

Do Personal AI Conversations Reduce Loneliness?

Evidence from a Pre-Registered Experiment

Louis Fréget¹, Ariell Reshef², and Claudia Senik³

¹Stone Centre, University College London ; Université Paris-Dauphine (PSL)

²CNRS, Paris 1 Pantheon-Sorbonne, Paris School of Economics, CESifo

³Sorbonne University; Paris School of Economics; IUF

Version: October 1, 2026

Abstract. Generative AI is increasingly promoted as a remedy for loneliness. We test this claim in a large-scale pre-registered experiment on a large sample of French adults. Participants assigned to treatment were asked to hold at least one personal conversation a day with a generative AI tool of their choice for 28 days; the control group continued as usual. The intervention raised socio-emotional AI use substantially. In response, subjective loneliness increased. Other measures of subjective well-being were also consistently adversely affected, although these effects are imprecisely estimated.

JEL: C93, I31, I12, O33. **Keywords:** Personal AI conversations ; Generative AI; Subjective well-being; Loneliness; Social connectedness.

1 Introduction

Modern societies are experiencing a loneliness epidemic. For example, the U.S. Surgeon General declared social isolation a public health emergency in 2023, noting that half of American adults experience measurable loneliness (Murthy, 2023). In the United States, the share of adults reporting no close friends rose from 3% in 1990 to 12% in 2021 (Cox, 2021) and up to 17% in 2024 (Cox & Pressler, 2024). More broadly, 19% of young adults across the world reported in 2023 that they have no one they could count on for social support, a 39% increase compared to 2006 (Pei and Zaki, 2025). The World Health Organization has described loneliness as a pressing health threat (WHO, 2023).¹

Alongside this crisis, generative AI has become pervasive. By mid-2025, more than half of U.S. adults aged 18–64 had used a generative AI tool (Bick et al., 2025). A recent survey found that the top use of AI is for therapy and companionship (Zao-Sanders, 2025). The fast growing use of these systems for emotional support, companionship, and relationship advice places them squarely in roles once occupied by human interlocutors. While technology leaders have proposed AI as a direct remedy for loneliness, there is little evidence to support this claim.²

In this paper, we ask whether encouraging individuals to converse with AI about their personal lives reduces loneliness or reinforces it. The question is of first-order importance. Generative AI has created a new form of socio-emotional interaction, and this use is diffusing rapidly. However, its consequences for well-being and social connection are largely unknown. If ordinary AI conversation improves the well-being of isolated adults at negligible cost, it could serve as a low-cost bridge to human support and help alleviate the loneliness epidemic. Our results, however, go in the opposite direction: we find that conversations with AI cause a significant increase in loneliness, consistent with the displacement of support, as well as other related negative outcomes.

The effect of conversations with AI on subjective well-being is theoretically ambiguous. These conversations may improve well-being: they offer companionship, allow users to articulate their feelings, and may prompt them toward human contact (complementarity). They may also deteriorate subjective well-being: repeated attention to personal difficulties can encourage rumination, which has been shown to amplify anxiety and depression (Nolen-Hoeksema, 1991); an always-available interlocutor may substitute for the effort of reaching

¹For example, social isolation is associated with mortality risks comparable to smoking fifteen cigarettes a day (Holt-Lunstad et al., 2010).

²For example, Mark Zuckerberg (CEO of Meta Platforms, a leading technology company) argues that most people want more social connections than they have and that AI might partially fill this gap, albeit not as well as human relationships. He also suggests that AI can help people prepare for difficult interpersonal conversations, for instance, with a partner or a manager (Zuckerberg, 2025).

out to others (substitution); and an accommodating model may validate grievances rather than help resolve them. Which mechanism dominates is an empirical question.

More generally, AI changes the relative price of social interaction. By providing an always-available and low-friction alternative to human conversation, it may reduce the incentives to invest in costly social ties, such as initiating contact, maintaining friendships, or repairing strained relationships. Because these investments shape the future availability and strength of one’s social network, short-run substitution away from human contact may generate longer-run effects on social connectedness.

We shed light on these questions by conducting a pre-registered randomized controlled trial on a large sample of French adults. Participants assigned to treatment were asked to hold at least one personal conversation per day with a generative AI tool of their choice for 28 consecutive days on topics related to their personal lives: their day, their feelings, their goals, their concerns, or any other aspect of their life they would like to share. Control participants were asked to continue their usual behavior. We measure subjective well-being (life satisfaction, loneliness, happiness, and depression), as well as relationship satisfaction with partners, family, and friends, and objective social interactions (meals taken alone and in-person meetings) at baseline and four weeks later. In addition, we record participants’ willingness-to-accept (WTA) to forgo AI for one week and how pleasant they found the conversations, as measures of participants’ own assessments. We estimate the intention-to-treat effect of encouraging personal AI conversations.

Our findings can be summarized as follows. Treated participants complied with the request, increasing the number of days on which they held personal conversations with AI by about 5 days. These conversations did not, however, improve their mental health. On the contrary, subjective loneliness increased. Other subjective well-being outcomes were also adversely affected, i.e. lower life satisfaction, higher depression, and more objective isolation, although these effects are imprecisely estimated. Taken together, the results are consistent with substitution, rumination, and AI validating grievances, rather than with support and complementarity. Prespecified heterogeneity analyzes yield larger point estimates for respondents who have a partner. *Ex post* exploratory heterogeneity analyzes also suggest that adverse well-being effects are more concentrated among respondents living in smaller urban units, where opportunities for offline social interaction may be more limited. Overall, our results suggest that the development of ordinary, unstructured personal AI conversations is more likely to worsen subjective well-being than to improve it.

This study is, to our knowledge, the first to test the causal effect of open-ended, naturalistic AI conversation on loneliness and subjective well-being, as opposed to scripted tasks or dedicated mental-health applications. The existing causal evidence focuses on guarded

settings: dedicated mental-health applications with clinical safeguards (Angelucci et al., 2026), companion-specific products or purpose-designed empathetic chatbots (De Freitas et al., 2026), experiments with assigned interaction modes, scripted prompts and tasks (Fang et al., 2025), and structured chatbots explicitly designed around psychological well-being strategies (Schöne et al., 2025; Li et al., 2026).³ In contrast, participants in our setting use ordinary generative AI, chosen by themselves, for unscripted personal conversation. We never observe the content of these conversations, which removes any concern that awareness of being observed shapes the nature of the interaction.

Our paper relates to the broader health economics literature on how digital technologies shape well-being and social behavior (Allcott et al., 2020; Braghieri et al., 2022; McDool et al., 2020; Donati et al., 2025; Myers & Hooper, 2026). It is also related to a large economic literature on loneliness, which has been described as an outcome of economic and social environments (Krekel et al., 2021; Kong & Lam, 2024; Babin et al., 2021; Morrish & Medina-Lara, 2021; Rebecchi et al., 2024), but also as a source of lower subjective well-being (De Neve et al., 2026) and as a factor associated with political behavior (Peterson et al., 2025; Berlingieri et al., 2025).

2 Data and Design

Setting and sample. We recruited adults residing in France from an online panel operated by the survey company Bilendi starting in January 2026.⁴ Wave 1 took place from January 5 to 19, 2026. Participants completed a baseline questionnaire covering demographics and all outcomes and were randomized to treatment or control in equal proportions. The treatment period ran from January 22 to February 18, 2026. In Wave 2 (February 19–28, 2026), the main cohort participants completed the endline questionnaire covering the same outcome

³One experiment in De Freitas et al. (2026) comes closer to our setting by using a neutral generalist chatbot rather than a purpose-built AI companion (Study 5). Important differences remain, however: their control group interacts with a chatbot restricted to arithmetic and grammar correction, and loneliness is measured within the same fifteen-minute session. Our setting instead studies self-chosen, open-ended use of ordinary generative AI and measures effects over a longer horizon. We discuss these differences in more detail in Section 3. Fang et al. (2025) also study a general-purpose chatbot, but their design imposes a minimum daily-use requirement and assigns daily conversation topics in the personal and non-personal arms, whereas our participants freely choose both whether and what to discuss; they also have no no-AI control group.

