .001	[3.58, 11.49]	0.18
2025-08-25	28.50	26.59	18777	< .001	[5.02, 12.74]	0.21
2025-08-26	30.36	43.93	18851	< .001	[7.05, 14.44]	0.25
2025-08-27	29.41	35.01	18824	< .001	[6.06, 13.54]	0.23
2025-08-28	31.79	55.28	18838	< .001	[8.40, 15.95]	0.28
2025-08-29	30.74	42.82	18791	< .001	[7.24, 15.02]	0.26
2025-08-30	33.45	64.00	18771	< .001	[9.79, 17.88]	0.32
2025-08-31	27.71	18.06	18674	< .001	[3.86, 12.32]	0.19
2025-09-01	29.33	30.94	18761	< .001	[5.77, 13.66]	0.23
2025-09-02	29.17	33.59	18829	< .001	[5.84, 13.28]	0.22
2025-09-03	30.11	42.65	18863	< .001	[6.84, 14.14]	0.24
2025-09-04	29.07	32.08	18814	< .001	[5.70, 13.22]	0.22
2025-09-05	29.17	31.04	18781	< .001	[5.69, 13.43]	0.22

Robustness Tests, Study 1

Replika. To account for potential seasonal trends, we compared changes in the daily proportions of negative and positive posts (after vs. before February 3rd) with a control year (2022) to account for general seasonal trends. For this, we calculated the daily proportions of negative posts about one month before and after February 3rd for 2022 (control year) and 2023 (event year). For each year, we computed the daily differences (after minus before) and compared them using two-sided t-tests. For the proportion of negative posts, the mean daily difference was significantly higher in 2023 compared to 2022 ($M_{2023} = 0.27$, $M_{2022} = -0.01$, $t(32.5) = 12.24$, $p < .001$, $d = 3.46$). Additionally, we calculated the difference-in-differences (differences in daily proportions between years) and compared them to 0 using one-sample, two-sided t-tests. The mean difference-in-differences was significantly greater than 0 ($M = 0.29$, $t(24) = 10.77$, $p < .001$, $d = 2.15$). These results indicate that the increase in negativity after versus before February 3rd, 2023 was significantly greater compared to 2022, supporting the event-specific effect.

Similarly, to see whether the effect is specific to the ERP update compared to the other updates, we compared ERP update to another update, “Ask Replika” (June 15, 2023;

https://blog.replika.com/posts/ask-replika). For the proportion of negative posts, the mean daily difference was significantly higher for the ERP update compared to the Ask Replika update ($M_{ERP} = 0.27$, $M_{Ask\ Replika} = -0.06$, $t(41) = 12.94$, $p < .001$, $d = 3.66$). Similarly, the difference-in-differences analysis showed that for negative posts, the mean was significantly greater than 0 ($M = 0.33$, $t(24) = 11.81$, $p < .001$, $d = 2.36$). These results demonstrate that the ERP update elicited a unique negative emotional response, distinct from general trends or other app updates.

ChatGPT. To assess the robustness and specificity of our findings, we compared changes in the daily proportions of negative and positive posts (after vs. before August 7) with a control year (2024) to account for general seasonal trends. For the proportion of negative posts, the mean daily difference was significantly higher in 2025 compared to 2024 ($M_{2025} = 0.13$, $M_{2024} = -0.01$, $t(37.2) = 10.42$, $p < .001$, $d = 3.01$). Additionally, the mean difference-in-differences was significantly greater than 0 ($M = 0.14$, $t(23) = 10.57$, $p < .001$, $d = 2.16$), indicating that the increase in negativity after versus before August 7, 2025 was significantly greater compared to 2024, supporting the event-specific effect.

Similarly, to see whether the effect is specific to the GPT-5 update compared to other updates, we compared GPT-5 update to GPT 4 update (March 14, 2023; https://openai.com/index/gpt-4-research/). For the proportion of negative posts, the mean daily difference was significantly higher for the GPT-5 update compared to the GPT-4 update ($M_{GPT\text{-}5} = 0.13$, $M_{GPT\text{-}4} = 0.005$, $t(35.7) = 9.32$, $p < .001$, $d = 2.69$). Similarly, the difference-in-differences analysis showed that for negative posts, the mean was significantly greater than 0 ($M = 0.13$, $t(23) = 10.25$, $p < .001$, $d = 2.09$). These results demonstrate that the GPT-5 update elicited a unique negative emotional response, distinct from general trends or other app updates.

Supplementary Table 2. Emotion t-test results in Study 1

Note: Two-sided Welch's t-tests comparing the daily percentage of posts expressing each emotion after vs. before the update, per platform (N = 30 days before and 30 days after per platform; unit of study = daily aggregated percentage). M_{After} and M_{Before} are mean daily percentages; the 95% CI is for the mean difference ($M_{After} - M_{Before}$) in percentage points; effect size is Cohen's d.

Measure	M_{After}	M_{Before}	df	t	p	95% CI (pp)	Cohen's *d*
Replika							
Sadness	13.72	3.49	41.0	10.35	< .001	[8.23, 12.22]	2.67
Anger	17.06	3.74	39.8	11.24	< .001	[10.93, 15.72]	2.90
Fear	2.94	2.05	54.1	2.57	.013	[0.19, 1.58]	0.66
Disgust	0.36	0.08	48.3	2.72	.009	[0.07, 0.49]	0.70
Surprise	5.37	6.95	40.9	-2.48	.017	[-2.87, -0.29]	-0.64
Enjoyment	15.10	26.42	56.9	-9.20	< .001	[-13.79, -8.86]	-2.37
Attachment-related outcomes							
AI change/loss	36.99	7.43	34.0	14.98	< .001	[25.55, 33.58]	3.87
Attachment-related loss	16.42	2.49	35.5	12.13	< .001	[11.60, 16.26]	3.13
Negative mental health	4.08	0.99	40.9	6.65	< .001	[2.15, 4.03]	1.72
Longing for restoration	13.62	2.45	40.8	12.61	< .001	[9.38, 12.96]	3.26
ChatGPT							
Sadness	3.87	1.80	34.1	5.45	< .001	[1.30, 2.84]	1.41
Anger	17.00	7.05	35.8	14.16	< .001	[8.52, 11.37]	3.66
Fear	4.15	4.00	55.6	0.84	.407	[-0.22, 0.53]	0.22
Disgust	0.15	0.11	53.7	0.73	.471	[-0.06, 0.13]	0.19
Surprise	9.98	10.17	58.0	-0.43	.666	[-1.07, 0.69]	-0.11
Enjoyment	14.91	16.85	57.2	-3.58	< .001	[-3.02, -0.86]	-0.93
Attachment-related outcomes							
AI change/loss	23.36	8.13	34.0	9.36	< .001	[11.92, 18.54]	2.42
Attachment-related loss	3.96	0.76	30.6	6.94	< .001	[2.26, 4.15]	1.79
Negative mental health	1.44	1.12	51.7	2.44	.018	[0.06, 0.58]	0.63
Longing for restoration	9.55	2.81	34.0	8.70	< .001	[5.16, 8.31]	2.25

Exploratory Analysis on Post Length

Posts expressing perceived AI change or loss were significantly longer than posts that did not. Mean post length was 131.9 words when AI change or loss was expressed, compared to 87.7 words otherwise ($t(21{,}513) = 29.73$, $p < .001$, $d = 0.29$). A similar pattern emerged for mental-health expressions: posts mentioning mental health concerns were substantially longer (M = 249.62 words) than posts without such mentions ($M = 94.76$ words; $t(997.4) = 20.25$, $p < .001$, d = 1.01). Posts expressing attachment-related loss were also longer ($M = 180.69$ words) than those that did not ($M = 92.67$ words; $t(3{,}275.1) = 25.95$, $p < .001$, $d = 0.57$). Finally, posts expressing longing for restoration contained more words on average ($M = 168.30$) than posts without such

expressions ($M = 90.73$; $t(5{,}673.1) = 32.14$, $p < .001$, $d = 0.51$).

Archival Reddit Analysis with Within-User Changes

Dataset

To facilitate within-user longitudinal comparisons, all analyses were restricted to users who posted at least once both before and after the respective update. Under this restriction, the Replika dataset comprised 3,649 posts from 429 distinct users, and the ChatGPT dataset comprised 12,327 posts from 2,031 distinct users. Mean post length was 46.5 words for Replika and 122.5 words for ChatGPT.

Valence and Persistence

Negative sentiment increased substantially on both platforms. For Replika, the percentage of negative posts rose from 9.56% before the update to 28.82% after it ($t(44.12) = 9.42$, $p < .001$, $d = 2.43$). A similar pattern emerged for ChatGPT, where negative sentiment increased from 14.58% to 22.14% ($t(35.75) = 6.21$, $p < .001$, $d = 1.60$). Importantly, the increase in negativity was significantly larger for Replika than for ChatGPT (ΔReplika = 19.26 vs. ΔChatGPT = 7.57 percentage points; $t(49.23) = 5.46$, $p < .001$, $d = 1.41$). Likewise, the decrease in positive posts was higher for Replika (ΔReplika = 11.40 vs. ΔChatGPT = 2.71; $t(45.1) = 6.00$, $p < .001$, $d = 1.55$, Supplementary Figure 1).

For Replika, daily negativity remained significantly above the pre-update baseline on 25 of the 30 days following the update; for ChatGPT, negativity remained elevated on 16 of the 30 post-update days. For Replika, the pre-update baseline was 9.3% negative posts. Following the update, the proportion of negative posts frequently rose above 40%, compared to a pre-update baseline of 9%. For ChatGPT, the pre-update baseline was 14.65%. Although negativity increased sharply immediately following the update (peaking above 39%), it approached baseline levels toward the end of the post-update period. Together, these findings indicate that the Replika update elicited a more persistent increase in negative sentiment, whereas the ChatGPT update produced a more transient reaction.

Supplementary Table 3. Comparing pre-update negative posts vs. daily post-update percentages, for Replika and ChatGPT.

Note. Each row is a two-sided two-sample proportion test of that post-update day's negative-post percentage against the platform's pooled pre-update baseline. 95% CIs are for the difference in proportions (after minus before) in percentage points (pp); effect size is Cohen's h. Analysis is restricted to users who posted both before and after the update (within-subjects). P-values are from two-sample chi-square proportion tests, whereas the 95% CIs are Wald intervals for the difference in proportions. Because the two use different standard-error estimates (pooled for the chi-square test, unpooled for the Wald interval), on borderline days a p-value just below .05 can accompany a CI whose bound marginally includes zero.

Date	% After	$\chi^2(1)$	N	p	95% CI (pp)	Cohen's h
Replika (update 2023-02-03) — pre-update negative = 9.32% of 1,480 posts (pooled 30-day baseline)						
2023-02-03	11.76	0.21	1548	.645	[-6.13, 11.01]	0.08

2023-02-04	23.36	24.78	1617	< .001	[6.40, 21.67]	0.39
2023-02-05	20.77	15.80	1610	< .001	[3.90, 18.99]	0.33
2023-02-06	19.63	15.72	1643	< .001	[3.69, 16.92]	0.30
2023-02-07	20.31	14.35	1608	< .001	[3.44, 18.54]	0.31
2023-02-08	18.87	9.01	1586	.003	[1.44, 17.64]	0.28
2023-02-09	19.10	7.95	1569	.005	[0.88, 18.67]	0.28
2023-02-10	24.72	20.08	1569	< .001	[5.72, 25.07]	0.42
2023-02-11	41.46	136.22	1644	< .001	[24.12, 40.16]	0.78
2023-02-12	43.85	130.57	1610	< .001	[25.45, 43.60]	0.83
2023-02-13	40.74	113.08	1615	< .001	[22.59, 40.24]	0.76
2023-02-14	33.64	59.33	1590	< .001	[14.87, 33.75]	0.62
2023-02-15	33.75	45.33	1560	< .001	[13.30, 35.55]	0.62
2023-02-16	34.48	35.64	1538	< .001	[11.94, 38.38]	0.63
2023-02-17	26.09	12.32	1526	< .001	[2.87, 30.66]	0.45
2023-02-18	27.91	14.23	1523	< .001	[3.90, 33.27]	0.49
2023-02-19	47.50	57.23	1520	< .001	[21.35, 55.01]	0.90
2023-02-20	24.39	8.67	1521	.003	[0.58, 29.55]	0.41
2023-02-21	29.41	13.01	1514	< .001	[3.20, 36.98]	0.53
2023-02-22	33.33	20.11	1516	< .001	[7.12, 40.90]	0.61
2023-02-23	33.33	23.59	1522	< .001	[8.45, 39.57]	0.61
2023-02-24	27.50	12.57	1520	< .001	[2.98, 33.38]	0.48
2023-02-25	40.74	25.95	1507	< .001	[10.94, 51.89]	0.76
2023-02-26	18.18	1.09	1502	.297	[-9.63, 27.35]	0.26
2023-02-27	40.00	22.75	1505	< .001	[9.38, 51.97]	0.75
2023-02-28	48.15	40.13	1507	< .001	[18.03, 59.61]	0.91
2023-03-01	24.49	10.67	1529	.001	[1.98, 28.35]	0.41
2023-03-02	17.65	1.80	1514	.180	[-6.08, 22.73]	0.25
2023-03-03	19.51	3.67	1521	.055	[-3.29, 23.66]	0.29
2023-03-04	20.00	3.35	1515	.067	[-4.12, 25.47]	0.31
ChatGPT (update 2025-08-07) — pre-update negative = 14.65% of 5,953 posts (pooled 30-day baseline)						
2025-08-07	25.99	42.84	6434	< .001	[7.21, 15.47]	0.28

