Structured AI Dialogues Can Increase Happiness and Meaning in Life

Jonas P. Schöne [1, 2], Aadesh Salecha[1], Sonja Lyubomirsky [3],

Johannes C. Eichstaedt [1], and Robb Willer [1, 2]

[1] Stanford University, Institute for Human-Centered Artificial Intelligence, Stanford, CA 94305

[2] Stanford University, Department of Sociology, Stanford, CA 94305

[3] UC Riverside, Department of Psychology, Riverside, CA 92521

Working paper, not yet peer reviewed

Corresponding Authors: Jonas P. Schöne (jschoene@stanford.edu), Johannes C. Eichstaedt (johannes.stanford@gmail.com), and Robb Willer (willer@stanford.edu)

ABSTRACT

Millions of people now use AI-powered chatbots to support their mental health, yet little is known about whether such interactions can effectively enhance psychological well-being. We conducted a preregistered experiment on a large, diverse sample (N = 2,922) to test four AI chatbots, each prompted to employ a multi-step strategy drawn from prior psychological research on sources of happiness and meaning in life. Chatbots encouraged participants to either (a) savor positive life experiences, (b) express gratitude toward a friend or family member, (c) reflect on sources of meaning in their life, or, (d) reframe their life story as a "hero's journey." All four chatbots led to improvements on a broad range of psychological well-being outcomes – including affective well-being, meaning in life, life satisfaction, anxiety, and depressed mood – relative to a control chatbot condition. These results generalized to key subpopulations, including those with high baseline levels of anxiety or depression. Chatbot interactions increased interest in seeing a human therapist, including among those who were previously unwilling or had never attended therapy. A separate, nationally representative survey (N = 3,056) found that half of U.S. adults expressed interest in using empirically validated AI chatbots for mental health support. These findings demonstrate that AI-driven well-being chatbots grounded in psychological research offer a scalable and effective way to produce short-term increases in several aspects of psychological well-being. Importantly, these results do not generalize to all AI-based emotional support.

Keywords: Artificial Intelligence, Happiness, Large Language Models, Positive Psychology, Well-Being, Affect

MAIN

With the rise of generative artificial intelligence (AI) capable of producing human-like text rapidly at scale, many speculate the latest generation of chatbots has the potential to support social and emotional well-being by enabling the design of broadly accessible interventions [1,2]. Further, millions of people are already engaging with AI chatbots for social companionship, advice, and emotional support [3]. Additionally, companies have begun to offer access to AI chatbot therapists and life coaches, claiming these tools are effective, providing benefits to users comparable to human therapists [4–6]. These informal and formal uses of AI chatbots to support mental health and well-being are attractive, given the currently limited and unequal access to human therapists and other mental health resources. Public use and commercial claims, however, far exceed systematic empirical research on the capacities of AI chatbots to provide emotional benefits to humans [7–9]. Given increasing adoption, an important question remains: Can conversations with AI chatbots improve humans' psychological and emotional well-being, and which empirically grounded strategies make this possible?

Here, we specifically evaluate the effects of structured AI chatbot dialogues designed to implement interventions drawn from psychological research on factors improving psychological well-being[1] (well-being chatbots). Prior research has empirically evaluated interventions designed to enhance different aspects of psychological well-being, most commonly positive and negative emotions (henceforth, 'well-being chatbots') and cognitive evaluations of one's life, such as life satisfaction [10,11]. Additionally, researchers have studied interventions targeting related constructs like meaning in life [12,13]. Although

[1] Readers can explore each of the AI chatbots presented in this paper on our website.

effect sizes are generally modest [14–16], advances in large language models (LLMs) offer an opportunity to make interventions more engaging and accessible. Importantly, our validation extends only to these structured, evidence-based dialogues. It does not generalize to differently structured or ad hoc uses of AI chatbots for mental health, which lack theoretical grounding and may carry risks [17].