⁴Bilendi was unable to meet its recruitment targets for 18–24-year-olds. The survey company hypothesized that younger adults might be less willing to commit to a month-long daily task, potentially due to less flexible schedules and higher opportunity costs of time. To address this underrepresentation, Bilendi enlisted an additional recruitment partner on January 20, 2026, in an attempt to recruit more young participants. These participants followed the same experimental protocol with a three-week delay, with the last endline questionnaire completed on March 20, 2026. However, this effort did not achieve the target age distribution, and only 106 additional participants were included in the final sample. We do not have access to the intermediate recruitment data.

measures.⁵ Randomization was at the individual level. Of those who passed the baseline attention check, 7,286 respondents (3,642 control and 3,644 treatment) completed both waves and form the analysis sample. Endline completion was virtually the same across treatment and control arms (74.8% vs. 74.7%; $p = 0.92$). Table 1 displays the covariate balance: all the Imbens–Rubin normalized differences for the primary analysis sample are below 0.04 in absolute value, well under the conventional threshold of 0.25. Table 2 compares the sample to the French population: our sample is older, more educated, and more likely to be partnered, and it reports somewhat lower subjective well-being at baseline.⁶

Intervention. Participants assigned to treatment were asked to hold at least one brief personal conversation each day for 28 consecutive days on personal topics such as their day, feelings, goals, or concerns. The instructions highlighted that they could use the generative AI tool of their choice (for example, ChatGPT, Anthropic, or Mistral). The treatment instructions were also briefly reiterated via emails at the midpoint of the intervention period. Control participants were only asked to maintain their usual behavior. To reduce basic usability frictions for participants with little prior experience, those unfamiliar with generative AI could watch a short soundless tutorial video (under one minute) showing a simple example of how to start a conversation with an AI assistant, beginning from a google search page.⁷ Yet take-up of the tutorial was low: only 12.3% of treatment participants actively clicked play at least once, and at most 34% watched more than five seconds (including cases of autoplay triggered automatically in certain browsers). We did not access the content of any conversation to minimize changes in behavior that could arise from knowing that their conversations were being observed. Compliance was self-reported at endline: participants indicated the total number of days on which they had held a personal AI conversation over the 28-day period. To help participants maintain an accurate count, both groups were provided with a daily online log in which they recorded whether they had spoken with an AI about their personal life each day; the log automatically tallied their running total. Weekly SMS reminders prompted both groups to update the log.

⁵During the experimental period, several capable generative AI tools were freely available, including ChatGPT (GPT-4o, OpenAI), Claude 3.5 Sonnet (Anthropic), Gemini 1.5 Flash (Google), and Le Chat (Mistral). Premium subscriptions, such as ChatGPT Plus, Claude Pro, and Gemini Advanced offered higher usage limits for around \$20 per month.

⁶The over-representation of older adults is consistent with feedback from the survey company, which noted that younger respondents were less likely to commit to a month-long daily task.

⁷The tutorial featured Mistral’s Le Chat as an illustrative example, as it offered a generous free tier well-suited to daily use, with no subscription requirement. To verify that exposure to Mistral branding did not drive the results, we replicated the main estimates, excluding the 1,319 participants who watched more than one second of the video (before the Mistral AI google search). The results on primary and secondary outcomes are stable in direction and magnitude, arguing against the tutorial being a mechanism in driving the results.

Outcomes. We report three first-stage outcomes, all of which are based on self-reported frequency of AI use by subjects during the 28-days of the experiment. They were reported by participants at endline. The first is the number of days they reported having held at least one personal AI conversation over the 28-day window. The second is a pre-specified broad index based on eight more specific questions about the frequency at which they used AI for personal conversations, planning, emotional support, companionship, relationship advice, information search, health information, and work tasks, each measured on a 1–6 frequency scale. Participants who did not use AI are assigned the minimum value (1 = never) on all items. The third index focuses on the questions related to socio-emotional AI use, i.e. emotional support, companionship, and relationship advice (with the same convention for non-users). This index was specified in the pre-analysis plan as the primary measure of socio-emotional AI use in the baseline descriptive analysis but was not pre-registered as a first-stage outcome; it is reported here as an exploratory measure.

Our primary outcomes measure subjective well-being: life satisfaction, happiness, and depression, each on a 0–10 scale, and loneliness, measured both with a single 0–10 item and with the three-item UCLA scale (Hughes et al., 2004). These integer scales are standard in the economics of subjective well-being (Kahneman & Deaton, 2010) and predict subsequent behavior in an approximately linear fashion (Kaiser & Oswald, 2022). Direct single-item measures of loneliness are also widely used in population surveys and official well-being measurement frameworks (OECD, 2013). They correlate with established multi-item loneliness scales and show the expected relationships with other measures of subjective well-being (Mund et al., 2023; Reinwarth et al., 2023).

Secondary outcomes measure social connectedness: satisfaction with partners, family, and friends, and objective loneliness, specifically the number of meals taken alone and the number of in-person meetings with friends or relatives in the preceding seven days, as well as perceived social support. The latter outcome is measured with a binary indicator equal to one if respondents report having at least one person on whom they can rely in case of need. Closely related measures of “someone to count on” are widely used as indicators of perceived social support in the Gallup World Poll and in the OECD *How’s Life?* framework, where social support is defined by whether respondents have relatives or friends they can count on in times of trouble. We also record willingness-to-accept (WTA) to forgo AI for one week and how pleasant participants found the conversations (0–10).

Empirical strategy. The primary estimand is the intention-to-treat (ITT) effect of assignment on endline outcomes. Our main specification is an analysis of covariance (ANCOVA):

$$Y_{i2} = \alpha + \beta T_i + \gamma Y_{i1} + \mathbf{X}_{i1}'\boldsymbol{\theta} + \varepsilon_{i2}, \quad (1)$$

where Y_{i2} is the outcome at endline, T_i is treatment assignment, Y_{i1} is the baseline value of the same outcome, and $\mathbf{X}_{i1}$ is a vector of pre-specified covariates (age, gender, education, socio-economic status, telework frequency, relationship status, and household composition).⁸ We report robust standard errors, and each outcome is estimated on its own complete cases, so the estimation sample varies across outcomes. We pre-specified three outcome families—subjective well-being, relationship quality, and objective interactions—and report Romano-Wolf stepdown p -values within each.⁹ As pre-specified robustness checks, we report a difference-in-means specification without baseline adjustment (Appendix Table A3), and Lee (2009) bounds to assess sensitivity to differential attrition (Appendix Table A4). Our results are largely robust to these checks.¹⁰ Finally, to assess heterogeneity, we augment specification (1) with a treatment-moderator interaction term and plot the estimated marginal effect of treatment within each subgroup.

3 Results

Baseline associations. We first document the cross-sectional negative relationship between socio-emotional AI use and well-being at baseline (Appendix Table A1). These associations are descriptive and subject to selection, as the decision to use AI for companionship is not random. In particular, reverse causality might be at play. Our experiment is designed to eliminate this selection through randomization.