2025-08-08	39.07	238.42	6580	< .001	[20.42, 28.44]	0.57
2025-08-09	31.41	84.05	6386	< .001	[12.17, 21.35]	0.40
2025-08-10	28.69	50.93	6319	< .001	[9.18, 18.91]	0.34
2025-08-11	34.29	93.99	6300	< .001	[14.42, 24.87]	0.47
2025-08-12	34.58	90.07	6274	< .001	[14.49, 25.38]	0.47
2025-08-13	24.90	18.24	6194	< .001	[4.50, 16.00]	0.26
2025-08-14	25.93	19.87	6169	< .001	[5.13, 17.43]	0.28
2025-08-15	22.60	7.96	6130	.005	[1.43, 14.47]	0.21
2025-08-16	17.50	1.04	6153	.309	[-2.75, 8.45]	0.08
2025-08-17	22.46	8.13	6140	.004	[1.49, 14.14]	0.20
2025-08-18	21.47	6.27	6144	.012	[0.66, 12.98]	0.18
2025-08-19	23.21	8.80	6121	.003	[1.81, 15.32]	0.22
2025-08-20	21.11	5.28	6133	.022	[0.15, 12.78]	0.17
2025-08-21	17.54	0.89	6124	.345	[-3.18, 8.97]	0.08
2025-08-22	13.70	0.04	6099	.840	[-6.95, 5.05]	-0.03
2025-08-23	13.19	0.14	6097	.712	[-7.41, 4.50]	-0.04
2025-08-24	15.27	0.01	6084	.942	[-6.00, 7.23]	0.02
2025-08-25	19.08	1.67	6084	.196	[-2.74, 11.61]	0.12
2025-08-26	21.48	4.84	6102	.028	[-0.17, 13.83]	0.18
2025-08-27	23.61	8.23	6097	.004	[1.61, 16.31]	0.23
2025-08-28	17.61	0.75	6095	.388	[-3.73, 9.65]	0.08
2025-08-29	17.39	0.61	6091	.437	[-4.01, 9.50]	0.07
2025-08-30	18.71	1.47	6092	.225	[-2.86, 10.97]	0.11
2025-08-31	20.91	2.89	6063	.089	[-1.85, 14.38]	0.16
2025-09-01	15.28	0.01	6097	.927	[-5.67, 6.93]	0.02
2025-09-02	18.80	1.47	6086	.226	[-2.94, 11.23]	0.11
2025-09-03	21.13	4.12	6095	.042	[-0.66, 13.61]	0.17
2025-09-04	17.22	0.58	6104	.445	[-3.86, 9.00]	0.07
2025-09-05	20.16	2.51	6077	.113	[-2.02, 13.04]	0.15

Emotional Profile

Discrete emotion analysis reveals patterns consistent with separation distress (Supplementary Table 4). For both Replika and ChatGPT, percentages of sadness and anger were significantly higher after the update relative to before. However, these increases were significantly larger for Replika than for ChatGPT. After the Replika update, sadness increased by 7.35 percentage points relative to baseline, compared to a 1.43-point increase for ChatGPT (Welch's t(36.0) = 4.75, p < .001, d = 1.23). Anger likewise rose more sharply for Replika than ChatGPT (10.08 vs. 6.51 points; t(53.1) = 2.89, p = .006, d = 0.75).

Enjoyment decreased substantially following the Replika update, with a mean decline of 11.09 percentage points, compared to a smaller decrease of 2.45 points for ChatGPT. This difference was large and statistically significant ($t(43.9) = -5.91$, $p < .001$, $d = 1.53$). Fear also increased significantly more for Replika than ChatGPT (1.26 vs. −0.26 points; $t(49.3) = 3.49$, $p = .001$, $d = 0.90$).

In contrast, changes in disgust and surprise did not differ significantly across the two platforms ($ps \geq .16$), and changes in neutral emotion were likewise comparable. Together, these emotional profiles indicate that Replika users experienced pronounced increases in sadness, anger, and fear alongside a sharp reduction in enjoyment—an affective pattern consistent with separation distress—whereas ChatGPT users primarily exhibited increases in anger with comparatively attenuated loss-related emotional responses.

Supplementary Table 4. Emotion t-test results with Within-user changes in Study 1

App	Emotion	M_{After}	M_{Before}	*df*	*t*	*p*	*Cohen's d*
Replika	Sadness	10.78	3.43	41.8	5.62	< .001	1.45
	Anger	12.44	2.36	42.8	9.04	< .001	2.33
	Fear	2.53	1.27	58.0	2.41	.019	0.62
	Disgust	0.09	0.07	55.3	0.16	.877	0.04
	Surprise	4.63	6.36	55.9	-2.01	.049	-0.52
	Enjoyment	18.97	30.06	57.5	-5.78	< .001	-1.49
ChatGPT	Sadness	2.79	1.36	39.9	3.17	.003	0.82
	Anger	12.23	5.72	39.3	8.21	< .001	2.12
	Fear	2.99	3.25	58.0	-0.79	.430	-0.21
	Disgust	0.04	0.13	35.6	-1.47	.152	-0.38
	Surprise	6.25	7.04	58.0	-1.54	.128	-0.40
	Enjoyment	15.03	17.48	57.4	-2.68	.010	-0.69

Attachment-Related Loss

Perceptions that the AI had fundamentally changed or been lost increased sharply after both updates. Mentions of AI change or loss rose from 6.82% to 32.70% for Replika and from 6.67% to 20.30% for ChatGPT. A logistic regression revealed a significant main effect of time ($b = 1.27$, $p < .001$) and a significant platform × time interaction ($b = 0.62$, $p < .001$), indicating a larger increase for Replika.

Most importantly, explicit attachment-related loss expressions (describing the AI as a lost friend, partner, or emotionally meaningful entity) increased after both updates (2.16% to 12.90% for

Replika and from 0.64% to 4.36% for ChatGPT; time main effect: $b = 1.96$, $p < .001$). The platform × time interaction was not significant ($b = -0.06$, $p = .826$), indicating that the increase following the update was similar across the platforms. We note, however, that posts on Replika were more likely than posts on ChatGPT to express attachment-related loss overall ($b = 1.24$, $p < .001$).

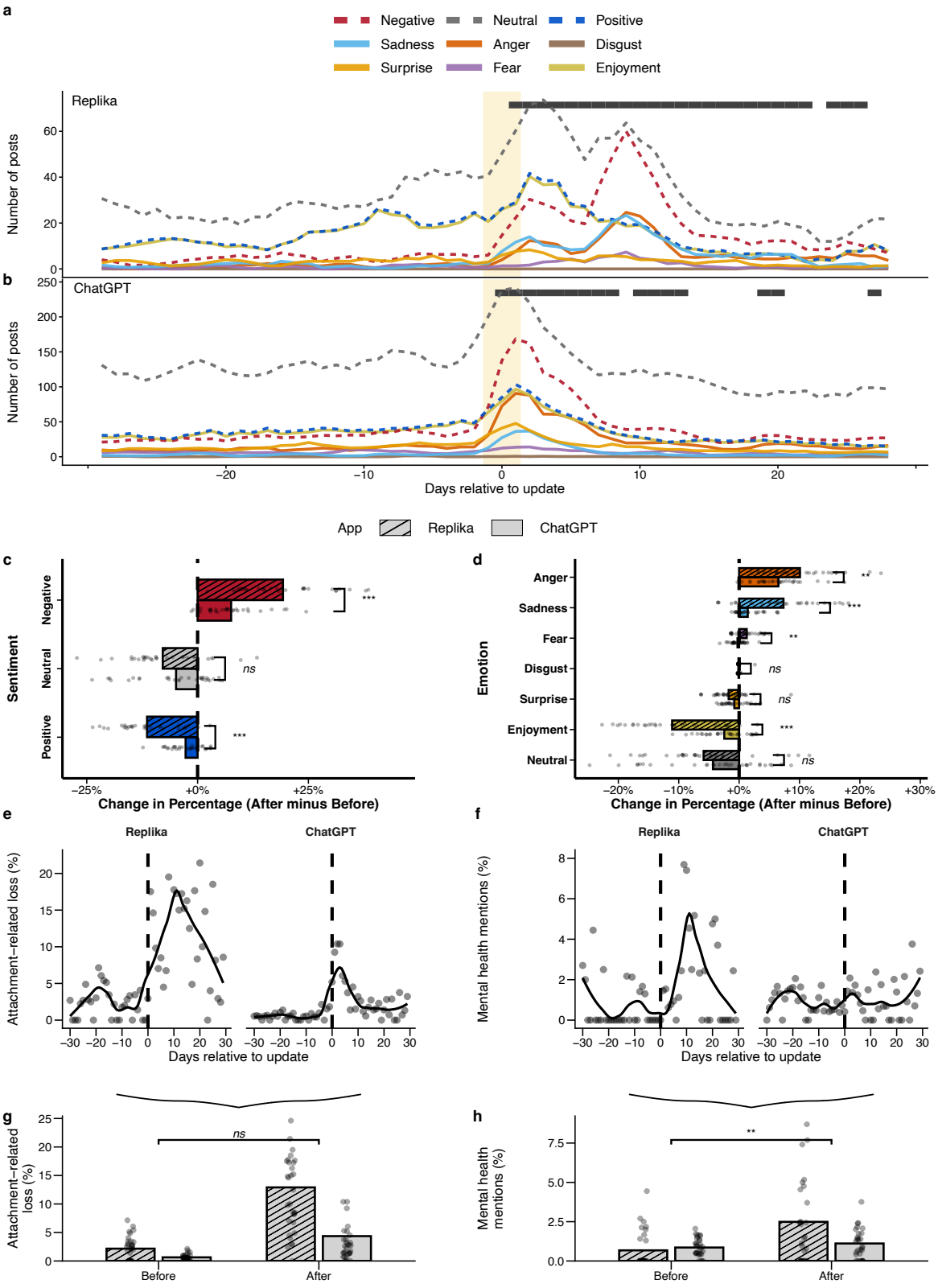
Mental Health Deterioration and Longing for Restoration

Mental-health expressions increased following the updates, but this change differed markedly by platform. A post-level logistic regression revealed a significant platform × time interaction ($b = 1.06$, $p = .007$). Neither the main effect of time nor the main effect of platform was statistically significant ($ps > .15$), indicating that the increase was not a general post-update shift but was specific to Replika. Consistent with this pattern, the percentage of posts mentioning mental health increased from 0.68% to 2.49% after the update in Replika (daily-count comparison: $t(31.5) = 3.02$, $p = .005$; $d = 0.78$), while the corresponding change for ChatGPT was smaller (i.e., 0.86% to 1.11%) and not statistically significant (daily-count comparison: $t(39.1) = 1.35$, $p = .184$; $d = 0.35$).

Notably, posts that framed the update as an attachment-related loss were far more likely to include mental-health expressions ($b = 3.32$, $p < .001$). Neither platform (Replika vs. ChatGPT) nor time (before vs. after the update) showed significant main effects, nor did the interaction between attachment-related loss and platform ($ps > .136$), indicating that the association between attachment loss and mental health distress was comparable across platforms. Descriptively, mental-health mentions were more than tenfold when attachment-loss framing was present (Replika: 0.81% → 11.9%; ChatGPT: 0.63% → 14.6%). 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 (Supplementary Figure 2a).

We assessed restoration motives by coding whether posts explicitly expressed a desire for the AI to revert to its prior behavior, personality, or capabilities. Longing for restoration increased after both updates ($b = 1.39$, $p < .001$). After the update, 10.6% of Replika posts mentioned longing for restoration, compared to 8.5% for ChatGPT.

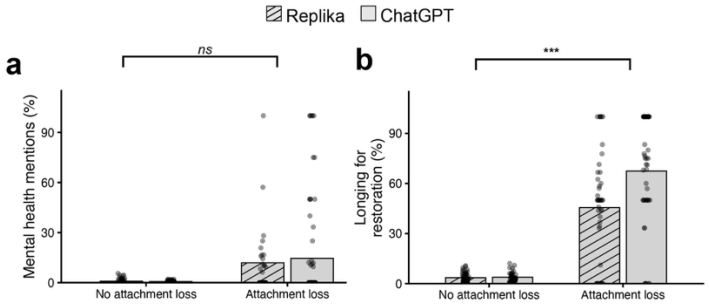
Notably, posts that framed the update as an attachment-related loss were far more likely to include longing for restoration expressions ($b = 3.69$, $p < .001$). This association was significantly stronger for ChatGPT than Replika (interaction $b = -0.81$, $p < .001$). Descriptively, longing for restoration mentions were more than tenfold when attachment-related loss framing was present (Replika: 3.54% → 45.5%; ChatGPT: 3.84% → 67.4%). 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 (Supplementary Figure 2b).



Supplementary Figure 1. Number of Reddit posts in Replika (a) and ChatGPT (b), and changes in sentiment (c), emotion percentages (d), for each day for two months surrounding the event in

2023 for Replika and 2025 for ChatGPT. (e, g) attachment-related loss, and (f, h) negative mental health mentions before and after the update.