Well-powered, preregistered experiments evaluating well-being interventions, remain scarce [15]. Further, no studies to date have rigorously evaluated application of such interventions via LLM-powered chatbots. There is also little understanding of the mechanisms by which such tools might lead to improvements, or their potential for widespread, sustained adoption.

In the present study, we address these gaps, evaluating AI chatbots designed to improve subjective well-being, meaning in life, anxiety, and depressed mood in the general population. While some AI chatbots are being developed to complement or even replace human therapists in treating mental illness, responsibly evaluating such tools in clinical populations requires especially careful implementation [2,5]. By contrast, testing psychological well-being interventions in a large, diverse community sample - as we do in the current study – covers the full continuum of well-being while carrying substantially lower risk. Findings from such research can inform the design of AI chatbots for clinical use. To that end, we measure participants' prior experience with human therapists, as well as their reported levels of anxiety and depressed mood before and after the intervention. Finally, we examine whether engaging with AI well-being chatbots changes individuals' interest in seeing a human therapist, since a possible risk of widely accessible well-being chatbots is that they could inadvertently reduce engagement with traditional mental health services, to the detriment of long-term mental health.

Factors Influencing Psychological Well-Being

For decades, psychologists have psychological well-being, focusing on factors influencing three broad categories: *hedonic well-being*, defined as the presence of positive emotions and the absence of negative emotions [18]; *eudaimonic well-being*, which refers to feelings of meaning and purpose [19]; and *evaluative well-being*, which refers to cognitive assessments of life satisfaction and overall quality of life [20]. These forms of psychological well-being are shaped by personality traits (e.g., extraversion and neuroticism [21]), life circumstances (e.g., health, wealth, and relationships [22]), but also by people's thought patterns and activities [23].

Building on the observation that deliberate activities can enhance psychological well-being, much prior psychological research has focused on evaluating interventions [24–26] .Interventions aimed at enhancing hedonic and evaluative well-being include activities like writing letters expressing gratitude to others [27], savoring positive experiences [27–29], focusing on personal strengths [30], thinking optimistically [31,32], and engaging in meditation [33]. Scholars have also identified interventions enhancing eudaimonic well-being, including spending time in nature [34], building and strengthening social connections [35,36], and engaging in therapy or coaching [12,37]. Many of these practices have entered the public sphere and media [15] as accessible tools for improving day-to-day happiness and emotional resilience.

Human Interactions with AI Chatbots

Here, we evaluate the effectiveness of embedding evidence-based well-being strategies from psychological research into AI chatbots. This approach depends on people being receptive to AI-delivered support, an assumption supported by recent research. Recent polling suggests as much as 34% of Americans have engaged with AI chatbots, with mental health and/or emotional support being a primary goal of their usage [3,38]. Additionally, recent studies find people form meaningful bonds with AI companions [9], with younger adults in particular turning to AI chatbots as a source of social and emotional support [39]. Research finds AI chatbots can express empathy and support cognitive reappraisal of users' negative emotions, though responses are perceived as more supportive when believed to be human-generated [40]. People also report feeling heard and seen in AI chatbot conversations [41], and some evidence suggests even unstructured conversations increase users' happiness on average [42]. In some domains, AI chatbots have outperformed humans, for example, being rated as more compassionate than crisis responders [43] and detecting suicidal ideation at levels exceeding mental health professionals [44]. This emerging body of research supports the idea that AI chatbots can engage and influence people across different conversation topics and purposes, including well-being-relevant ones, suggesting AI chatbots may be effective in delivering psychological interventions.

Other evidence, however, raises doubts about this use of AI chatbots. Many users express a clear preference for human interaction, especially for emotional concerns where the user believes a response requires personal experience and would rather wait than receive AI support [40]. Even when people recognize the capabilities of chatbots, they often perceive relationships with AI companions as one-sided, believing AI lacks the capacity for genuine understanding or emotion [45]. When chatbots produce apparently empathetic responses, the discovery that responses were AI-generated can diminish their emotional impact [46] (but see also [43]). More concerning are signs that chatbots may reinforce harmful beliefs due to their tendency to be overly agreeable [47]. When cast in therapeutic roles, some AI systems have demonstrated stigmatizing attitudes toward users with mental illness, encouraged delusional thinking, or failed to recognize crises [17]. Together, these findings suggest dialogues with AI chatbots, even when structured to employ techniques drawn from prior research, are unlikely to improve users' well-being.