First stage. The intervention shifted behavior as intended, as reported in Table 3. Assignment to treatment increased the self-reported number of days with personal AI conversations by about 5.2 days over the 28-day window (compared to a control median of 2 days).¹¹ As

⁸ANCOVA dominates difference-in-differences in efficiency when the baseline-endline autocorrelation is below one (McKenzie, 2012).

⁹We present the pre-specified experimental analyzes in Section 4.2 of the pre-analysis plan.

¹⁰The difference in-means specification yields directionally identical results across nearly all outcomes (Appendix Table A3), though point estimates and standard errors differ somewhat without baseline adjustment. Lee (2009) bounds confirm the robustness of the primary results to differential attrition, which is very low in our setting (−0.09 percentage points). Both bounds are positive and statistically significant for the single-item loneliness measure; the upper bound is marginally significant for UCLA-3 and depressive affect, while the lower bounds remain positive but imprecisely estimated.

¹¹Respondents who could not recall their exact number of personal AI conversation days were allowed to skip this question, resulting in 518 missing values (7.1% of the analysis sample). Under mean imputation

pre-specified, we also report the effect of assignment to treatment on a broad index of AI use frequency across all domains; this index increased by 0.43 standard deviation over the time of the experiment ($F = 343$). In addition, we report the effect of being in the treatment group on a subscale of the broad index measuring the usage of AI for socio-emotional purposes, such as emotional support, companionship, and relationship advice; this index increased by 0.46 standard deviations ($F = 324$).

Subjective well-being. The intention-to-treat estimates are reported in Table 4. Assignment to daily personal AI conversation robustly increases subjective loneliness on the single-item measure. The effect is statistically significant at conventional levels of significance, even after applying the Romano-Wolf correction within this family of outcomes ($p = 0.001$). The remaining outcomes all move in the same adverse direction: loneliness on the UCLA-3 scale increases, life satisfaction declines, happiness declines, and depression rises. However, these effects are imprecisely estimated and are not statistically significant at conventional significance levels. The point estimates are modest in size but indicate lower subjective well-being across several dimensions.

Our results contrast with the estimates of De Freitas et al. (2026), who find a *reduction* in loneliness after a single fifteen-minute session with a neutral generalist chatbot (in their study 5). Their control condition, however, is not normal daily life but an interaction with a chatbot capable only of arithmetic and grammar correction. Moreover, their outcome is measured within the same fifteen-minute session, capturing a momentary mood shift rather than a lasting change in subjective well-being.¹² Similarly, Angelucci et al. (2026) find gains of 0.3 SD in mental health from a purpose-built AI therapy application administered to women in psychological distress. The difference with our results is due to differences in both treatment and control. Our treatment consists of free-conversation, while the control group is neutral; in contrast, their treatments are structured, and their control groups have specific settings.

We benchmark our estimates against those of another randomized health economics experiment involving digital technology. Allcott et al. (2020) find that a four-week Facebook deactivation intervention improved subjective well-being by approximately 0.08 SD and re-

with a missing value indicator, as pre-specified in the pre-analysis plan, the first-stage coefficient remains large and highly significant (4.77, SE = 0.24, $p < 0.001$, $N = 7,286$), though mechanically attenuated toward zero relative to the complete-cases estimate (5.16).

¹²De Freitas et al. (2026) also find reductions in loneliness using populations and AI tools that differ greatly from those in our setting. They measure loneliness immediately after an interaction in a single session with an AI companion (their studies 3 and 6). They also extend the intervention to seven daily sessions of fifteen minutes each, but with a purpose-built empathetic AI companion rather than an ordinary general-purpose AI (Study 4).

duced loneliness by approximately 0.03 SD, although the latter effect is not statistically significant in their study.¹³ Our nudge, which increased the number of days with personal AI conversations by only about five over a 28-day period, produces an adverse effect on loneliness of roughly 0.06 SD, twice the magnitude of their estimated effect on loneliness and three-quarters of their effect on overall subjective well-being.

Relationships and objective contact. The same broad adverse pattern appears in measures of real-world connection with others, as reported in Table 5. On average, treated participants report more lunches and dinners taken alone, fewer in-person meetings with friends or relatives, and lower satisfaction with partner and family relationships. Friendship satisfaction is essentially unchanged. Similarly, we find no detectable effect on perceived social support, measured by whether respondents have someone they can rely on in case of need. These estimates are imprecisely estimated and are not statistically significant at conventional significance levels. Overall, the estimates provide no evidence of improved social connectedness, and if anything, they point toward greater isolation.

Participants’ assessment of AI. Participants tend to perceive AI as siding with them in interpersonal conflicts in both the treatment and control groups (6.36 on a 0 – 10 scale). This suggests that AI validation may play a role even if the treatment does not significantly increase this perception: by encouraging more frequent personal conversations, it may increase participants’ exposure to such validation.

The adverse effect on loneliness does not appear to reflect a negative experience with the tool. The results reported in Table 6 indicate that treated participants rate the conversations as slightly more pleasant than those in the control group (0.10 SD; $p < 0.001$) although their subjective well-being declined. These outcomes are only asked of participants who used AI during the experiment, as pre-specified in the registered questionnaire. However, baseline covariates remain well balanced within each estimation sample, with all standardized differences below 0.10 (see Appendix Table A2).¹⁴

Pre-specified heterogeneity analysis. Figure 1 reports marginal treatment effects on single-item loneliness (the outcome most robustly affected by the intervention) across all

¹³The reduced-form effects (ITT) in Allcott et al. (2020) are approximated by multiplying the reported IV estimates by the first stage: $ITT = LATE \times (0.90 - 0.02)$, yielding approximately 0.08 SD for the composite index of subjective well-being and 0.03 SD for loneliness. The reported IV estimate for loneliness is not statistically significant ($p = 0.54$).

¹⁴These outcomes are estimated without baseline adjustment, as the baseline is only defined for participants who had already used AI for the relevant type of conversation prior to the experiment. Results are qualitatively unchanged when controlling for the baseline value for individuals who answered this question in the baseline.

pre-specified moderator subgroups: baseline relationship status, gender, baseline loneliness, telework at baseline, and age.

The most notable heterogeneity concerns relationship status: the effect is concentrated among participants in a relationship at baseline, whereas the estimate for single participants is small and statistically indistinguishable from zero. The same pattern appears when using the alternative measure of loneliness (UCLA-3; see Figure A1). The difference in treatment effects between partnered and single respondents (as measured by the interaction term) is statistically significant for the UCLA-3 loneliness scale ($\hat{\beta} = -0.075$, $p = 0.032$), but falls short of conventional thresholds for the single-item measure ($\hat{\beta} = -0.054$, $p = 0.176$).¹⁵ One possible mechanism is partial substitution away from communication with the partner. The negative, though imprecise, estimate on relationship quality is consistent with this interpretation. No clear heterogeneity can be detected along the other pre-specified dimensions.

Exploratory heterogeneity analysis. Beyond the pre-specified moderators, we conduct two additional exploratory heterogeneity analyzes motivated by potential differences in initial exposure to AI and opportunities for offline social interaction.

First, we test whether the effect of the treatment reflects novelty among previously inexperienced users, or could depend on the type of AI subjects used before the experiment. Figure 2 examines this potential heterogeneity by distinguishing individuals with pre-treatment main usage of ChatGPT, Claude, Mistral, other AI tools, or no AI usage. The marginal effect of the treatment on loneliness is qualitatively similar for respondents who had not used AI at all before (39% of the sample), those who had mainly used ChatGPT (47% of the sample), and those who had mainly used Mistral (2.7% of the sample). Subjects who had used other AI tools (typically Google Gemini or Microsoft Copilot) prior to the experiment represent a minority (11% altogether). A joint test rejects the hypothesis that the effect of the treatment is driven by prior familiarity with a particular AI platform, which could be due to small sample sizes.