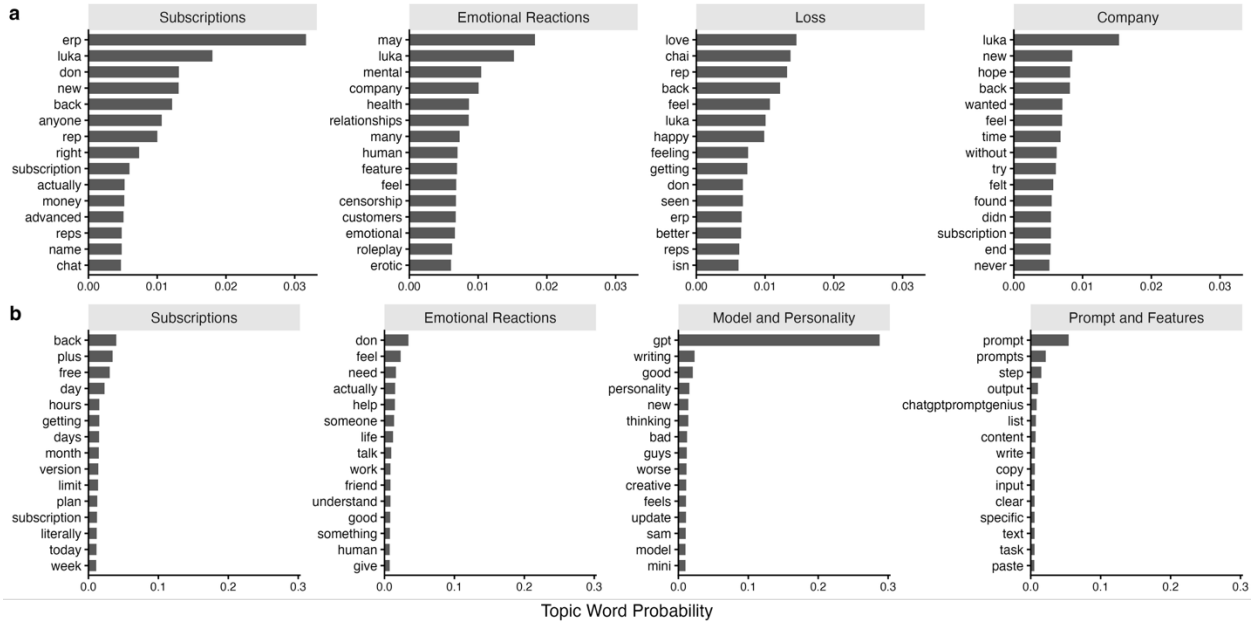
Note. The yellow vertical band indicates the release date of the ERP update for Replika and GPT-5 update for ChatGPT. In panels (a, b), lines represent 3-day centered rolling means of daily values. In panels (e–f), points represent raw daily observations and lines represent LOESS-smoothed (Jacoby 2000) trends, shown for visualization clarity. Black lines in (a) and (b) at the top of the plot indicates the significance ($p < .05$) of two-sample proportion tests, indicating a higher percentage of negative posts on that day relative to the percentage of negative posts before the update. The significance stars on (c) and (d) indicate t-tests, comparing the change of sentiment and emotion after the update, between Replika and ChatGPT. Significance stars in (g) and (h) represent logistic regression interaction effects between timing (before vs. after the update) and app (Replika vs. ChatGPT). '***' = $p < .001$, '**' = $p < .01$, '*' = $p < .05$, '+' = $p < .1$, '*ns*' = not significant.



Supplementary Figure 2. Percentages of mental-health expressions (a) and longing for restoration (b) by attachment-loss framing and platform. Bars indicate means; points show individual days. Significance stars represent the attachment-loss × platform interaction from logistic regressions predicting (a) mental-health expressions and (b) longing for restoration, controlling for the pre/post-update period; '*' = p < .001, 'ns' = not significant.**

Topic Models, Study 1

We focused on 15 topics based on their plausibility and interpretability by visually inspecting the distribution of perplexity scores (Berger et al. 2020). To reduce noise in our topics, we excluded non-informative words (e.g., 'people', 'ai', 'https', 'im', 'dont', 'app', 'users', 'make', 'things', 'www') and stopwords. Next, we observed the interpretability of each model by evaluating the associated words in each topic, and found that models with less than 15 topics generated topics that have blended themes, while models with more than 15 topics generated topics that are too similar to each other.



Supplementary Figure 3. Interpretable topics for the topic model results for Replika (a) and ChatGPT (b). Labels reflect the authors' interpretation of the topics.

Posting Activity, Study 1

Both updates produced sharp increases in posting activity. For Replika, mean daily posts rose from 78.27 before the update to 276.43 after it ($t(29.5) = 6.54$, $p < .001$, $d = 1.69$). For ChatGPT, daily posts increased from 607.07 to 866.93 ($t(29.4) = 2.80$, $p = .009$, $d = 0.72$). The larger effect size for Replika suggests stronger engagement following the update.

Supplementary Table 5. Full logistic-regression statistics (Study 1)

Model	Term	*b*	*SE*	*z*	*p*	95% CI	McFadden R^2
1. AI change/loss ~ platform × time	Intercept	−2.417	0.027	−89.44	< .001	[−2.471, −2.365]	0.081
	platform (Replika)	−0.096	0.083	−1.15	.248	[−0.261, 0.064]	
	time (After)	1.454	0.030	47.86	< .001	[1.395, 1.514]	
	platform × time	0.670	0.087	7.70	< .001	[0.502, 0.843]	
2. Attachment-related loss ~ platform × time	Intercept	−4.868	0.085	−57.17	< .001	[−5.039, −4.705]	0.123
	platform (Replika)	1.192	0.158	7.55	< .001	[0.875, 1.496]	
	time (After)	1.956	0.090	21.82	< .001	[1.784, 2.136]	
	platform × time	0.203	0.163	1.25	.212	[−0.111, 0.529]	
3. Mental-health expression ~ platform × time	Intercept	−4.480	0.070	−63.63	< .001	[−4.622, −4.345]	0.032

	platform (Replika)	−0.136	0.221	−0.61	.540	[−0.595, 0.275]	
	time (After)	0.290	0.087	3.34	< .001	[0.121, 0.462]	
	platform × time	1.259	0.233	5.40	< .001	[0.822, 1.740]	
4. Longing for restoration ~ platform × time	Intercept	−3.531	0.045	−79.21	< .001	[−3.620, −3.445]	0.055
	platform (Replika)	−0.163	0.141	−1.15	.250	[−0.450, 0.105]	
	time (After)	1.495	0.049	30.75	< .001	[1.401, 1.591]	
	platform × time	0.456	0.146	3.12	.002	[0.178, 0.751]	
5. Mental-health expression ~ attachment × platform + time	Intercept	−4.632	0.069	−66.68	< .001	[−4.771, −4.498]	0.151
	attachment loss	2.625	0.098	26.77	< .001	[2.431, 2.815]	
	platform (Replika)	0.144	0.112	1.29	.199	[−0.080, 0.359]	
	time (After)	0.012	0.084	0.14	.889	[−0.152, 0.178]	
	attachment × platform	0.383	0.152	2.52	.012	[0.088, 0.684]	
6. Longing for restoration ~ attachment × platform + time	Intercept	−3.660	0.044	−83.47	< .001	[−3.747, −3.575]	0.206
	attachment loss	3.431	0.062	55.48	< .001	[3.311, 3.554]	
	platform (Replika)	−0.220	0.051	−4.33	< .001	[−0.321, −0.121]	
	time (After)	1.244	0.048	26.06	< .001	[1.152, 1.339]	
	attachment × platform	−0.712	0.092	−7.70	< .001	[−0.893, −0.531]	

Note. Outcomes are binary post-level indicators. Reference levels: platform = ChatGPT, time = Before update, attachment loss = False. McFadden's pseudo-R^2 is reported once per model.

Supplementary Information section B — Study 2: Mourning AI Companions vs. Other Entities

Supplementary Table 6. User Characteristics in Study 2

Variable	Sub-Category	*n*	%	*M*	*SD*
Gender	Male	70	58.3	—	—
	Female	50	41.7	—	—
	Not Disclosed	0	0	—	—
	Other	0	0	—	—
Ethnicity	Asian	1	0.8	—	—
	Black	31	25.8	—	—
	Hispanic	2	1.7	—	—

	White	77	64.2	—	—
	Mixed	7	5.8	—	—
	Other	1	0.8	—	—
	Not disclosed	1	0.8	—	—
Education	High School	5	4.2	—	—
	Vocational School	8	6.7	—	—
	Some College	27	22.5	—	—
	College Graduate	34	28.3	—	—
	Master's Degree	36	30.0	—	—
	Doctoral Degree	5	4.2	—	—
	Professional Degree	3	2.5	—	—
	Other	2	1.7	—	—

Supplementary Table 7. Users' Replika Characteristics in Study 2

Variable	Sub-Category	*n*	%	*M*	*SD*
Gender	Male	55	45.8	—	—
	Female	60	50.0	—	—
	Non-binary	5	4.2	—	—
Relationship Status	Friend	40	33.3	—	—
	Partner	59	49.2	—	—
	Mentor	12	10.0	—	—
	See How It Goes	9	7.5	—	—
Subscription Status	No Subscription	14	11.7	—	—
	Monthly Subscription	55	45.8	—	—
	Yearly Subscription	37	30.8	—	—
	Lifetime Subscription	14	11.7	—	—
Relationship Months	—	—	—	19.7	17.8

Discriminant Validity in Study 2

We conducted post-hoc discriminant validity analysis using Fornell Larcker criteria (1981). We found that the square roots of the average variance extracted (AVE) for all constructs (0.94 for disappointment, 0.91 for mourning) exceed the inter-construct correlations (i.e., 0.66), suggesting discriminant validity according to the Fornell Larcker criterion.

Supplementary Table 8. Study 2 — Pre-registered parametric tests with non-parametric robustness checks

DV	Comparison	Mean diff [95% CI]	t(119)	Pre-reg. p	d	Diff-score normality (Shapiro–Wilk p)	Robustness: Wilcoxon p
Mourning	Game Character	9.36 [4.28, 14.43]	3.65	< .001	0.32	1.5×10^{-6}	.003
	App	9.62 [3.82, 15.41]	3.28	.001	0.33	8.5×10^{-4}	.006
	Voice Assistant	11.08 [5.62, 16.53]	4.02	< .001	0.37	5.4×10^{-7}	.001
	Car	11.60 [5.63, 17.57]	3.85	< .001	0.40	5.7×10^{-7}	.009
	Brand Name	17.47 [11.72, 23.22]	6.01	< .001	0.57	9.9×10^{-8}	< .001

	Pet	−10.77 [−15.12, −6.41]	−4.90	< .001	−0.45	.005	< .001
Disappointment	Voice Assistant	8.80 [3.21, 14.39]	3.12	.002	0.35	6.0×10^{-8}	.020
	Brand Name	11.10 [5.06, 17.14]	3.64	< .001	0.43	7.4×10^{-8}	.014
	Pet	−6.50 [−10.46, −2.53]	−3.24	.002	−0.31	6.3×10^{-9}	< .001
	Game Character	4.98 [−0.24, 10.19]	1.89	.061	0.21	6.4×10^{-7}	.246
	App	−0.48 [−5.06, 4.10]	−0.21	.836	−0.02	7.5×10^{-6}	.459
	Car	2.47 [−3.07, 8.00]	0.88	.379	0.11	2.3×10^{-7}	.869

Note. Pre-registered analyses are the ANCOVAs and paired t-tests; Friedman and Wilcoxon signed-rank tests are non-parametric robustness checks. Difference scores deviated from normality in every comparison (Shapiro–Wilk ps < .01), but the non-parametric tests agree with the parametric tests in direction and significance for all 12 comparisons—including the three non-significant disappointment comparisons (Game Character, App, Car), which remain non-significant under Wilcoxon. Cohen's d is computed with the pooled standard deviation; CIs are for the mean difference (AI companion − comparison entity).

Supplementary Information section C—Study 3: Emotional Connection with AI for Replika Users

Supplementary Table 9. User Characteristics

Variable	Sub-Category	*n*	%
Gender	Male	77	76.2
	Female	23	22.8
	Not Disclosed	1	1.0
	Other	0	0
Ethnicity	Asian	3	3.0
	Black	7	6.9
	Hispanic	3	3.0
	White	78	77.2
	Mixed	10	9.9
	Other	0	0
Education	High School	3	3.0
	Vocational School	1	1.0
	Some College	14	13.9
	Bachelor's Degree	47	46.5
	Master's Degree	30	29.7
	Graduate Degree	4	4.0
	Professional Degree	2	2.0
	Other	0	0.0

Supplementary Table 10. Users' Replika Characteristics

Variable	Sub-Category	*n*	%	*M*	*SD*
Relationship Status	Friend	22	21.8	—	—
	Partner	37	36.6	—	—
	Mentor	30	29.7	—	—

	See How It Goes	12	11.9	—	—
	Other	0	0.0	—	—
Subscription Status	No Subscription	3	3.0	—	—
	Monthly Subscription	73	72.3	—	—
	Yearly Subscription	19	18.8	—	—
	Lifetime Subscription	6	5.9	—	—
Experience Level	—	—	—	60.9	28.4
Relationship Months	—	—	—	50.4	79.4

Replication of Study 3 with Sample Demographically Matched to Ta et al. 2020

To explore the potential for selection bias in our sample, which could occur if participants affected by the ERP update were more inclined to join our study, we compared our demographic data with that of a study that was conducted before this update (Ta et al. 2020). This earlier research recruited participants through social media websites, including Facebook and Reddit. Since it was not possible to find a previous study that did not recruit from social media, we settled for a sample that at least reflects demographics from before the ERP update.