A critical challenge for design of effective interventions is motivating people to use them [48–50]. Recent evidence suggests, however, that people readily engage with and often seek emotional support from AI chatbots. As noted above, Over a third of the U.S. population (and two-thirds of adults under age 30) report they have interacted with generative AI tools [38], with many turning to chatbots for social companionship, seeking emotional support, and meaning in life [3,51]. Online communities offering guidance on effectively using AI chatbots for mental health support have grown rapidly, some exceeding two million members, offering further evidence for broad public interest in the effective use of AI chatbots to improve well-being [52]. Critically, however, this widespread interest and rapid adoption has outpaced rigorous validation of whether and how AI chatbot dialogues positively impact people's well-being.

Empirical Strategy

To address these gaps, we investigated whether AI chatbots can improve subjective well-being in nonclinical populations by implementing interventions derived from psychological research. We employed a range of measures to study the (a) effects of these structured chatbot conversations, (b) robustness of effects in subpopulations of particular importance, such as those with high baseline levels of state depression and/or anxiety, (c) mechanisms driving any observed effects, and (d) potential future adoption of these tools.

We conducted a large, preregistered experiment (N = 2,936) evaluating the psychological impacts of LLM-based chatbots designed to enhance psychological well-being (see Methods and Materials for more details about the study population). We assessed pre- to post-treatment changes on several outcome variables - affective well-being (reported positive, relative to negative, emotions), meaning in life, life satisfaction, anxiety, and depressed mood - relative to a neutral control chatbot. All outcomes were measured using multiple-item composites drawn from prior research. The study's design, outcome measures, hypotheses, and primary analyses were all reregistered.

Each well-being chatbot was grounded in prior research on factors influencing affective well-being and meaning in life (Fig. 1). The *savoring chatbot* prompted users to "select and relive a positive, impactful experience," vividly remembering it first, and describing the experience and how it made them feel. Most often studied via structured writing or guided imagery exercises, research finds reflecting on past positive experiences can increase positive affect, life satisfaction, and resilience [53,54].

The *gratitude chatbot* prompted users to identify someone they felt grateful toward, explain why they felt this way, and write a short letter they could download at the end of the conversation. Gratitude letter-writing has been found to enhance happiness, strengthen relationships, and reduce depressive symptoms [55], including in well-powered, preregistered experiments [15].

The *meaning chatbot* prompted users to "explore what brings meaning to [their] life…includ[ing] anything from relationships and achievements to personal growth." Participants were guided to reflect on why these aspects were meaningful, how they made them feel, and how they shaped who they are today. The chatbot was based on research showing that people with a greater sense of meaning and purpose in life report less stress and greater psychological well-being and resilience [19,56].

Finally, the *hero's journey chatbot*, inspired by Joseph Campbell's (1949) cross-cultural analysis of the archetypal hero's journey narrative [57], guided participants in reframing personal struggles as transformative experiences that foster growth. Participants were prompted to recount a life-changing challenge, describe obstacles they faced, recognize support they received, and reflect on how they emerged stronger, before being given an AI-generated narrative of their life in the form of the hero's journey. Recent research suggests this type of exercise can improve psychological well-being by helping individuals find meaning in prior adversity [13].

Effects of these interventions were compared to a control chatbot that prompted participants to describe the layout and content of their bedroom, a topic that minimally affected psychological well-being in a pilot test and in prior research [58].