Second, we examine heterogeneity by the size of the urban unit of residence of respondents, which we view as a proxy for the local opportunity set for offline social interactions. Using the *taille d'unité urbaine* (TUU) classification of the National Institute of Statistics and Economic Studies (INSEE), we distinguish respondents living in rural areas or in urban units with fewer than 100,000 inhabitants (TUU categories 0–5) from those living in urban units with at least 100,000 inhabitants, including the Paris urban unit (TUU categories 6–8). The 100,000-inhabitant threshold corresponds to an existing boundary in the INSEE

¹⁵These heterogeneity tests are mechanically underpowered relative to the main ITT estimates due to smaller sample sizes.

classification and divides the sample into two similarly sized groups (52.4% and 47.6%).

As shown in Figure 3, the adverse effects of the treatment on subjective well-being appear concentrated in less densely urbanized areas. Among respondents living in rural areas or in urban units with fewer than 100,000 inhabitants, treatment lowers life satisfaction and increases loneliness on both the UCLA-3 scale and the single-item measure. By contrast, the corresponding estimates for respondents in urban units of 100,000 inhabitants or more are small and statistically indistinguishable from zero. Finally, in both groups, the effects on happiness and depression scores are small and not statistically significant. One possible interpretation is that daily AI conversation is more harmful in environments where in-person social contact requires greater effort. Note, however, that this pattern is only suggestive, as the difference in treatment effects between the two groups is not statistically significant in formal interaction tests.

4 Discussion

Our experiment provides an exogenous manipulation of socio-emotional AI use. The intervention increased the frequency of personal conversations with AI and had a robust adverse effect on subjective loneliness. Most remaining outcomes move in the same adverse direction, i.e., lower life satisfaction, higher depression, and greater objective isolation, although these effects are imprecisely estimated. The results are consistent with substitution, rumination, and AI validating grievances rather than with complementarity and support.

A limitation of our study is the duration of the experiment, which lasted only one month, whereas the effects of longer exposure may differ. Moreover, because we encourage rather than mandate use, our estimates measure the ITT effects of a nudge toward AI conversation, not the effects of intensive use. In addition, our design does not distinguish the effect of AI conversation *per se* from the act of daily personal reflection that the treatment also induces. An alternative self-reflection control arm, such as diary-writing or discussion groups, could help disentangle the two, and we encourage future work to include one. At the same time, the volume of personal reflection itself may be part of the mechanism: an always-available, responsive interlocutor may induce people to dwell on their problems more than they otherwise would.

Overall, the results temper an optimistic view of personal AI conversations. The proposition that ordinary AI conversation can serve as a low-cost remedy for loneliness is rejected by a naturalistic test: encouraging daily AI conversation left participants somewhat lonelier. Our results suggest that the development of ordinary, unstructured personal AI conversations is more likely to worsen subjective well-being than to improve it.

Registration and Ethics

This study was pre-registered on the AEA RCT Registry (AEARCTR-0016749). It was approved by the IRB-PSE (protocol 2025-044).

Table 1 Baseline balance between treatment and control

	Control	Treatment	Norm. diff
Age (years)	51.87	51.71	-0.011
Female (%)	55.1	54.6	-0.011
Higher education, long (%)	25.4	24.3	-0.024
In a relationship (%)	70.5	70.8	+0.005
Household size	2.43	2.41	-0.017
Child <15 at home (%)	26.0	26.8	+0.018
Economically active (%)	62.3	63.5	+0.025
Life satisfaction (0–10)	6.42	6.42	+0.002
Loneliness, UCLA-3 (3–9)	4.80	4.83	+0.017
Loneliness, single item (0–10)	3.52	3.58	+0.019
Happiness (0–10)	6.41	6.45	+0.018
Depression (0–10)	3.04	3.09	+0.017
Observations	3,642	3,644	7,286

Notes. Means by arm at baseline on the primary analysis sample ($N = 7,286$). The last column reports the Imbens–Rubin (2015) normalized difference, calculated as treatment - control: $\Delta = (\bar{X}_T - \bar{X}_C)/\sqrt{(s_T^2 + s_C^2)/2}$.

Table 2 Sample characteristics vs. French population

	Sample	France	Difference
<i>Sociodemographics</i>			
Women (%)	54.8	52.3	+2.5
Age 18–34 (%)	15.3	25.0	−9.7
Age 35–54 (%)	37.5	32.2	+5.3
Age 55+ (%)	47.2	42.8	+4.4
Secondary education (%)	40.9	57.8	−16.9
Higher education, short (%)	34.3	26.6	+7.7
Higher education, long (%)	24.9	15.6	+9.3
Partnered (%)	70.6	57.6	+13.0
Household size	2.42	2.14	+0.28
<i>Well-being at baseline</i>			
Life satisfaction (0–10)	6.42	6.6	−0.18
Happiness yesterday (0–10)	6.43	7.2	−0.77
Depression yesterday (0–10)	3.07	2.1	+0.97
Loneliness, single item (0–10)	3.55	2.5	+1.05
Observations	7,286		

Notes. Sociodemographic characteristics use INSEE’s 2023 census for France entière (excluding Mayotte): [age and sex](#), [education](#), and [marital status and households](#). Age and sex shares are recalculated for adults aged 18+. Education shares cover non-enrolled persons aged 15+ with at least one diploma, excluding those with no diploma; secondary includes brevet, CAP/BEP and baccalauréat; short higher education includes bac+2 through bac+4; long higher education is bac+5 or above (Master or PhD). The partnered benchmark sums married, PACS and cohabiting persons among those aged 15+. Household size is persons in households divided by households.

Well-being metrics for France are taken from the June 2026 wave of the quarterly survey conducted by the CEPREMAP Well-Being Observatory, which is appended to INSEE’s monthly household survey (Enquête mensuelle de conjoncture auprès des ménages, CAMME). [\[LINK\]](#). The dashboard’s 7.9 for anxiety/depression and 7.5 for solitude are reversed as $10 - x$ to match the sample’s coding (higher means more distress). The sample’s loneliness question asks about the previous *two weeks*, versus the dashboard’s *seven days*.

Table 3 First stage: effect of assignment on AI use

	Days of personal conversation (0–28)	Broad AI use (z)	Socio-emotional AI use (z)
Treatment	5.155*** (0.252)	0.429*** (0.023)	0.460*** (0.026)
Median (control group)	2	—	—
First-stage F	417.9	343.2	324.0
Observations	6,768	7,286	7,286

Notes. ITT effect of assignment on compliance, measured at endline. Days of personal conversations with AI are self-reported. The broad AI-use index combines eight endline items measuring the frequency of generative AI use across domains covered in the survey (e.g., work, emotional support, companionship, and health information). The socio-emotional AI-use index combines three endline items (emotional support, companionship, and relationship advice), each measured on a 1–6 frequency scale. For both indices, respondents who did not use AI at endline are assigned the minimum value (1 = never), so the measures are defined on the full sample. Both indices are standardized using the control-group mean and standard deviation. All regressions control for pre-specified covariates. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 Effect of daily personal conversations with AI on subjective well-being