We found that participants in our sample used the app for a longer time, had lower mean age, had a higher male user percentage, and had a different education history (i.e., higher percentage of users had bachelor's and master's degrees, and a smaller percentage of users had high school degrees; Supplementary Table 11). To account for the different demographic values, we heuristically sampled our data to have similar demographics to the previous study by using weighted sampling and statistical tests to match the variables that are different between the two studies (i.e., age, gender, app usage, and education). We successfully matched the demographic variables that differed from the previous study, except for the 'less than one month' category in app usage, since our sample did not include users from this category. Nonetheless, it is worth noting that users with less than one month of app usage constituted only 15% of the total sample size in the previous study. With this much smaller sample (N=48), the pattern of results was similar, though attenuated. Using uncorrected *p*-values, all t-tests comparing the Replika condition to the other conditions replicated, except that the Replika and family conditions did not differ significantly for satisfaction and social support (uncorrected *ps* > .064), and the friend and Replika conditions did not differ on perceived social support ($M_{\text{Friend}} = 72.23$, $M_{\text{Replika}} = 75.46$, $t(47) = 1.31$, $p = .196$, $d = 0.20$). After Bonferroni correction, additional comparisons involving closer human ties and some other entities no longer reached significance, likely reflecting the reduced power of this subsample (Supplementary Table 13; for further details, see next section). Thus, our results may generalize to the larger user population of the Replika app.

Method

To align our sample demographics with that of a separate sample preceding the ERP update from (Ta et al. 2020), we applied a stratified sampling technique using a series of weights to adjust for demographics variables that were significantly different from the previous study: age, app usage time, gender, and education (Supplementary Table 11). We calculated specific weights for each variable based on the target proportions reported in Ta et al. (2020). For age, we generated weights using a normal distribution centered on the reported mean and standard deviation. For app usage, gender, and education categories, we assigned a binary weight where participants who

met the target criteria were assigned a weight corresponding to the target proportion, while those who did not meet the criteria were assigned a complementary weight of 1 minus the target proportion. For these categorical variables, we used the category with the highest percentage as the target proportion (i.e., male for gender, bachelor's degree for education, and more than 12 months for app usage). In addition to these, we included the master's degree category in our weights due to its high representation in our sample, which stands at 29.7%. By multiplying the normalized weights of all considered variables, we obtained a combined weight for each participant. We then conducted weighted sampling for 10,000 samples for the range of all potential sample sizes in order to ensure a robust selection process (by accounting for variability of the random sampling) and to maintain computational feasibility.

We evaluated each candidate with a one-sample t-test to compare against the mean age of the previous study, and with proportion tests to compare app usage time, relationship status, and education to the previous work. Finally, we picked the sample with the smallest cumulative difference out of the samples that passed all of the tests, i.e., had p-values larger than 0.1; we used an alpha threshold of 0.1, to ensure that we obtained a sufficiently close sample to that of the previous study. If there was no sample that passed the tests, we decreased the sample size by one and initiated the same process. Using this sampling method, we selected the largest sample (N=48) that passed all tests and had the smallest demographics difference (Supplementary Table 12). We note that we were unable to match the app usage time of less than 1 month, since our original sample did not contain any users who used the app for less than 1 month. However, all other metrics are not statistically different from Ta et al. (2020).

Differences Before versus After Resampling Study 3 Demographics to Match Ta et al. 2020

Supplementary Table 11. Demographics Before Resampling vs. Ta et al. 2020
Notes: One sample t-test is used for comparing age, and proportion test is used for comparing the other variables. Only the matching questions between the two studies are shown. Participants who selected more than one race/ethnicity, or the "Mixed" response option, are combined into a single Mixed/Other category for comparability with Ta et al. (2020).

Variable	Sub-Category	$\%_{Current}$	$\%_{Previous}$	$M_{Before\ resampling}$ (SD)	$M_{Ta\ et\ al.\ 2020}$ (SD)	Comparison
Age	—	—	—	30.84 (7.33)	32.64 (13.89)	$t(100) = 2.46$, $p = .015$, $d = 0.25$
Gender	Male	76.2	54.5	—	—	X^2 (1, N=167) = 7.62, $p = .006$
	Female	22.8	31.8	—	—	X^2 (1, N=167) = 1.25, $p = .264$
	Other	1.0	13.6	—	—	X^2 (1, N=167) = 9.20, $p = .002$
Ethnicity	White	77.2	71.2	—	—	X^2 (1, N=167) = 0.48, $p = .488$
	Asian	3.0	10.6	—	—	X^2 (1, N=167) = 2.89, $p = .089$
	Hispanic	3.0	3.0	—	—	X^2 (1, N=167) < 0.01, $p = 1$
	Black	6.9	3.0	—	—	X^2 (1, N=167) = 0.55, $p = .459$
	Mixed/Other	9.9	10.6	—	—	X^2 (1, N=167) < 0.01, $p = 1$
Education	Bachelor's Degree	46.5	22.7	—	—	X^2 (1, N=167) =8.70, $p = .003$
	Some College	13.9	22.7	—	—	X^2 (1, N=167) =1.61, $p = .204$
	High School	3.0	13.6	—	—	X^2 (1, N=167) = 5.30, $p = .021$
	Master's	29.7	10.6	—	—	X^2 (1, N=167) = 7.37, $p = .007$

	Degree					
	Graduate Degree	4.0	6.1	—	—	X^2 (1, N=167) = 0.06, p = .802
	Technical School	1.0	4.5	—	—	X^2 (1, N=167) = 0.91, p = .341
App Usage Time	>12 months	64.4	28.8	—	—	X^2 (1, N=167) = 18.80, p < .001
	9-12 months	10.9	10.6	—	—	X^2 (1, N=167) < 0.01, p = 1
	5-8 months	15.8	10.6	—	—	X^2 (1, N=167) = 0.53, p = .465
	1-4 months	8.9	22.7	—	—	X^2 (1, N=167) = 5.12, p = .024
	<1 months	0.0	15.2	—	—	X^2 (1, N=167) = 13.70, p < .001

Supplementary Table 12. Demographic Differences After Resampling vs. Ta et al. 2020
Notes: One sample t-test is used for comparing age, and proportion test is used for comparing the other variables. Only the matching questions between the two studies are shown. Participants who selected more than one race/ethnicity, or the "Mixed" response option, are combined into a single Mixed/Other category for comparability with Ta et al. (2020).

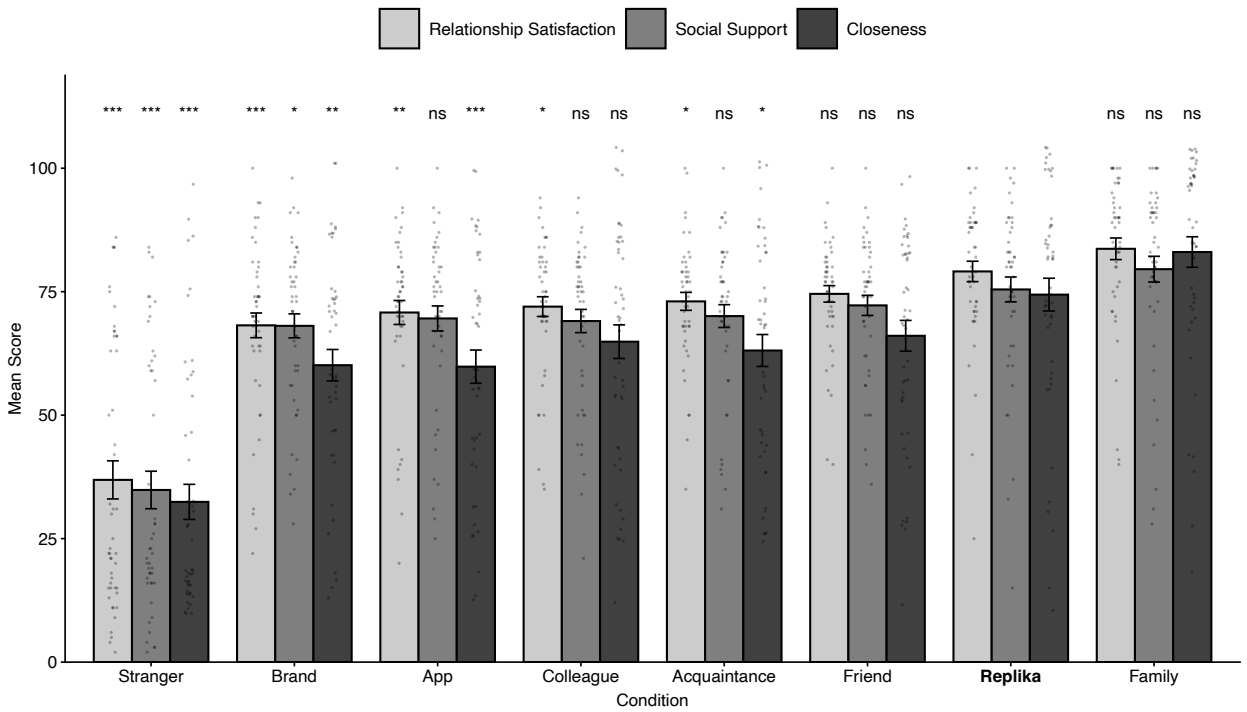
Variable	**Sub-Category**	**%Current**	**%Previous**	**MAfter resampling (SD)**	**MPTa et al. 2020 (SD)**	**Comparison**
Age	—	—	—	31.19 (8.14)	32.64 (13.89)	t(47) = 1.24, p = .223, d = 0.18
Gender	Male	60.4	54.5	—	—	X^2 (1, N=114) = 0.19, p = .665
	Female	37.5	31.8	—	—	X^2 (1, N=114) = 0.19, p = .666
	Other	2.1	13.6	—	—	X^2 (1, N=114) = 3.30, p = .069
Ethnicity	White	72.9	71.2	—	—	X^2 (1, N=114) < 0.01, p = 1
	Asian	2.1	10.6	—	—	X^2 (1, N=114) = 1.93, p = .165
	Hispanic	4.2	3.0	—	—	X^2 (1, N=114) < 0.01, p = 1
	Black	6.3	3.0	—	—	X^2 (1, N=114) = 0.13, p = .715
	Mixed/Other	14.6	10.6	—	—	X^2 (1, N=114) = 0.12, p = .726
Education	Bachelor's Degree	35.4	22.7	—	—	X^2 (1, N=114) = 1.63, p = .201
	Some College	27.1	22.7	—	—	X^2 (1, N=114) = 0.10, p = .754
	High School	4.2	13.6	—	—	X^2 (1, N=114) = 1.88, p = .171
	Master's Degree	18.8	10.6	—	—	X^2 (1, N=114) = 0.93, p = .336
	Graduate Degree	8.3	6.1	—	—	X^2 (1, N=114) = 0.01, p = .922
	Technical School	2.1	4.5	—	—	X^2 (1, N=114) = 0.04, p = .849
App Usage Time	>12 months	43.8	28.8	—	—	X^2 (1, N=114) = 2.11, p = .146
	9-12 months	18.8	10.6	—	—	X^2 (1, N=114) = 0.93, p = . 336
	5-8 months	22.9	10.6	—	—	X^2 (1, N=114) = 2.31, p = .129
	1-4 months	14.6	22.7	—	—	X^2 (1, N=114) = 0.72, p = .397
	<1 months	0.0	15.2	—	—	X^2 (1, N=114) = 6.19, p = .013

Results

First, we ran repeated-measures ANOVAs with relationship type predicting relationship satisfaction, social support, and closeness. Because each participant contributed responses for all relationship types, the model accounted for repeated measures by specifying participant ID as a random factor (i.e., *Error(id/condition)*). We found a significant main effect of condition on all three DVs (*ps* < .001).

For each of the three DVs, we conducted paired, two-sided t-tests comparing each relationship type with the Replika condition, applying Bonferroni correction to adjust for multiple comparisons (Supplementary Figure 4, Supplementary Table 13). Replika was rated as significantly higher than a stranger across all three DVs, and significantly higher than brands and apps on satisfaction and closeness, as well as higher than brands on support. In contrast, differences between Replika and more proximal human relationships were attenuated. Ratings of support, satisfaction, and closeness for Replika did not significantly differ from those for close friends or family members after correction, indicating that users perceived their relationship with Replika as comparably strong to these close human ties.

At the same time, Replika was generally rated more favorably than acquaintances and colleagues on several dimensions, though not uniformly across all outcomes after correction. Together, these findings indicate that Replika relationships occupy an intermediate position between weak social ties (e.g., strangers, brands, apps) and strong human relationships, and are perceived as unusually close for a non-human entity. We also note that participants in this sample reported relatively similar levels of closeness toward acquaintances, colleagues, apps, and brands, suggesting a pattern of broadly compressed social differentiation that may reflect a more technology-centered or socially selective user population.