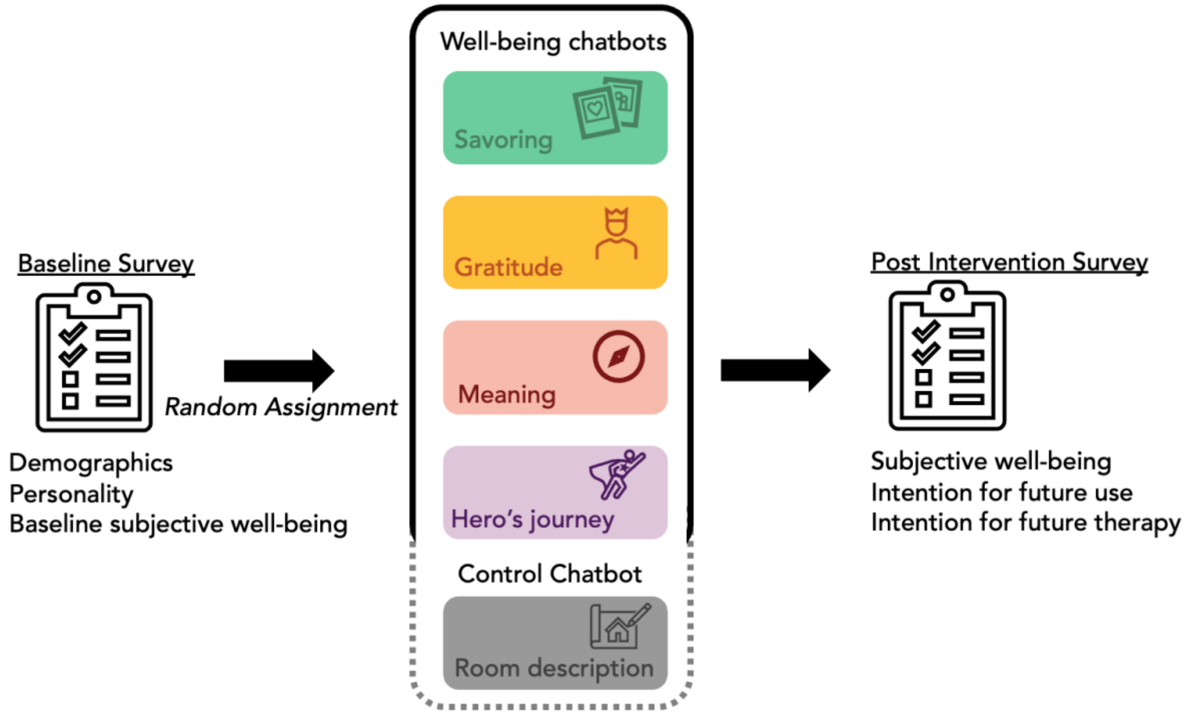


Fig. 1 Design of experiment testing the effectiveness of well-being chatbots, compared to a control condition, with dependent variables administered pre- and post-treatment. After completing the pre-treatment survey, participants were assigned to one of four well-being chatbot interventions or to a control chatbot. Following dialogue with the assigned chatbot, participants completed the post-treatment survey.

RESULTS

Participants' dialogues with a randomly assigned chatbot lasted an average of 9.56 minutes (*SD* = 7.08). Duration of dialogues were longest for participants assigned to the hero's journey chatbot (*M* = 10.97, SD = 8.27) and shortest for those assigned to the control condition (*M* = 8.26, *SD* = 5.56).[2] Participants wrote an average of 814.28 words (*SD* =

[2] Estimated effects of well-being chatbots on psychological well-being outcomes were robust in models controlling for duration of dialogue (see SI Appendix Section 3.4.2).

307.11) across 14.66 conversational turns (*SD* = 5.01), with the AI chatbots producing an average of 790.84 words (SD = 403.15) across 13.65 turns (SD = 5.01).