Outcome	ITT (orig. units)	ITT (SD)	Control mean	RW p	N
Life satisfaction (0–10)	-0.056 (0.036)	-0.029 (0.019)	6.96	0.344	7,286
Loneliness, UCLA-3 (3–9)	0.045 (0.028)	0.025 (0.016)	4.68	0.344	7,286
Loneliness, single item (0–10)	0.163*** (0.052)	0.056*** (0.018)	3.47	0.001	7,286
Happiness (0–10)	-0.027 (0.040)	-0.013 (0.019)	6.97	0.495	7,286
Depression (0–10)	0.081 (0.054)	0.029 (0.020)	2.98	0.344	7,286
Baseline outcome & covariates	Yes	Yes			

Notes. ANCOVA ITT estimates (Eq. 1): each row is a separate regression of the endline outcome on treatment assignment, controlling for the baseline value of the outcome and pre-specified covariates. Control means are endline means in the control group on the estimation sample. Robust standard errors in parentheses. Column 4 reports Romano–Wolf stepdown p -values within the subjective well-being family (4,999 bootstrap replications). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5 Effect of daily personal AI conversations on social connectedness

Outcome	ITT	Control mean	RW <i>p</i>	<i>N</i>
<i>Relationship quality and perceived social support</i>				
Partner satisfaction (0–10)	-0.065 (0.043)	7.48	0.415	5,056
Family satisfaction (0–10)	-0.056 (0.038)	7.21	0.415	7,224
Friendship satisfaction (0–10)	0.004 (0.037)	6.97	0.926	7,286
Has someone to rely on (0/1)	0.002 (0.006)	0.91	0.926	7,286
<i>Objective contacts (days out of 7)</i>				
Lunches alone	0.061 (0.045)	2.84	0.307	7,286
Dinners alone	0.056 (0.038)	2.17	0.307	7,286
Meetings with friends/relatives	-0.061 (0.039)	2.14	0.307	7,286
Baseline outcome & covariates	Yes			

Notes. ANCOVA ITT estimates in original units, each estimated on its own complete-case sample (*N* in the last column). Partner satisfaction is estimated on the subsample of respondents partnered at baseline (*N* = 5,056). Family satisfaction is estimated among respondents with a non-missing baseline family-satisfaction outcome. Perceived social support is an indicator equal to one if the respondent reports having at least one person on whom they can rely in case of need. Romano–Wolf stepdown *p*-values are computed separately within the relationship-quality and perceived-social-support family and within the objective-contact family. Robust standard errors in parentheses. **p* < 0.10, ***p* < 0.05, ****p* < 0.01.

Table 6 Effect on how participants value and experience AI

Outcome	ITT (SD)	Control mean/median	<i>N</i>
Conversations pleasant (0–10)	0.094*** (0.029)	6.59 (mean)	4,541
AI “took my side” in conflicts (0-10)	0.021 (0.034)	6.34 (mean)	3,363
WTA to forgo AI for one week	-0.014 (0.034)	50 (median)	3,240
Covariates	Yes		

Notes. ITT estimates on standardized endline outcomes, controlling for pre-specified covariates without baseline adjustment. The estimation samples are smaller than the primary analysis sample because these questions were only asked of participants who had used generative AI during the experiment, and respondents could select “not applicable” for the pleasantness and conflict items if they had not had the relevant type of conversation. For willingness to accept (WTA), the estimation sample is smaller because the question was only asked to AI users, and a substantial share of users did not respond. Because the distribution of willingness-to-pay is heavily right-skewed by a small number of extreme values, we report the control median rather than the mean. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

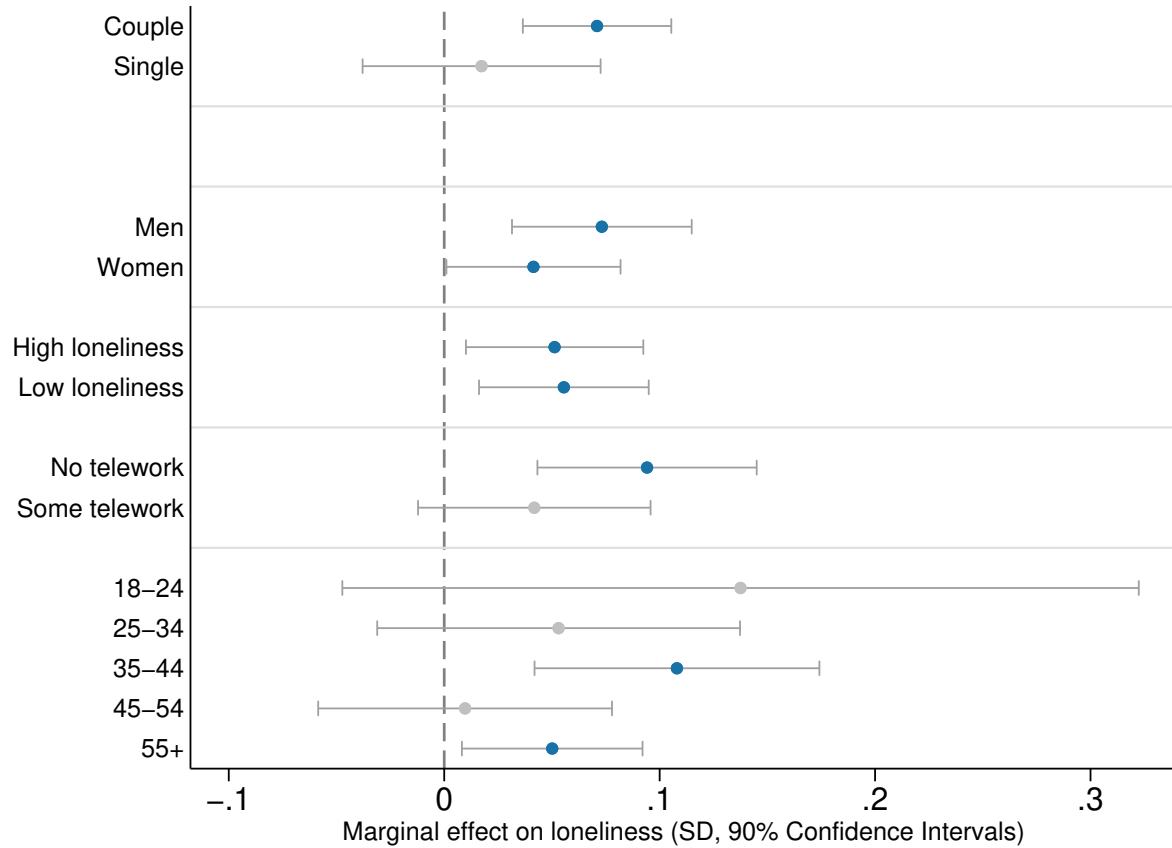


Figure 1 Heterogeneity of treatment effect on loneliness (single-item), by moderators at baseline

Notes. Each panel interacts treatment with one pre-specified moderator in the main ANCOVA on the single-item loneliness measure (0–10), controlling for the baseline outcome and all primary covariates. Dots indicate marginal treatment effects (in standard deviations); horizontal bars are 90% confidence intervals based on robust standard errors. Filled dots denote effects significant at the 10% level. *Remote work* — respondents who never telework vs. those who telework at least occasionally (among employed respondents only, median split).