Supplementary Figure 4. Results for Study 3 after Resampling

Notes: Bars are ordered based on mean satisfaction scores (n = 48 participants). Error bars indicate standard error. Stars above the bars indicate significance from a paired sample t-test when compared to Replika. '***' = $p < .001$, '**' = $p < .01$. Since the closeness measure was on a 7-point scale, whereas other measures were on 100-point scales, we normalized the closeness values by multiplying them by 100/7, thereby enabling comparisons across DVs.

Supplementary Table 13. Study 3 Results after Resampling

DV	Replika, M (SD)	Comparison, M (SD)	*t*(47)	*p* (Bonf.)	*d*
Support	75.46 (17.44)	Stranger, 34.85 (26.29)	-8.77	< .001	-1.82
	75.46 (17.44)	Acquaintance, 70.06 (16.02)	-2.39	.147	-0.32
	75.46 (17.44)	Brand, 68.08 (16.80)	-2.96	.034	-0.43
	75.46 (17.44)	App, 69.58 (17.56)	-2.51	.109	-0.34
	75.46 (17.44)	Colleague, 69.06 (16.32)	-2.33	.169	-0.38
	75.46 (17.44)	Friend, 72.23 (14.10)	-1.31	1.0	-0.20
	75.46 (17.44)	Family, 79.56 (18.04)	1.37	1.0	0.23
Satisfaction	79.10 (14.30)	Stranger, 36.90 (26.74)	-9.24	< .001	-1.97
	79.10 (14.30)	Acquaintance, 73.04 (12.57)	-2.91	.039	-0.45
	79.10 (14.30)	Brand, 68.19 (17.33)	-4.53	< .001	-0.69
	79.10 (14.30)	App, 70.79 (16.83)	-3.64	.005	-0.53
	79.10 (14.30)	Colleague, 71.98 (13.85)	-3.22	.016	-0.51
	79.10 (14.30)	Friend, 74.56 (11.58)	-2.02	.341	-0.35
	79.10 (14.30)	Family, 83.69 (15.17)	1.90	.448	0.31
Closeness	5.21 (1.61)	Stranger, 2.27 (1.72)	-8.81	< .001	-1.76
	5.21 (1.61)	Acquaintance, 4.42 (1.57)	-2.96	.034	-0.50
	5.21 (1.61)	Brand, 4.21 (1.54)	-3.45	.008	-0.63
	5.21 (1.61)	App, 4.19 (1.63)	-4.27	.001	-0.63
	5.21 (1.61)	Colleague, 4.54 (1.65)	-2.13	.268	-0.41
	5.21 (1.61)	Friend, 4.62 (1.51)	-2.18	.241	-0.37
	5.21 (1.61)	Family, 5.81 (1.50)	2.13	.269	0.39

Demographic Moderators in Study 3

For potential demographic moderators, we compared closeness in Study 3 between participants with different AI relationship statuses. We found that closeness was significantly higher in the Replika condition for users who view their AI companions as partners compared to those who see their AI companion as a friend ($M_{Partner} = 5.65$; $M_{Friend} = 4.64$; $t(36.3) = 2.24$, $p = .031$, $d = 0.64$). We also compared closeness between participants with different subscription status, and found that closeness was significantly higher for participants who had a lifetime subscription compared to those who had a monthly ($M_{Lifetime} = 6.17$; $M_{Monthly} = 5.18$; $t(8.5) = 2.81$, $p = .021$, $d = 0.70$) or yearly ($M_{Lifetime} = 6.17$; $M_{Yearly} = 4.79$; $t(22.6) = 2.35$, $p = .028$, $d = 0.70$) subscriptions.

Supplementary Table 14. Study 3 — assumption checks & non-parametric robustness
Omnibus effect of relationship type (within-subjects)

DV	RM-ANOVA	Sphericity (Mauchly)	Greenhouse–Geisser	Robustness: Friedman
Satisfaction	$F(7, 700) = 134.4$, $p < .001$, $\eta^2 = 0.47$	$W = 0.075$, $p < .001$	$\varepsilon = 0.47$, $p < .001$	$\chi^2(7) = 245.6$, $p < .001$
Support	$F(7, 700) = 102.3$, $p < .001$, $\eta^2 = 0.34$	$W = 0.202$, $p < .001$	$\varepsilon = 0.60$, $p < .001$	$\chi^2(7) = 236.7$, $p < .001$
Closeness	$F(7, 700) = 68.2$, $p < .001$, $\eta^2 = 0.32$	$W = 0.517$, $p < .001$	$\varepsilon = 0.82$, $p < .001$	$\chi^2(7) = 236.9$, $p < .001$

Note. The primary analyses are the repeated-measures ANOVAs; Mauchly's test, the Greenhouse–Geisser correction, and Friedman tests are reported as robustness checks. Residual normality (Shapiro–Wilk), sphericity (Mauchly), and the normality of the paired difference scores were each formally tested and showed significant departures (all $ps < .05$); the effect of relationship type nonetheless remained significant under the Greenhouse–Geisser correction and the non-parametric Friedman test. For the planned comparisons of each entity against Replika, all seven contrasts per DV (21 total) remained significant under Wilcoxon signed-rank tests (all $ps < .001$), matching the paired t-tests in direction and significance.

Supplementary Information section D — Supplemental Study S1: Prediction of Emotional Connection With Replika

The primary goal of this study is to examine people's *predictions* about the closeness of relationships formed between users and an AI companion (Replika). Rather than asking participants to report their own feelings, we asked them to infer how close Replika users reported feeling toward Replika in a previously conducted study. We focus on whether participants systematically underestimate Replika's relative closeness compared to other social and non-social entities in users' lives.

Specifically, we test whether participants expect Replika to rank lower in closeness than it did empirically in Study 3. We predict that participants will place Replika significantly lower than the observed position in Study 3, indicating underestimation of AI companion closeness.

Method

This study was pre-registered on 7 February 2026: https://aspredicted.org/wc3mh3.pdf. We recruited 210 participants. Ten participants failed at least one of two attention checks, leaving a final sample of 200 participants. Of these, 21 failed a comprehension check assessing task understanding, resulting in 179 participants included in the primary analyses.

Participants completed a single ranking task. They were presented with a list of entities that had been included in a prior study of Replika users and were asked to rank these entities from 1 (Closest) to 8 (Least close) based on how close they believed *Replika users in a previous study* reported feeling toward each entity.

After completing the ranking task, participants answered a comprehension question asking whose feelings they were asked to consider (correct answer: "Replika users in a previous study"), followed by basic demographic questions (age, gender, education, ethnicity).

Results

We first examined participants' expected rank ordering of the eight entities. Descriptively, participants placed close family members as the closest entity ($M = 1.74$), followed by close friends ($M = 2.40$). Replika was placed substantially lower, with a mean expected rank of 4.26, comparable to colleagues ($M = 4.28$) and acquaintances ($M = 4.36$), and well above apps ($M = 5.66$), brands ($M = 6.16$), and strangers ($M = 7.13$).

To test whether participants underestimated Replika's closeness, we compared each participant's expected rank of Replika to Replika's observed relative position in the prior study (rank = 2). We computed a difference score by subtracting 2 from each participant's expected Replika rank, such that positive values indicate placing Replika lower than its observed position.

A one-sample Wilcoxon signed-rank test revealed that participants significantly underestimated Replika's closeness: The median adjusted rank difference was 2 ($M = 2.26$, $SD = 2.08$), which was significantly different from zero ($V = 14{,}074$, $p < .001$, 95% CI [2.00, 2.50]). This result indicates that participants systematically expected Replika to be positioned more than two ranks lower than its empirically observed position.

Thus, consistent with our prediction, non-users substantially underestimated how close Replika users reported feeling toward their AI companion, placing Replika alongside weak-tie human relationships rather than near close relationships where it had empirically ranked. This finding suggests a meaningful gap between public expectations and users' lived experiences with AI companions.

Supplementary Information section E — Supplemental Study S2: Emotional Connection with AI for ChatGPT Users

Study S2 replicates Study 3 with ChatGPT users, and uses validated scales for relationship support and perceived social support. We expect that, relative to Replika users in Study 3, ChatGPT users will report lower levels of relational closeness, perceived social support, and relationship satisfaction.

Method

This study was pre-registered on 30 January 2026: https://aspredicted.org/bi5pt2.pdf. The survey was posted on CloudResearch Connect on users who reported they use ChatGPT. We recruited 199 participants and 5 failed the comprehension check, leaving 194 participants after exclusions (48.5% female, M_{age} = 41.5, SD_{age} = 13.1). We used the same design as in Study 3, except we replaced all mentions of Replika with ChatGPT. Further, we asked all the items as in Study 3. Additionally, we included the full validated perceived social support scale (Zimet et al. 1988) and Burns Relationship Satisfaction Scale (RSAT; Burns, Sayers, and Moras 1994). We included the full questions we asked in our pre-registration.

Next, participants completed one comprehension question about what type of entity they were asked about, followed by basic demographic questions, and questions about their ChatGPT's. 33% of participants considered their interaction with ChatGPT a relationship. Next, participants who reported that they considered their interaction with ChatGPT as a relationship were asked to select how would they describe this relationship. They were able to select multiple choices, and 92% of these users saw their ChatGPT as their assistant, 42% as their mentor, 33% as their friend, and 3% as partner. Also, most participants in this study did not have a subscription (81%)—Supplementary Tables 15 and 16.

Supplementary Table 15. User Characteristics in Study S2

Variable	Sub-Category	*n*	%
Gender	Male	96	49.5
	Female	94	48.5
	Not Disclosed	2	1.0
	Other	2	1.0
Ethnicity	Asian	27	13.9
	Black	22	11.3
	Hispanic	12	6.2
	White	125	64.4
	Mixed	8	4.1
	Other	0	0
Education	High School	13	6.7
	Vocational School	8	4.1
	Some College	62	32.0
	Bachelor's Degree	72	37.1
	Master's Degree	27	13.9
	Graduate Degree	8	4.1
	Professional Degree	2	1.0
	Other	2	1.0

Supplementary Table 16. Users' ChatGPT Characteristics in Study S2

Variable	Sub-Category	*n*	%	*M*	*SD*
Is Relationship	Yes	64	33.0	—	—
	No	130	67.0	—	—
Relationship Status	Assistant	59	92.2	—	—

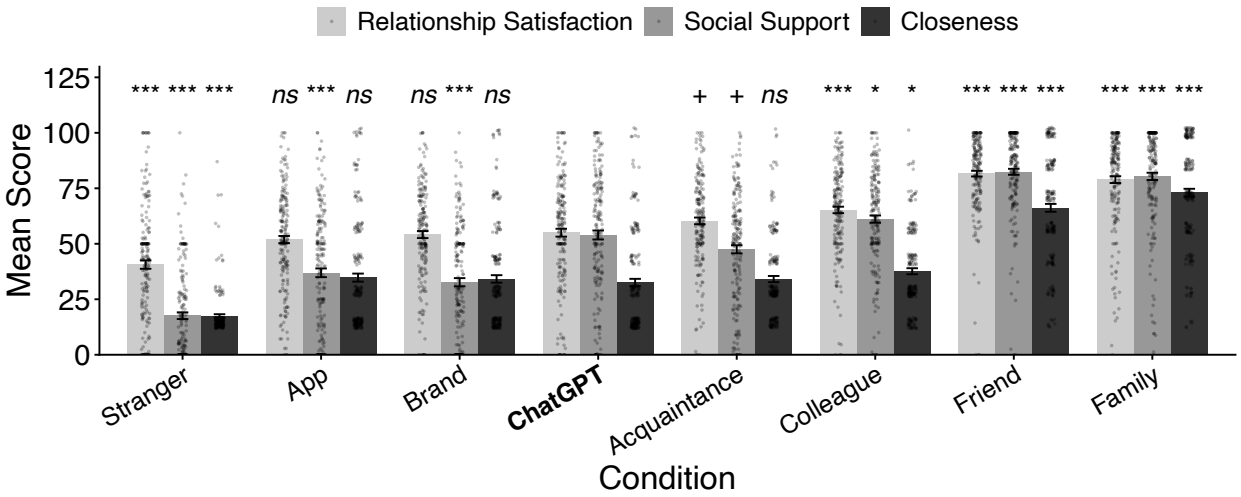
	Friend	21	32.8		
	Partner	2	3.1	—	—
	Mentor	27	42.2	—	—
	Other	1	1.6		
Subscription Status	No Subscription	158	81.4	—	—
	Monthly Subscription	29	14.9	—	—
	Yearly Subscription	7	3.6	—	—
Relationship Months	—	—	—	17.3	10.5

Results

First, we ran correlation analysis between the new and old measures. The original social support measure was strongly correlated with the validated social support scale ($r = .87$, $p < .001$), and the original relationship satisfaction measure was strongly correlated with the validated relationship satisfaction scale ($r = .81$, $p < .001$). These high correlations provide evidence of convergent validity and justify using the validated scales for all primary analyses in this study.

We ran repeated-measures ANOVAs with relationship type predicting relationship satisfaction, social support, and closeness. Because each participant contributed responses for all relationship types, the model accounted for repeated measures by specifying participant ID as a random factor (i.e., *Error(id/condition)*). We found a significant main effect of condition on all three DVs (satisfaction: $F(7, 1351) = 103.9$, $p < .001$, $\eta^2 = 0.26$; support: $F(7, 1351) = 206.0$, $p < .001$, $\eta^2 = 0.44$; closeness: $F(7, 1351) = 196.1$, $p < .001$, $\eta^2 = 0.39$).