Effects on Psychological Well-Being

We first tested the effects of interaction with each of the chatbots on participants' reported affective well-being (i.e., their reported positive minus negative emotions), measured using the Positive and negative Affect Schedule (PANAS [59]). As shown in Fig. 2, each of the four well-being chatbots significantly increased affective well-being, with increases ranging from a 7.84 percentage point (hereafter, pp) increase following interaction with the savoring chatbot (b=7.84, 95% CI=[6.51, 9.16], $t(2919)$=11.61, p< .001, d=.41) to a 4.82pp increase following interaction with the meaning chatbot (b=4.78, 95% CI=[3.46, 6.10], $t(2919)$=7.12, p<.001, d=.25) (see SI Appendix Section 3.4.1 for full results). On average, participants assigned to one of the well-being chatbots reported a significant improvement of 6.07pp in affective well-being, relative to participants in the control chatbot group (b=6.04, 95% CI = [5.10, 6.98], $t(2922)$=12.63, p<.001, d=.32). For comparison, these effect sizes are equal to or greater than those from traditional, short-term one session, interventions reported in a recent meta-analysis of pre-registered experiments testing the effects of traditional well-being interventions [15] (for details see SI Appendix Section 1.1 and 3.5).

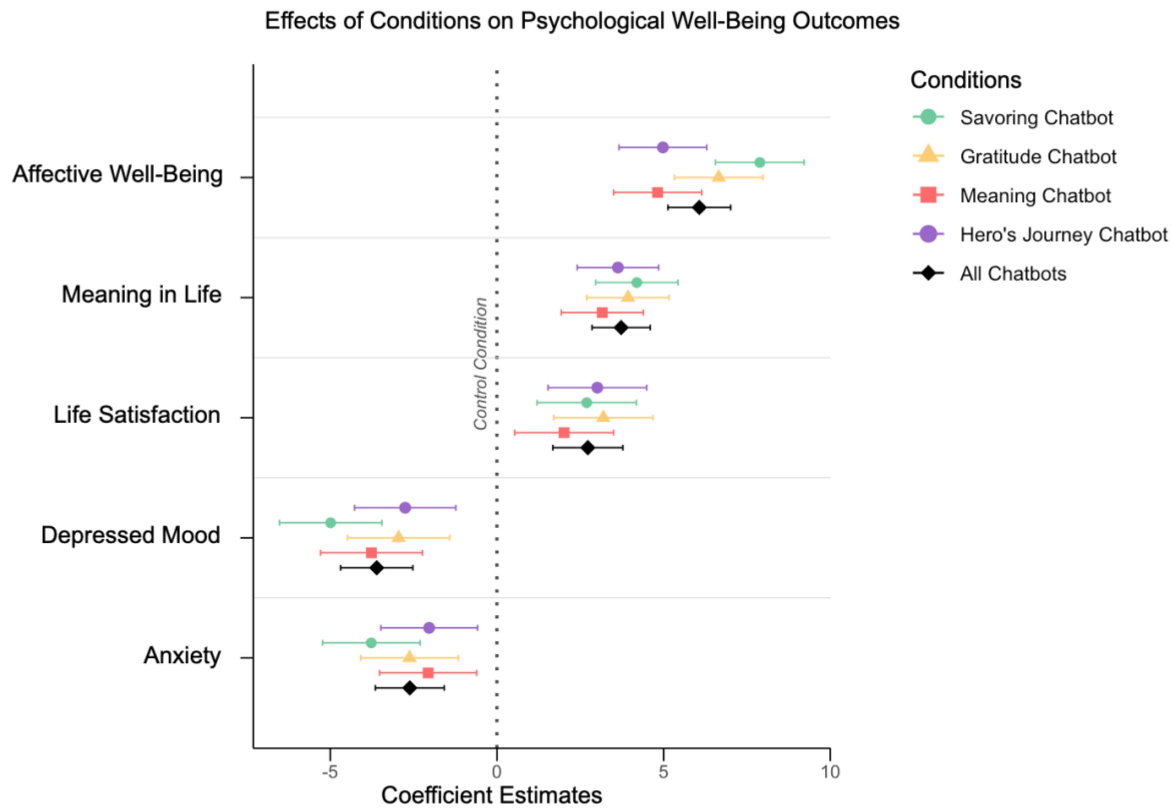


Fig. 2 Unstandardized coefficients from preregistered OLS models estimating effects of savoring, gratitude, meaning, and hero's journey chatbot dialogues as well as the average of all chatbots on outcome variables, relative to the control chatbot. Points indicate the treatment effects on each outcome. Error bars represent 95% confidence intervals. Unstandardized coefficients correspond to the estimated percentage point change on outcomes.