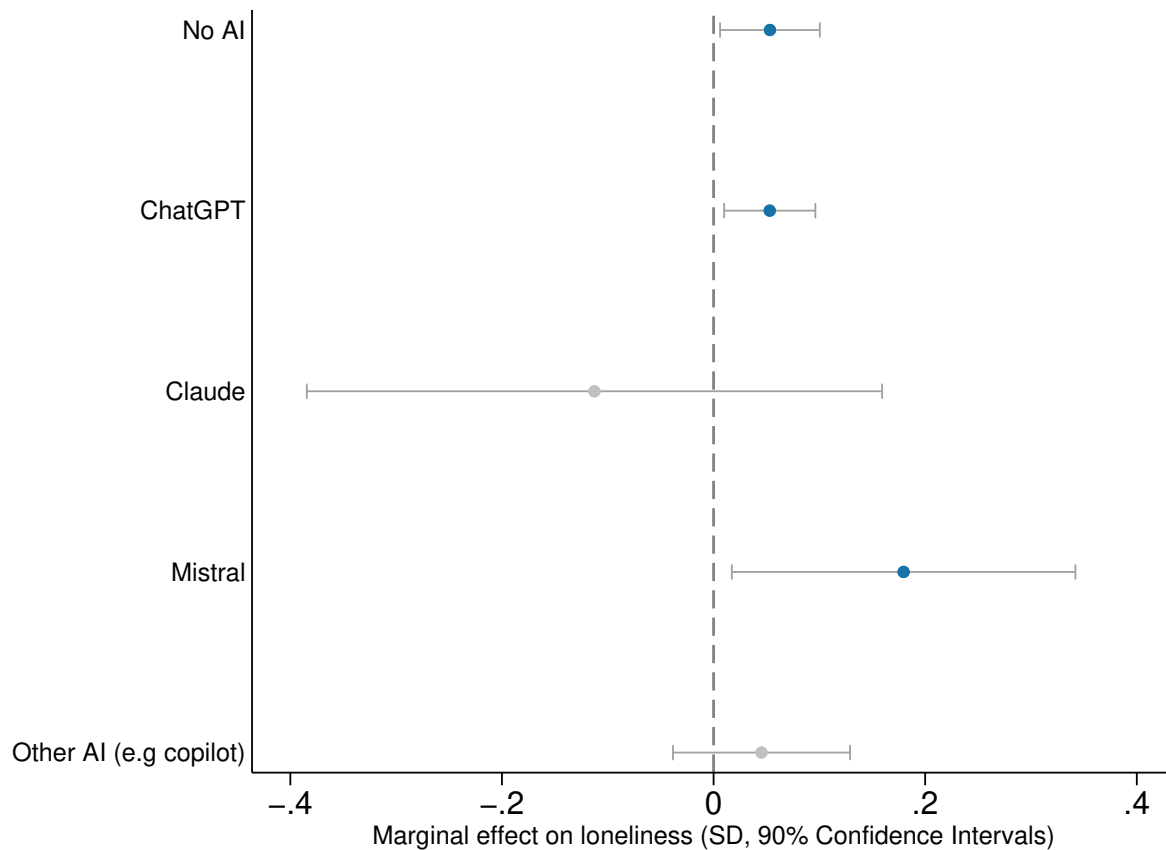


Figure 2 Heterogeneity of treatment effects by baseline AI used most of the time

Notes. Treatment effects are average marginal effects on the standardized single-item loneliness outcome. All regressions control for baseline loneliness and pre-specified covariates. Robust standard errors are reported. This heterogeneity analysis was not pre-specified and is exploratory. 39% of participants had not used AI at all before the experiment, 47% mainly used had used ChatGPT, 2.7% mainly used Mistral, and around 1% mainly used Claude. 11% of participants had used other AI tools (typically Google Gemini or Microsoft Copilot) prior to the experiment.

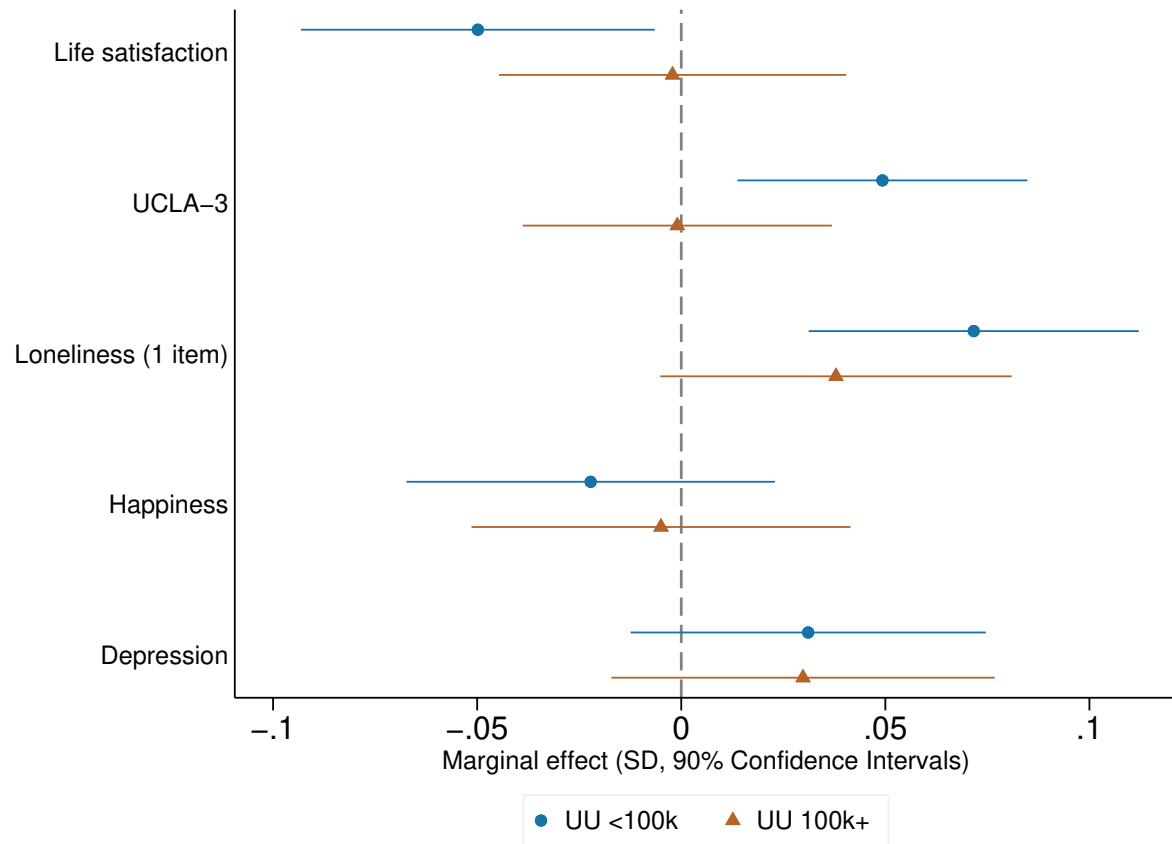


Figure 3 Heterogeneity analysis by urban area population (primary outcomes, family 1)

Notes. Each panel interacts treatment with an indicator for urban units with fewer than 100,000 inhabitants, controlling for the baseline outcome and all prespecified covariates. Standard errors are robust. The urban-unit split follows the INSEE TUU classification: categories 0–5 correspond to rural areas and urban units below 100,000 inhabitants, while categories 6–8 correspond to urban units with at least 100,000 inhabitants, including the Paris urban unit. This heterogeneity analysis was not pre-specified and is exploratory.