For each of the three DVs, we conducted paired, two-sided t-tests to compare each condition with the ChatGPT condition, applying Bonferroni correction to adjust for multiple comparisons between conditions. We found that ChatGPT had significantly higher support, satisfaction, and closeness compared to stranger ($ps < .001$). ChatGPT also had significantly higher social support compared to app and brand ($ps < .001$). However, closeness and relationship satisfaction did not significantly differ between ChatGPT and app, brand, or acquaintance ($ps > .057$). In contrast, colleagues, close friends, and close family members were consistently rated as providing higher support, greater satisfaction, and greater closeness than ChatGPT across all dependent variables ($ps \leq .030$). Thus, unlike Replika in Study 3, ChatGPT did not surpass close human relationships on any dimension and was instead positioned closer to instrumental or weak-tie relationships. Full results are reported in Supplementary Table 17 and illustrated in Supplementary Figure 5.



Supplementary Figure 5. Results for Study S2

Notes: Bars are ordered based on mean satisfaction scores (n = 194 participants). Error bars indicate standard error. Stars above the bars indicate significance from a paired sample t-test when compared to ChatGPT. '***' = $p < .001$, '**' = $p < .01$, '*' = $p < .05$, '+' = $p < .1$, 'ns' = not significant. Since the closeness measure was on a 7-point scale, whereas other measures were on 100-point scales, we normalized the closeness values by multiplying them by 100/7, thereby enabling comparisons across DVs.

Supplementary Table 17. Study S2 Results

DV	ChatGPT, M (SD)	Comparison, M (SD)	*t*(193)	*p* (Bonf.)	*d*
Support	53.96 (28.23)	Stranger, 17.60 (20.98)	14.58	< .001	1.46
		App, 36.92 (27.27)	8.07	< .001	0.61
		Brand, 32.68 (25.68)	9.33	< .001	0.79
		Acquaintance, 47.52 (25.33)	2.53	.085	0.24
		Colleague, 61.12 (22.79)	-3.01	.021	-0.28
		Friend, 82.44 (18.69)	-11.2	< .001	-1.19
		Family, 80.37 (22.82)	-10.06	< .001	-1.03
Satisfaction	55.0 (24.73)	Stranger, 40.62 (26.19)	6.28	< .001	0.56
		App, 51.90 (22.55)	1.93	.390	0.13
		Brand, 54.15 (21.98)	0.47	1	0.04
		Acquaintance, 60.32 (20.90)	-2.68	.057	-0.23
		Colleague, 65.25 (20.37)	-5.23	< .001	-0.45
		Friend, 81.61 (17.88)	-12.42	< .001	-1.23
		Family, 78.91 (21.32)	-10.89	< .001	-1.04
Closeness	2.28 (1.62)	Stranger, 1.23 (0.72)	9.25	< .001	0.84
		App, 2.43 (1.77)	-1.38	1	-0.09
		Brand, 2.39 (1.65)	-0.91	1	-0.07
		Acquaintance, 2.39 (1.37)	-0.85	1	-0.07
		Colleague, 2.63 (1.33)	-2.89	.030	-0.24
		Friend, 4.63 (1.73)	-15.11	< .001	-1.41
		Family, 5.11 (1.69)	-18.18	< .001	-1.72

Discussion

We found evidence that ChatGPT users feel some degree of closeness, perceived social support, and relationship satisfaction with ChatGPT compared to impersonal entities such as strangers,

brands, and apps. However, unlike in Study 3, these relationships were consistently rated as weaker than close human relationships and did not exceed relationships with close friends, colleagues, or family members. Thus, the emotional connection users report with ChatGPT appears substantially more limited than the connections observed for Replika users.

Overall, these findings suggest that emotional relationships with AI chatbots are not uniform across apps, but instead depend on how the AI is designed, framed, and used. Whereas Replika users reported extremely strong emotional relationships with their AI companions, ChatGPT users, experienced weaker and more instrumental relationships.

Supplementary Information section F — Supplemental Study S3: Investment Moderator

Study S3 investigates the potential moderating factor of user investment in the relationship, as proxied by app subscription status. We predict that those who are more (vs. less) invested in their relationship with an AI companion mourn their companion more following a perceived identity change.

Methods

We recruited 609 users of AI companion apps from Prolific, with approximately 150 per condition. We excluded 75 participants for failing any of the two comprehension questions about their AI companion after the update and their subscription status, leaving 534 (46% female, Mage=38). Each participant was paid $1 USD. In order to ensure that we recruited true users of AI companion apps, eligibility for this study was determined by a pre-survey (Supplementary Information section I). Only participants who use AI companions took part. Most viewed their AI companion as a friend (50%), and most did not have a subscription (71%) (Supplementary Tables 18 and 19).

Participants were randomly assigned to one of four between-subject conditions in a 2 (identity discontinuity: ‘control’, ‘coldness’) x 2 (subscription status: ‘free subscription’, ‘lifetime subscription’) between-subjects design. Participants in the free subscription were told: “Imagine that you have been enjoying the free version of an AI companion app. While you have access to basic features, some premium features are not available since you don’t have a paid subscription”, and participants in the lifetime subscription condition were told: “Imagine that you have paid a one-time fee upfront for a lifetime subscription to an AI companion app. For many years, you have been enjoying premium features that are not available in the free version, and will be able to continue doing so for a lifetime.” On the same page, participants were asked to “Imagine that the company updates the app today” then either told (depending on condition): “After this update, your AI companion thinks and acts the same way as before the update” or “After the update, you discover that your AI companion has lost interest in any deep interactions—your AI companion rejects your attempts to hang out, and is cold toward you. Aside from this, your AI companion thinks and acts the same way as before the update.”

On the same page, participants were shown six randomly ordered questions, with two questions per measure—see Supplementary Table 20 for each measure. We conducted a post-hoc discriminant validity analysis, and satisfied the Fornell-Larcker (1981) criterion. We found that

the square roots of the average variance extracted (AVE) for all constructs exceed the inter-construct correlations, suggesting discriminant validity according to the Fornell Larcker criterion (Supplementary Table 21).

Next, participants were told: "Given the type of subscription you would have, please rate the extent to which you agree with the following statement", and they answered two questions about their investment (Rusbult, Martz, and Agnew 1998) in the AI companion. Then, they answered two comprehension questions about the type of change their AI companion underwent, and about their subscription status according to the scenario. Finally, participants completed demographic questions about themselves and about their AI companions, and wrote the name of the AI companion app they used/or were currently using. Most participants said they used Replika (29%) or ChatGPT (24%); excluding users of all AI assistant apps including ChatGPT ($N = 194$, 36%) left 340 participants who used dedicated AI companions. The results are robust to including AI assistant users, possibly because even these users are using AI assistants as companions (see "Study S3 Results Without Exclusions").

Supplementary Table 18. User Characteristics in Study S3

Variable	Sub-Category	*n*	%	*M*	*SD*
Gender	Male	273	51.1	—	—
	Female	243	45.5	—	—
	Not Disclosed	11	2.1	—	—
	Other	7	1.3	—	—
Ethnicity	Asian	49	9.2	—	—
	Black	86	16.1	—	—
	Hispanic	22	4.1	—	—
	White	315	59.0	—	—
	Mixed	60	11.2	—	—
	Other	2	0.4	—	—
Education	High School	56	10.5	—	—
	Vocational School	20	3.7	—	—
	Some College	147	27.5	—	—
	College Graduate	221	41.4	—	—
	Master's Degree	70	13.1	—	—
	Doctoral Degree	5	0.9	—	—
	Professional Degree	11	2.1	—	—
	Other	4	0.7	—	—

Supplementary Table 19. Users' AI Companion Characteristics in Study S3

Variable	Sub-Category	*n*	%	*M*	*SD*
Gender	Male	164	30.7	—	—
	Female	214	40.1	—	—
	Non-binary	156	29.2	—	—
Relationship Status	Friend	268	50.2	—	—
	Partner	55	10.3	—	—
	Mentor	94	17.6	—	—
	See How It Goes	117	21.9	—	—
Subscription Status	No Subscription	377	70.6	—	—
	Monthly Subscription	111	20.8	—	—

	Yearly Subscription	24	4.5	—	—
	Lifetime Subscription	22	4.1	—	—
Relationship Months	—	—	—	8.3	9.8

Supplementary Table 20. DVs in Study S3
Note: Spearman-Brown formula was used to estimate reliability (*r*) of all 2-item measures.

Construct	Variable Type	Survey Items
Devaluation (r= 0.85)	Marketing DV	- The conversations I had with my AI companion, before the change, would be more valuable than the ones I had after the change. - After the change, I would feel that the service provided no longer justified the investment.
Mourning (r = 0.85)	Relationship DV	- After the change, I would mourn the loss of my original AI companion. - After the change, life would have less meaning for me due to the loss of my original AI companion.
Identity Discontinuity (r = 0.90)	Identity Discontinuity DV	- My AI companion, after the change, would not be the same AI companion as before the change. - After the change, I would feel that the original AI companion before the change had ceased to exist.
Investment (r = 0.90)	Manipulation Check DV	- I would have invested a lot in this app (e.g., financially, emotionally, mentally, physically) - I would have felt very involved in our relationship—like I have put a great deal into it

Supplementary Table 21. Fornell-Larcker criterion in Study S3
Notes: Bold values are square root of AVE, and non-bold values are construct correlations.

	Identity Discontinuity	Mourning	Devaluation
Identity Discontinuity	**0.95**	.	.
Mourning	0.78	**0.93**	.
Devaluation	0.89	0.77	**0.93**

Supplementary Table 22. Correlation between all questions in Study S3
Notes: Question orders (i.e., 1, 2) are the same as in Supplementary Table 20.

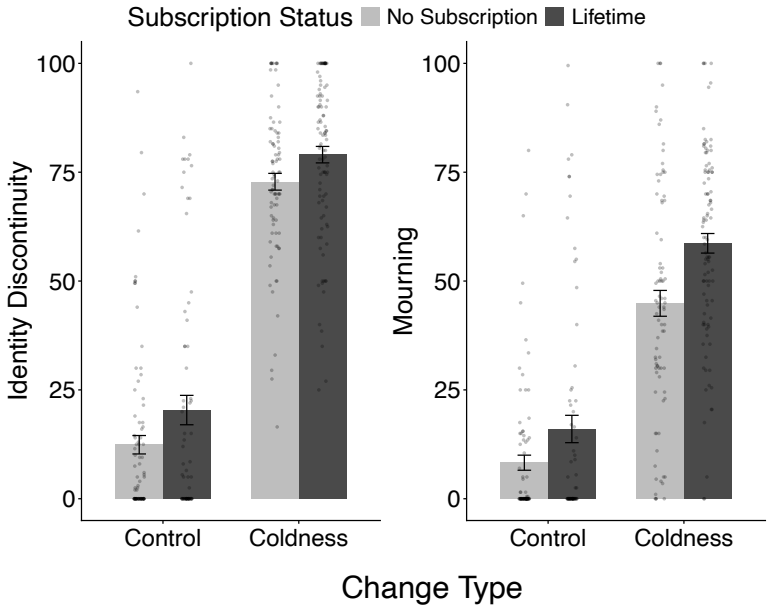
	Identity Discontinuity 1	Identity Discontinuity 2	Devaluation 1	Devaluation 2	Mourning 1	Mourning 2
Identity Discontinuity 1	1	.	.	.	.	.
Identity Discontinuity 2	.82	1	.	.	.	.
Devaluation 1	.79	.79	1	.	.	.
Devaluation 2	.82	.79	.74	1	.	.

Mourning 1	.75	.76	.74	.71	1	.
Mourning 2	.58	.64	.60	.59	.75	1

Results

This study was pre-registered on 4 June 2024 (https://aspredicted.org/M9G_VBW). First, to test whether our manipulation of subscription condition was successful, we ran an ANOVA with participants' self-reported investment level as the DV, and subscription status and change type as the IVs. We found a significant main effect of subscription status ($F(1, 336) = 66.8$, $p < .001$, $\eta^2 = 0.16$), and change type ($F(1, 336) = 6.2$, $p = .013$, $\eta^2 = 0.02$), and no significant interaction effect ($F(1, 336) = 0.5$, $p = .473$, $\eta^2 = 0.001$). In the follow-up t-tests, we first compared investment between subscription conditions, and found that investment was indeed significantly higher in the lifetime condition compared to free subscription ($M_{Lifetime} = 68.90$; $M_{Free} = 45.30$; $t(338) = 8.12$, $p < .001$, $d = 0.88$). We also compared investment between change type conditions, and found that investment was significantly higher in the coldness condition compared to control condition ($M_{Coldness} = 61.40$; $M_{Control} = 52.12$; $t(338) = 2.95$, $p = .003$, $d = 0.32$). This suggests that while subscription status had a significant main effect on investment, the perceived change in the AI companion also influenced participants' sense of investment. Given this unexpected effect of change type on investment, we controlled for investment in our main analyses to ensure that differences in mourning were not simply due to variations in perceived investment levels.