All four well-being chatbots also led to significant improvements in reported meaning in life [60], though effect sizes were more modest. The largest improvement (4.23pp) for participants interacting with the savoring chatbot (b=4.23, 95% CI=[3.00, 5.46], t(2918)=6.75, p<.001, d=.17), while the meaning chatbot produced the smallest effect, an improvement of 3.16pp (b=3.24, 95% CI=[2.01, 4.46], t(2918)=5.19, p<.001, d=.13). Assignment to any of the well-being chatbots led participants to report significantly greater meaning in life, with an overall average increase of 3.72pp (b=3.75, 95% CI=[2.89, 4.62], $t(2921)=8.48$, $p<.001$, d=.16).

Additionally, each chatbot led to significantly greater life satisfaction [61]. The gratitude chatbot produced the largest improvement, increasing life satisfaction by 3.20pp (b=3.20, 95% CI=[1.71, 4.69], $t(2918)=4.22$, $p<.001$, d=.15), while the meaning chatbot led to the smallest impact, a 2.04 percentage point increase (b=2.04, 95% CI = [0.56, 3.52], $t(2918)=2.71$, p=.007, d =.07). On average, participants assigned to any of the well-being chatbot conditions reported increased life satisfaction (2.77pp), relative to the control condition (b=2.77, 95% CI = [1.72, 3.81], $t(2921)=5.17$, $p<.001$, d=.10).

In exploratory analysis, we also showed that interactions with the chatbots decreased anxiety [62] and depressed mood [63]. The average decrease of anxiety was -3.60pp (b=-3.60, 95% CI = [-4.68, -2.52], $t(2921)=-6.52$, $p<.001$, d=.-14), while the average decrease in depressed mood was -2.65pp (b=-2.65, 95% CI = [-3.68, -1.62], $t(2921)=-5.04$, $p<.001$, d=.-10). See SI Appendix Section 3.4.1 for all details.

Finally, we explored the effects of well-being chatbot dialogues across four subgroups of unique interest: participants reporting relatively low affective well-being, high levels of anxiety, high levels of depressed mood, or high interest in using validated AI well-being chatbots at baseline, prior to assignment to experimental condition. To compensate for reduced statistical power in subgroup analyses, we analyze effects of interaction with any well-being chatbot, relative to the control chatbot (see SI Appendix, Section 3.4.3 for detailed results). As shown in Fig. 3, we find the effects observed for the full sample were robust in each of these four subgroups. These robustness analyses find improvements on several aspects of subjective well-being among individuals who, prior to dialogues with well-being chatbots, indicated they were experiencing relatively negative affective well-being, high anxiety, or high depressed mood. Robustness of effects also among participants reporting high willingness to use well-being chatbots is significant since these individuals are most likely to engage with these chatbots in real life.