References

- Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3), 629–676.
- Angelucci, M., Fabregas, R., & Vazquez, A. (2026). *The well-being effects of digital mental health care*. IZA Discussion Paper No. 18538.
- Babin, J. J., Foray, M., & Hussey, A. (2021). Shelter-in-place orders, loneliness, and collaborative behavior. *Economics & Human Biology*, 43, 101056.
- Berlingieri, F., d’Hombres, B., & Kovacic, M. (2025). *Disentangling loneliness and trust in populist voting behaviour in Europe*. Ca’ Foscari University of Venice, Department of Economics Research Paper Series No. 11/2025. <https://doi.org/10.2139/ssrn.5368679>.
- Bick, A., Blandin, A., & Deming, D. (2025, November 13). The state of generative AI adoption in 2025. *Federal Reserve Bank of St. Louis, On the Economy Blog*. <https://www.stlouisfed.org/on-the-economy/2025/nov/state-generative-ai-adoption-2025>.
- Braghieri, L., Levy, R., & Makarin, A. (2022). Social media and mental health. *American Economic Review*, 112(11), 3660–3693.
- Cox, D. A. (2021). *The state of American friendship: Change, challenges, and loss*. Survey Center on American Life, American Enterprise Institute. <https://www.americansurveycenter.org/research/the-state-of-american-friendship-change-challenges-and-loss/>
- Cox, D. A., & Pressler, S. (2024). *Disconnected: The growing class divide in American civic life*. Survey Center on American Life, American Enterprise Institute. <https://www.americansurveycenter.org/research/disconnected-places-and-spaces/>
- De Freitas, J., Oğuz-Uğuralp, Z., Uğuralp, A. K., & Puntoni, S. (2026). AI companions reduce loneliness. *Journal of Consumer Research*, 52(6), 1126–1148.
- De Neve, J.-E., Dugan, A., Kaats, M., & Prati, A. (2026). Sharing meals is associated with greater wellbeing. *Scientific Reports*, 16, 18741. <https://doi.org/10.1038/s41598-026-46771-9>.
- Donati, D., Durante, R., Sobbrío, F., & Zejcirovic, D. (2025). Lost in the net? Broadband internet and youth mental health. *Journal of Health Economics*, 103.

- Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). *How AI and human behaviors shape psychosocial effects of extended chatbot use: A longitudinal randomized controlled study* [Working paper]. MIT Media Lab / OpenAI.
- Giuliano, P., & Wacziarg, R. (2020). *Who Voted for Trump? Populism and Social Capital*. NBER Working Paper No. 27651. National Bureau of Economic Research. <https://doi.org/10.3386/w27651>.
- Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: A meta-analytic review. *PLOS Medicine*, 7(7), e1000316.
- Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys:: Results From Two Population-Based Studies. *Research on Aging*, 26(6), 655–672.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
- Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38), 16489–16493.
- Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences*, 119(42), e2210412119.
- Kong, N., & Lam, J. (2024). Physical isolation and loneliness: Evidence from COVID lockdowns in Australia. *Journal of Economic Behavior & Organization*, 224, 598–623.
- Krekel, C., De Neve, J.-E., Fancourt, D., & Layard, R. (2021). A local community course that raises wellbeing and pro-sociality: Evidence from a randomised controlled trial. *Journal of Economic Behavior & Organization*, 188, 322–336.
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3), 1071–1102.
- Li, R.-N., Folk, D., Singh, A., Ungar, L., & Dunn, E. (2026). Is a random human peer better than a highly supportive chatbot in reducing loneliness over time? *Journal of Experimental Social Psychology*, 125, 104911.
- McDool, E., Powell, P., Roberts, J., & Taylor, K. (2020). The internet and children’s psychological wellbeing. *Journal of Health Economics*, 69, 102274.

- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2), 210–221.
- Morrish, N., & Medina-Lara, A. (2021). Does unemployment lead to greater levels of loneliness? A systematic review. *Social Science & Medicine*, 287, 114339.
- Mund, Marcus, Marlies Maes, Pia M. Drewke, Antonia Gutzeit, Isabel Jaki, and Pamela Qualter. 2023. “Would the Real Loneliness Please Stand Up? The Validity of Loneliness Scores and the Reliability of Single-Item Scores.” *Assessment* 30(4): 1226–1248. <https://doi.org/10.1177/10731911221077227>.
- Murthy, V. H. (2023). *Our epidemic of loneliness and isolation: The U.S. Surgeon General’s advisory on the healing effects of social connection and community*. U.S. Department of Health and Human Services.
- Myers, C. K., & Hooper, E. (2026). *Is the iPhone birth control? Causal evidence from AT&T’s 2007–2011 carrier monopoly*. NBER Working Paper No. 35310.
- Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology*, 100(4), 569–582.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192.
- OECD. 2013. *OECD Guidelines on Measuring Subjective Well-being*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264191655-en>.
- Pei, Rui, and Jamil Zaki. 2025. “Connecting with Others: How Social Connections Improve the Happiness of Young Adults.” In *World Happiness Report 2025*, edited by John F. Helliwell, Richard Layard, Jeffrey D. Sachs, Jan-Emmanuel De Neve, Lara B. Aknin, and Shun Wang. Oxford: Wellbeing Research Centre, University of Oxford. <https://doi.org/10.18724/whr-zc5n-4f96>.
- Peterson, D., Rooduijn, M., Hopp, F. R., Schumacher, G., & Bakker, B. N. (2025). Loneliness is positively associated with populist radical right support. *Social Science & Medicine*, 366, 117676. <https://doi.org/10.1016/j.socscimed.2025.117676>.
- Rebecchi, A., Lepinteur, A., Clark, A. E., Rohde, N., Vögele, C., & D’Ambrosio, C. (2024). Loneliness during the COVID-19 pandemic: Evidence from five European countries. *Economics & Human Biology*, 55, 101427.

- Reinwarth, Anna Celine, Mareike Ernst, Lina Krakau, Elmar Brähler, and Manfred E. Beutel. 2023. “Screening for Loneliness in Representative Population Samples: Validation of a Single-Item Measure.” *PLOS ONE* 18(3): e0279701. <https://doi.org/10.1371/journal.pone.0279701>.
- Schöne, J. P., Salecha, A., Lyubomirsky, S., Eichstaedt, J., & Willer, R. (2025). *Structured AI dialogues can increase happiness and meaning in life* [Working paper]. <https://psyarxiv.com/2bf7t>
- World Health Organization. (2023, November 15). *WHO launches commission to foster social connection*. <https://www.who.int/news/item/15-11-2023-who-launches-commission-to-foster-social-connection>
- Zao-Sanders, M. (2025, April 9). How People Are Really Using Gen AI in 2025 *Harvard Business Review*.
- Zuckerberg, M. (2025, April 29). *Interview with Dwarkesh Patel* [Video podcast]. YouTube. <https://www.youtube.com/watch?v=rYXeQbTuVl0>