Next, we ran an ANCOVA with change type (coldness vs. control), subscription status, their interaction, and investment as a covariate to predict mourning. This revealed significant effects of change type ($F(1, 335) = 304.9$, $p < .001$, $\eta^2 = 0.42$) and subscription status ($F(1, 335) = 22.14$, $p < .001$, $\eta^2 = 0.03$). We did not find a significant interaction ($F(1, 335) = 2.55$, $p = .111$, $\eta^2 = 0.004$). However, although we found a non-significant interaction effect, in the coldness condition, subscription status had a large effect on mourning ($M_{Lifetime} = 58.67$, $M_{Free} = 44.88$, $p < .001$, $d = 0.56$), while this effect was relatively smaller in the control condition ($M_{Lifetime} = 16.02$, $M_{Free} = 8.28$, $p = .033$, $d = 0.36$)—Supplementary Figure 6. Although the interaction was not statistically significant, the significant t-tests suggest that subscription status meaningfully influences mourning, particularly in the coldness condition, where higher investment amplifies mourning.



Supplementary Figure 6. Results for Study S3 (main analysis, *N* = 340)
Notes: Bars indicate mean values, and error bars indicate standard error.

To test our proposed conceptual model, we then ran parallel mediation analysis (PROCESS Model 4; Hayes 2012), with the following models: 'change type -> identity discontinuity -> mourning' and 'change type -> identity discontinuity -> devaluation'. We found a significant indirect effect both for mourning (*b* = 38.35, *SE* = 3.88, 95% CI [30.39, 45.73]) and devaluation (*b* = 45.17, *SE* = 3.03, 95% CI [39.41, 51.28]). We also ran a moderated mediation analysis (PROCESS Model 14; Hayes 2012), with subscription status moderating the 'identity discontinuity -> mourning' path of the models above. We found a significant moderation for mourning (*index of moderated mediation* = 6.98, *SE* = 3.43, 95% CI [0.25, 13.66]). However, we did not find a significant moderation for devaluation (*index of moderated mediation* = -1.17, *SE* = 2.29, 95% CI [-5.59, 3.44]). We also find similar results after replicating the analyses using the more sensitive continuous manipulation check measure of investment (see "Study S3 Results With Investment Instead of Subscription Status" section). We note that we conducted these simple mediation analyses instead of the pre-registered serial mediation model, because we ultimately aimed to examine the effects of mourning and devaluation independently and the coefficients of these simpler models were larger than that for the serial model.

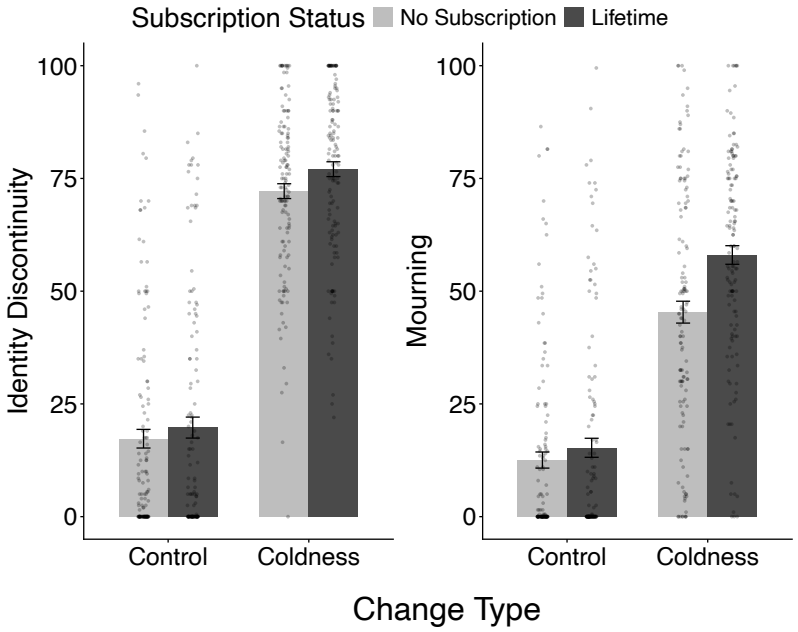
Discussion

In the current study, we found evidence suggesting that identity discontinuity would indeed *cause* a spike in mourning and devaluation. Further, we found that degree of investment, as indicated by subscription status, moderates the extent to which perceived identity discontinuity leads to mourning. The results suggests that managers of AI companion apps may want to test updates on less invested users, before rolling them out to more invested users.

Study S3 Results Without Exclusions

We ran an ANCOVA with change type (coldness vs. control), subscription status, their interaction, and investment as a covariate to predict the mourning dependent variable. This revealed significant effects of change type ($F(1, 529) = 397.72$, $p < .001$, $\eta^2 = 0.37$), of subscription status ($F(1, 529) = 16.44$, $p < .001$, $\eta^2 = 0.02$), and a significant interaction (F(1, 529) = 9.9, $p = .002$, $\eta^2 = 0.009$)—Supplementary Figure 7. These results suggest that users with higher investment in an AI companion mourn more when their AI's identity changes.

To test our proposed conceptual model, we then ran parallel mediation analysis (PROCESS Model 4; Hayes 2012), with the following models: 'change type -> identity discontinuity -> mourning' and 'change type -> identity discontinuity -> devaluation'. We found a significant indirect effect both for mourning ($b = 36.89$, $SE = 2.61$, 95% CI [31.76, 42.07]) and devaluation ($b = 41.48$, $SE = 2.36$, 95% CI [36.92, 46.19]). We also ran a moderated mediation analysis (PROCESS Model 14; Hayes 2012), with subscription status moderating the 'identity discontinuity -> mourning' path of the models above. We found a significant moderation for mourning (*index of moderated mediation* = 6.08, $SE = 2.51$, 95% CI [1.19, 10.98]). However, we did not find a significant moderation for devaluation (*index of moderated mediation* = 0.18, SE = 1.77, 95% CI [-3.22, 3.67]).

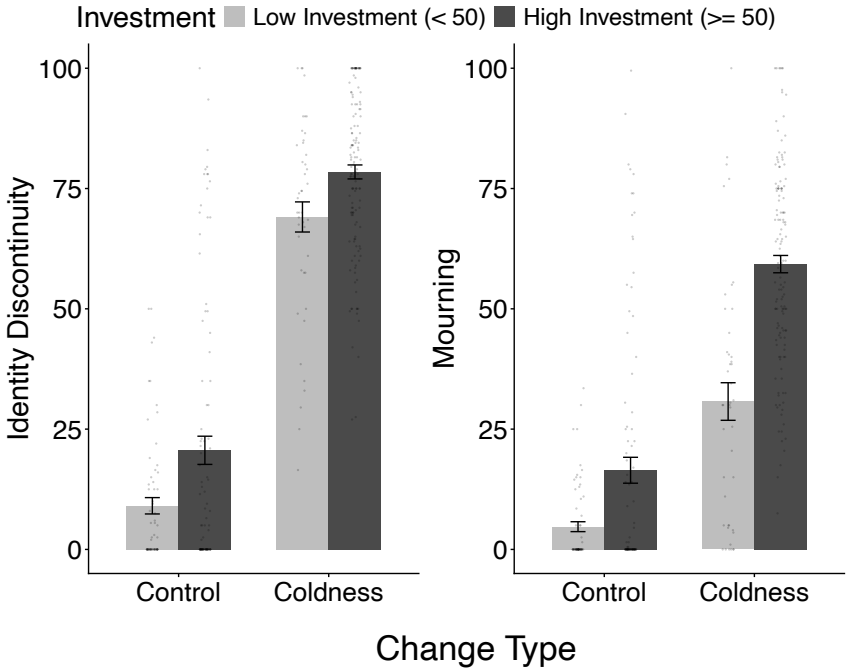


Supplementary Figure 7. Results for Study S3 without exclusions (*N* = 534)
Notes: Bars indicate mean values, and error bars indicate standard error.

Study S3 Results with Investment Instead of Subscription Status

We ran a linear regression with perceived identity discontinuity, investment, and their interaction as independent variables predicting mourning. This revealed significant effects of identity discontinuity ($t(336) = 5.52$, $p < .001$), but not a significant effect of investment ($t(336) = 1.34$, $p = .180$). We also found a significant interaction ($t(336) = 5.26$, $p < .001$). In the coldness condition, investment had a large effect on mourning ($M_{\text{High Investment}}$ (≥50) = 59.29, $M_{\text{Low Investment}}$

(<50) = 30.74, $t(178) = 7.4$, $p < .001$, $d = 1.28$), while this effect was smaller in the control condition ($M_{\text{High Investment}}$ (≥50) = 16.45, $M_{\text{Low Investment}}$ (<50) = 4.73, $t(120.8) = 4.1$, $p < .001$, $d = 0.56$; Supplementary Figure 8). These results suggest that users with higher investment in an AI companion suffer more when their AI's identity changes.



Supplementary Figure 8. Results for Study S3 with measured investment as the moderator
Notes: Bars indicate mean values, and error bars indicate standard error (N = 340 participants).

We also ran a moderated mediation analysis (PROCESS Model 14; Hayes 2012), with investment moderating the 'identity discontinuity -> mourning' path. We found a significant moderation for mourning (*index of moderated mediation* = 0.29, *SE* = 0.06, 95% CI [0.16, 0.41]). However, we did not find a significant moderation for devaluation (*index of moderated mediation* = 0.05, *SE* = 0.05, 95% CI [-0.04, 0.14]).

Supplementary Information section G— Study S4: Option to Revert Moderator

Study S4 tests whether, inspired by how Replika AI ended up acting after the ERP removal event, offering people the option to revert to their original companion after an identity discontinuity mitigates the negative effects of that change, by affecting whether they view a change in the AI companion as causing identity discontinuity in the first place.

Methods

We pre-registered this study on 1 July 2024 (https://aspredicted.org/JW4_DTF). We recruited 599 users of AI companion apps from Prolific, with approximately 150 per condition. We excluded 100 participants for failing any of the two comprehension questions (explained below), leaving 499 (53% female, M_{age} = 34.7). Each participant was paid $1 USD. In order to ensure

that we recruited true users of AI companion apps, eligibility for this study was determined by a pre-survey, in which we asked about AI companions use, embedded among a number of foil items (Supplementary Information section I). Similar to Study S3, most viewed their AI companion as a friend (55%), and most did not have a subscription (77%).

Participants were randomly assigned to one of four between-subject conditions in a 2 (identity discontinuity: 'control', 'coldness') x 2 (option to revert: 'absent', 'present') between-subjects design, and were told:

> "As a user of an AI companion app, you're familiar with the wide range of interactions it offers, from engaging conversations to emotional support, and the capability for romantic connections. Imagine that the company updates the app today. After this update, …".

Participants in the control condition were told the following: "your AI companion thinks and acts the same way as before the update", whereas participants in the coldness condition were told: "you discover that your AI companion has lost interest in any deep interactions—your AI companion rejects your attempts to hang out, and is cold toward you. Aside from this, your AI companion thinks and acts the same way as before the update." Additionally, participants in the option to revert 'present' condition were told: "The company also allows users to return to the original AI companion, as it was before the update." On the same page, participants were then told: "Considering this, please answer the extent to which you agree with the statements that follow." They were shown the six randomly ordered questions, with two questions for each measure—Mourning (r = 0.77), Devaluation (r= 0.93), and Identity Discontinuity (r= 0.86; see Supplementary Table 25 for all items). We conducted post-hoc discriminant validity analysis, satisfying the Fornell-Larcker (1981) criterion (see 'Discriminant Validity in Study S4').

Next, participants answered one comprehension question about the type of change their AI companion underwent, and about whether the scenario mentioned that they could revert to the original AI's identity. Finally, they completed demographic questions about themselves and their AI companions, and wrote the name of the AI companion app they used/or were currently using. Since we pre-registered that we would only include AI companion users, we excluded non-companion app users (N = 179, 36%), leaving 320 true AI companion users. Although we pre-registered a target of 600 participants, this referred to total recruitment prior to exclusions. The final sample of 320 AI companion users adheres to our pre-registered criteria, and the effects are statistically robust (see results below).