Online Appendix

A Additional results

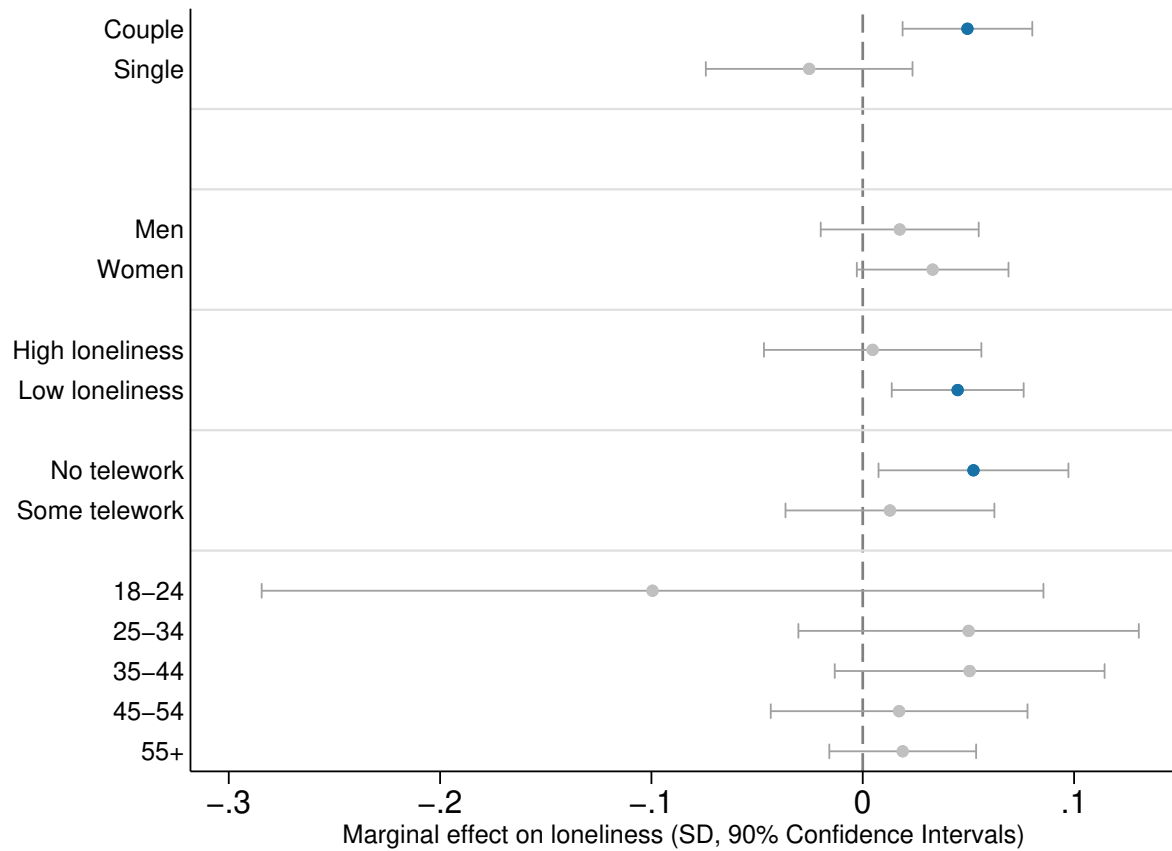


Figure A1 Heterogeneity of treatment effects on loneliness (UCLA-3)

Notes. Each panel interacts treatment with one pre-specified moderator in the main ANCOVA on the UCLA-3 measure, controlling for the baseline outcome and all primary covariates. Dots indicate marginal treatment effects (in standard deviations); horizontal bars are 90% confidence intervals based on robust standard errors. The *Remote work* categories refer to respondents who never telework vs. those who telework at least occasionally (among employed respondents only, median split).

Table A1 Socio-emotional AI use and well-being at baseline (correlational)

	Life sat. (0–10)	Loneliness UCLA-3	Loneliness item (0–10)	Happiness (0–10)	Depression (0–10)
Socio-emotional AI use (z)	−0.050* (0.029)	0.192*** (0.022)	0.456*** (0.034)	−0.026 (0.028)	0.391*** (0.036)
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.059	0.083	0.098	0.039	0.070
Observations	7,286	7,286	7,286	7,286	7,286

Notes. OLS at baseline on the primary analysis sample. Each column is a separate regression of the outcome on the standardized socio-emotional AI-use index, controlling for age, gender, education, relationship status, number of friends at age fifteen, and household size. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2 Baseline covariate for samples in Table 6

	Pleasantness ($N = 4,541$)	Sided with me ($N = 3,363$)	WTA ($N = 3,240$)
Age (years)	+0.037	+0.033	+0.079
Female (%)	+0.005	+0.029	+0.004
Higher education, long (%)	+0.007	+0.020	−0.016
In a relationship (%)	+0.013	+0.041	+0.018
Household size	−0.026	−0.009	−0.004
Children under 15	−0.005	+0.018	−0.018
Economically active (%)	+0.021	+0.014	+0.025
Life satisfaction (0–10)	+0.064	+0.076	+0.023
Loneliness, UCLA-3	−0.053	−0.068	−0.004
Loneliness, single item (0–10)	−0.062	−0.074	+0.008
Happiness (0–10)	+0.070	+0.081	+0.074
Depression (0–10)	−0.057	−0.077	−0.032

Notes. Imbens–Rubin normalized differences (treatment – control) within each estimation sample in Table 6.

Table A3 Robustness: difference-in-means vs. ANCOVA

Outcome	ANCOVA (primary spec.)			Difference-in-means		
	ITT	(SE)	N	ITT	(SE)	N
<i>Subjective well-being (Family 1)</i>						
Life satisfaction (0–10)	-0.056	(0.036)	7286	-0.051	(0.046)	7286
Loneliness, UCLA-3 (3-9)	0.045	(0.028)	7286	0.069	(0.042)	7286
Loneliness, single item (0–10)	0.163***	(0.052)	7286	0.201***	(0.068)	7286
Happiness (0–10)	-0.027	(0.040)	7286	-0.009	(0.048)	7286
Depression (0–10)	0.081	(0.054)	7286	0.114*	(0.066)	7286
<i>Relationship quality (Family 2)</i>						
Partner satisfaction (0–10)	-0.065	(0.043)	5056	-0.095	(0.060)	5056
Family satisfaction (0–10)	-0.056	(0.038)	7224	-0.064	(0.051)	7224
Friendship satisfaction (0–10)	0.004	(0.037)	7286	-0.011	(0.050)	7286
Has someone to rely on (0/1)	0.002	(0.006)	7286	-0.004	(0.007)	7286
<i>Objective contact (Family 3)</i>						
Lunches alone (0–7)	0.061	(0.045)	7286	0.115*	(0.061)	7286
Dinners alone (0–7)	0.056	(0.038)	7286	0.074	(0.064)	7286
Meetings with friends/relatives (0–7)	-0.061	(0.039)	7286	-0.024	(0.044)	7286

Notes. The ANCOVA specification (columns 1–3) controls for the baseline value of the outcome and pre-specified covariates. The difference-in-means specification (columns 4–6) regresses the endline outcome on treatment assignment only, without baseline adjustment (pre-specified robustness check). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4 Lee (2009) bounds on treatment effects (subjective well-being)

Outcome	Lower bound		Upper bound	
	Coef.	95% CI	Coef.	95% CI
<i>Subjective well-being (Family 1)</i>				
Life satisfaction (0–10)	-0.055	[-0.157, 0.047]	-0.043	[-0.169, 0.082]
Loneliness, UCLA-3 (3-9)	0.064	[-0.040, 0.167]	0.071*	[-0.013, 0.154]
Loneliness, single item (0–10)	0.193**	[0.034, 0.353]	0.205***	[0.060, 0.350]
Happiness (0–10)	-0.012	[-0.121, 0.097]	-0.001	[-0.134, 0.133]
Depression (0–10)	0.106	[-0.057, 0.269]	0.117*	[-0.017, 0.251]

Notes. Attrition rates are 25.32% in the control group and 25.24% in the treatment group, a difference of -0.09 percentage points ($p = 0.92$). The corresponding Lee trimming proportion is 0.12%. Bootstrap standard errors, 2,000 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Discrepancies between versions of this working paper

The current version of this working paper differs from the previous one by a smaller number of observations. This is because one of the intermediate files that were communicated to us by Bilendi contained several lines for the same individual respondent. We became aware of this mistake later and corrected the estimation sample for this version. We could have suspected the problem by noticing that the sample was larger than initially planned in the PAP. However, we attributed this to the fact that Bilendi had to contact more respondents than planned in order to try and maintain the desired sample stratification, particularly to recruit younger respondents. Note that the decision to extend the sample was made prior to any endline data collection, and we did not access data before the endline was completed.