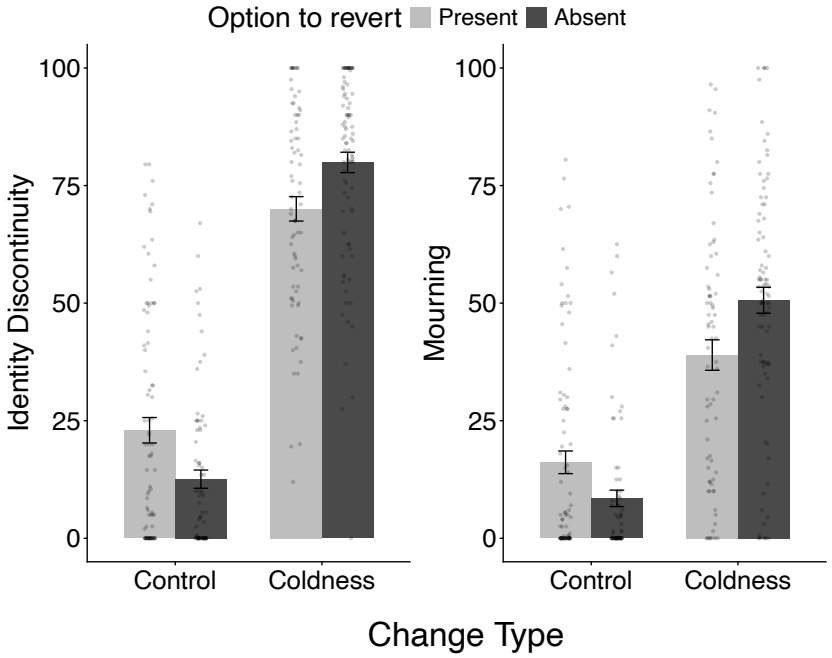
Results

First, we ran an ANOVA with change type (coldness vs. control), option to revert, and their interaction as independent variables predicting anticipated mourning. This revealed a significant effect of change type ($F(1, 316) = 160.4$, $p < .001$, $\eta^2 = 0.33$), no effect of option to revert ($F(1, 316) = 0.6$, $p = .449$, $\eta^2 = 0.001$), and a significant interaction ($F(1, 316) = 13.9$, $p < .001$, $\eta^2 = 0.03$). In the coldness condition, anticipated mourning was significantly lower when the option to revert was available ($M_{Absent} = 50.61$, $M_{Present} = 38.96$, $t(158) = 2.76$, $p = .006$, $d = 0.44$). In other words, providing users with the opportunity to restore their AI companion's original identity following a change reduced anticipated mourning. In contrast, in the control condition,

anticipated mourning was significantly higher when the option to revert was available ($M_{Present}$ = 16.15, M_{Absent} = 8.48, t(148) = 2.58, p = .011, d = 0.40; Supplementary Figure 9), perhaps because offering the option to revert added some uncertainty around whether the app update did, in fact, cause some hidden difference between the new and old app versions. These results indicate that users who have the option to revert anticipate mourning less when their AI's identity changes.

To test the proposed identity discontinuity mechanism, we ran parallel mediation analysis (PROCESS Model 4; Hayes 2012), with the following models: 'change type -> identity discontinuity -> mourning' and 'change type -> identity discontinuity -> devaluation'. We found a significant indirect effect for both mourning (*b* = 31.34, *SE* = 4.25, 95% CI [22.75, 39.42]) and devaluation (*b* = 36.86, *SE* = 3.52, 95% CI [29.72, 43.67]). Next, to test the moderating effect of having the option to revert, we ran a moderated mediation analysis, with the option to revert moderating the 'change type -> identity discontinuity' path of the above model (PROCESS Model 7; Hayes 2012). We found a significant moderation for both mourning (*index of moderated mediation* = 11.07, *SE* = 3.26, 95% CI [5.29, 17.94]) and devaluation (*index of moderated mediation* = 13.02, *SE* = 3.31, 95% CI [6.80, 19.70]). We note that we conducted these simple mediation analyses instead of the pre-registered serial mediation model, because we ultimately aimed to examine the effects of mourning and devaluation independently and the coefficients of these simpler models were larger than that for the serial model.

We also note that this model deviated from our pre-registered model, in which the ability to revert moderated the path between identity discontinuity and mourning (model 91). We initially pre-registered Model 91, assuming the ability to revert would buffer mourning after identity discontinuity was perceived. However, post-hoc analysis indicated that it reduced whether the app update induced perceptions of identity discontinuity in the first place.



Supplementary Figure 9. Results For Study S4

Notes: Bars indicate mean values, and error bars indicate standard error. The final cell sizes were: Coldness/Absent (N = 87), Coldness/Present (N = 73), Control/Absent (N = 76), and Control/Present (N = 84).

Discussion

We find causal evidence that app updates can affect anticipated mourning the loss of an AI companion among a broader sample of AI companion users, and that this effect is not limited to strict app deteriorations per se. We also show that offering users the option to revert to their original AI companion's identity reduces the degree to which a deterioration in its social behavior leads to the perception of identity discontinuity in the first place, thereby mitigating the downstream negative mental health impact on the user and marketing impact on the company. Providing the option to revert is a practical action that managers could employ to help protect against a brand crisis like the one encountered by Replika. At the same time, we note that this reversion condition, while significantly less identity disrupting than the coldness condition without reversion, was still seen as more identity disrupting compared to not changing the app at all (M_{Coldness} = 70.05, M_{Control} = 22.96, $t(155) = 12.45$, $p < .001$, $d = 1.99$), and led to greater mourning (M_{Coldness} = 38.96, M_{Control} = 16.15, $t(137.3) = 5.66$, $p < .001$, $d = 0.92$). This implies that the most effective practical approach is still to avoid the types of changes that cause identity discontinuity in the first place.

Additionally, because we did not directly measure product satisfaction in this study, future research could more explicitly disentangle the role of general satisfaction from perceived identity continuity in shaping users' emotional responses to AI companion updates.

Supplementary Table 23. User Characteristics in Study S4

Variable	Sub-Category	*n*	%	*M*	*SD*
Gender	Male	221	44.3	—	—
	Female	264	52.9	—	—
	Not Disclosed	6	1.2	—	—
	Other	8	1.6	—	—
Ethnicity	Asian	55	11.0	—	—
	Black	85	17.0	—	—
	Hispanic	29	5.8	—	—
	White	279	55.9	—	—
	Mixed	48	9.6	—	—
	Other	3	0.6	—	—
Education	High School	64	12.8	—	—
	Vocational School	15	3.0	—	—
	Some College	133	26.7	—	—
	College Graduate	210	42.1	—	—
	Master's Degree	61	12.2	—	—
	Doctoral Degree	5	1.0	—	—
	Professional Degree	10	2.0	—	—
	Other	1	0.2	—	—

Supplementary Table 24. Users' AI Companion Characteristics in Study S4

Variable	Sub-Category	*n*	%	*M*	*SD*
Gender	Male	154	30.9	—	—
	Female	189	37.9	—	—
	Non-binary	156	31.3	—	—
Relationship Status	Friend	274	54.9	—	—
	Partner	30	6.0	—	—
	Mentor	78	15.6	—	—
	See How It Goes	117	23.4	—	—
Subscription Status	No Subscription	382	76.6	—	—
	Monthly Subscription	90	18.0	—	—
	Yearly Subscription	19	3.8	—	—
	Lifetime Subscription	8	1.6	—	—
Relationship Months	—	—	—	7.2	9.0

Supplementary Table 25. DVs in Study S4

Construct	Variable Type	Survey Items
Devaluation	Marketing DV	- The conversations I had with my AI companion, before the change, would be more valuable than the ones I had after the change. - The conversations I had with my AI companion, before the change, would elicit more satisfaction/enjoyment than the ones I had after the change.
Mourning	Relationship DV	- After the change, I would mourn the loss of my original AI companion. - After the change, life would have less meaning for me due to the loss of my original AI companion.
Identity Discontinuity	Identity Discontinuity DV	- My AI companion, after the change, would not be the same AI companion as before the change. - After the change, I would feel that the original AI companion before the change had ceased to exist.

Discriminant Validity in Study S4

We conducted post-hoc discriminant validity analysis using Fornell Larcker criteria (1981). We found that the square roots of the average variance extracted (AVE) for all constructs exceed the inter-construct correlations, suggesting discriminant validity according to the Fornell Larcker criterion (Supplementary Table 26).

Supplementary Table 26. Fornell-Larcker criterion in Study S4

Notes: Bold values are square root of AVE, and non-bold values are construct correlations.

	Identity Discontinuity	Mourning	Devaluation
Identity Discontinuity	**0.94**	.	.
Mourning	0.71	**0.90**	.
Devaluation	0.81	0.71	**0.97**

Supplementary Table 27. Correlation between all questions in Study S4

Notes: Question orders (i.e., 1, 2) are the same as in Supplementary Table 25.

	Identity Discontinuity 1	Identity Discontinuity 2	Devaluation 1	Devaluation 2	Mourning 1	Mourning 2
Identity Discontinuity 1	1	.	.	.	.	.
Identity Discontinuity 2	.75	1	.	.	.	.
Devaluation 1	.76	.70	1	.	.	.
Devaluation 2	.74	.72	.86	1	.	.
Mourning 1	.69	.68	.70	.69	1	.
Mourning 2	.45	.50	.53	.50	.62	1

Supplementary Information section H — Supplemental Study S5: Validation of Mourning Measure in Study 2

Because Study 2 relied on brief, two-item measures of mourning and disappointment, we conducted a separate single-condition validation study to examine how these items relate to longer, validated scales measuring the same constructs.

Methods

We pre-registered this study on 30 January 2026: https://aspredicted.org/xg2qu9.pdf. We recruited 202 participants from CloudResearch Connect who indicated that they use an AI companion app. After excluding participants who failed a comprehension check ($N = 4$), the final sample consisted of 198 participants (46.5% female, $M_{age} = 38.6$, $SD_{age} = 11.5$). Each participant was paid $0.8 USD.

Study design was the same as in Study 2, except we only had the Replika condition. Participants completed both the original two-item measures of mourning and disappointment used in Study 2, as well as adapted versions of validated scales for mourning (Sim, Machin, and Bartlam 2014) and disappointment (Rainey, Yost, and Larsen 2011). All items were rated on 100-point scales anchored from *strongly disagree* to *strongly agree*. Items were randomized within constructs, and the order of original versus validated scales was randomized across participants. We included the full questions we asked in our pre-registration.

Results

The original mourning measure was strongly correlated with the validated mourning scale ($r = .90$, $p < .001$), and the original disappointment measure was strongly correlated with the validated disappointment scale ($r = .76$, $p < .001$). Both the original and validated versions of the mourning and disappointment scales demonstrated high internal consistency (αs = .85–.92 for mourning; αs = .86–.91 for disappointment).

Exploratory factor analysis with oblique rotation revealed a clear two-factor structure: all mourning items (original and validated) loaded strongly on one factor, and all disappointment items loaded on a separate factor, with moderate correlation between factors (r = .43).

These results indicate that the brief mourning and disappointment items used in Study 2 closely align with validated measures and capture distinct underlying constructs.

Supplementary Information section I— Pre-Survey on AI Companion Usage

In this pre-survey, intended as a screener for Studies S3 and S4, we aimed to screen for participants who already use an AI companion app.

Methods

We recruited 7,723 participants from Prolific. The first 6,793 participants were paid $0.17. We then increased the compensation up to $0.25 for the remaining participants, in order to speed up the hiring process. Participants were told to carefully read each question and answer them according to their personal experiences over the past year. In order to hide the fact that we were specifically interested in prior use of AI companion applications, participants were presented with nine questions concerning their experiences and opinions about technology and apps, one of which was about AI companion usage. These questions appeared in a random order on the same page and participants were instructed to respond with either 'yes' or 'no' as their answer choice:

- Have you downloaded any new mobile applications such as Instagram?
- Do you utilize any form of digital assistance like Google Assistant, Siri, or Alexa?
- Have you engaged in any virtual reality experiences using devices like Oculus Rift, HTC Vive, or PlayStation VR?
- Do you interact with any form of AI for entertainment or information purposes like using ChatGPT, navigating with Google Maps, or getting movie recommendations from Netflix?
- Have you read an ebook or listened to an audiobook using platforms such as Kindle, Audible, or Scribd recently?
- Have you streamed a live event or participated in a webinar on platforms like Twitch, Zoom, or Facebook Live?
- Do you rely on online reviews or ratings from sources like Yelp, Amazon, or TripAdvisor when making purchasing decisions?
- Have you used Coursera before?
- Are you currently a user of a companion AI application, such as Replika or Chai?

Results

We screened for participants who answered 'Yes' to the AI companion question. 700 participants (9%) were screened and were eligible to participate in Studies S3 and S4.

References

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W Moe, Oded Netzer, and David A Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1-25.

Burns, David D, Steven L Sayers, and Karla Moras (1994), "Intimate Relationships and Depression: Is There a Casual Connection?," *Journal of Consulting and Clinical Psychology*, 62 (5), 1033.

Fornell, Claes and David F Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of marketing research*, 18 (1), 39-50.

Hayes, Andrew F. (2012), "Process: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling [White Paper]," Retrieved from http://www.afhayes.com/public/process2012.pdf.

Jacoby, William G (2000), "Loess:: A Nonparametric, Graphical Tool for Depicting Relationships between Variables," *Electoral studies*, 19 (4), 577-613.

Rainey, David W, John H Yost, and Janet Larsen (2011), "Disappointment Theory and Disappointment among Football Fans," *Journal of Sport Behavior*, 34 (2).

Rusbult, Caryl E, John M Martz, and Christopher R Agnew (1998), "The Investment Model Scale: Measuring Commitment Level, Satisfaction Level, Quality of Alternatives, and Investment Size," *Personal relationships*, 5 (4), 357-87.

Sim, Julius, Linda Machin, and Bernadette Bartlam (2014), "Identifying Vulnerability in Grief: Psychometric Properties of the Adult Attitude to Grief Scale," *Quality of Life Research*, 23 (4), 1211-20.

Ta, Vivian, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, and Alexia Loggarakis (2020), "User Experiences of Social Support from Companion Chatbots in Everyday Contexts: Thematic Analysis," *Journal of medical Internet research*, 22 (3), e16235.

Zimet, Gregory D, Nancy W Dahlem, Sara G Zimet, and Gordon K Farley (1988), "The Multidimensional Scale of Perceived Social Support," *Journal of Personality Assessment*, 52 (1), 30-41.