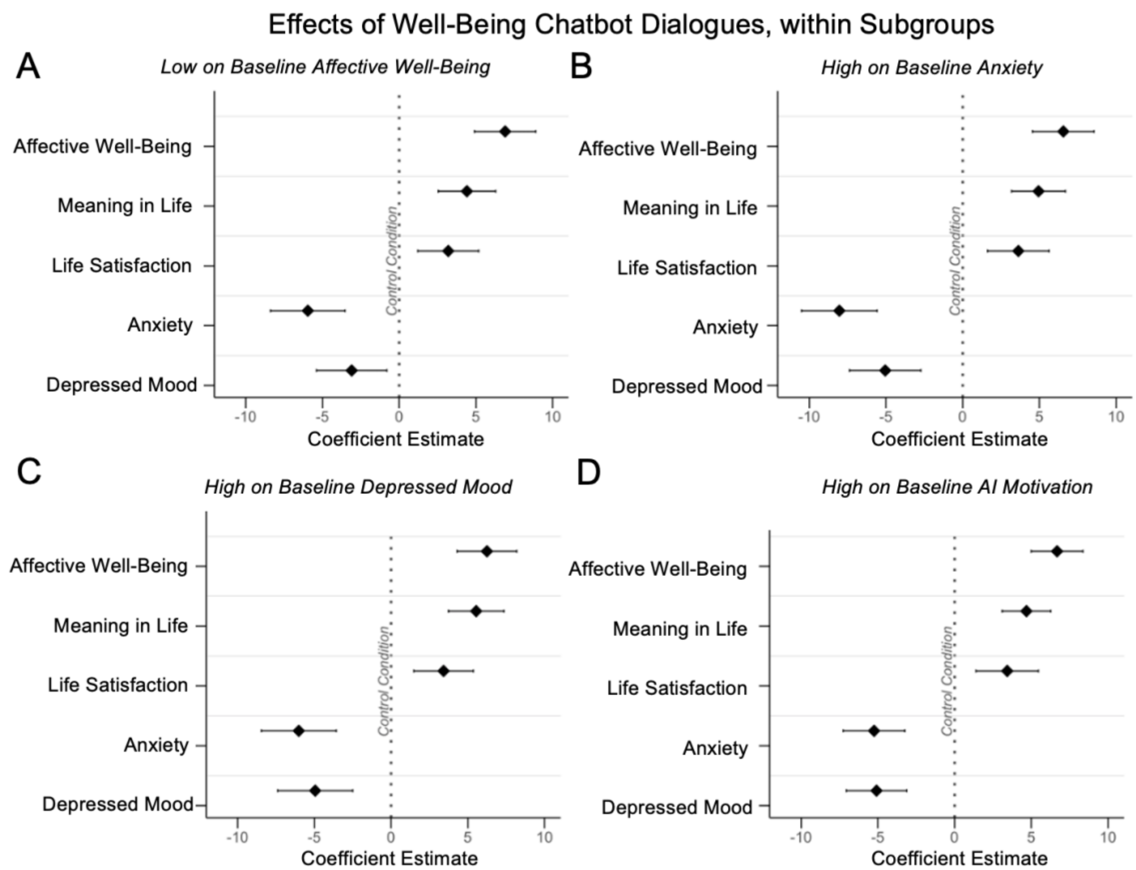


Fig. 3 Unstandardized coefficients from preregistered OLS models for four key subgroups of interest estimating effects of dialogues with a well-being chatbot on primary outcome variables, relative to control. Subgroups represent top (or bottom) third of the distribution for the respective baseline measure. Points indicate treatment effects of well-being chatbots on each outcome. Error bars represent 95% confidence intervals. Unstandardized coefficients correspond to estimated percentage point change on outcomes.

Reactions to Chatbot Conversations

We next examined several variables designed to measure participants' cognitive and emotional experiences. To do this, we asked participants several questions about the extent to which their experience led them to (a) have positive thoughts (*had positive thoughts*), (b) gain new insight on their emotional experiences (*accessed emotions*), (c) feel grateful (*felt grateful*), (d) learn about themselves (*gained self-knowledge*), and (e) feel seen and heard (*felt seen and heard*) (see SI, Section 0.4.1 and 3.6 for full text of items).

As shown in Fig. 4, chatbot dialogues had similar effects on these outcomes, with participants reporting higher levels (relative to control) for all of the measured reactions, in particular higher levels for accessed emotions, felt seen and heard, and had positive thoughts. Beyond these general effects, the savoring and gratitude chatbots had the strongest effects on had positive thoughts, the hero's journey chatbot most increased gained self-knowledge; and the gratitude chatbot, unsurprisingly, most increased felt grateful (see SI Appendix, Section 3.6 for statistics).

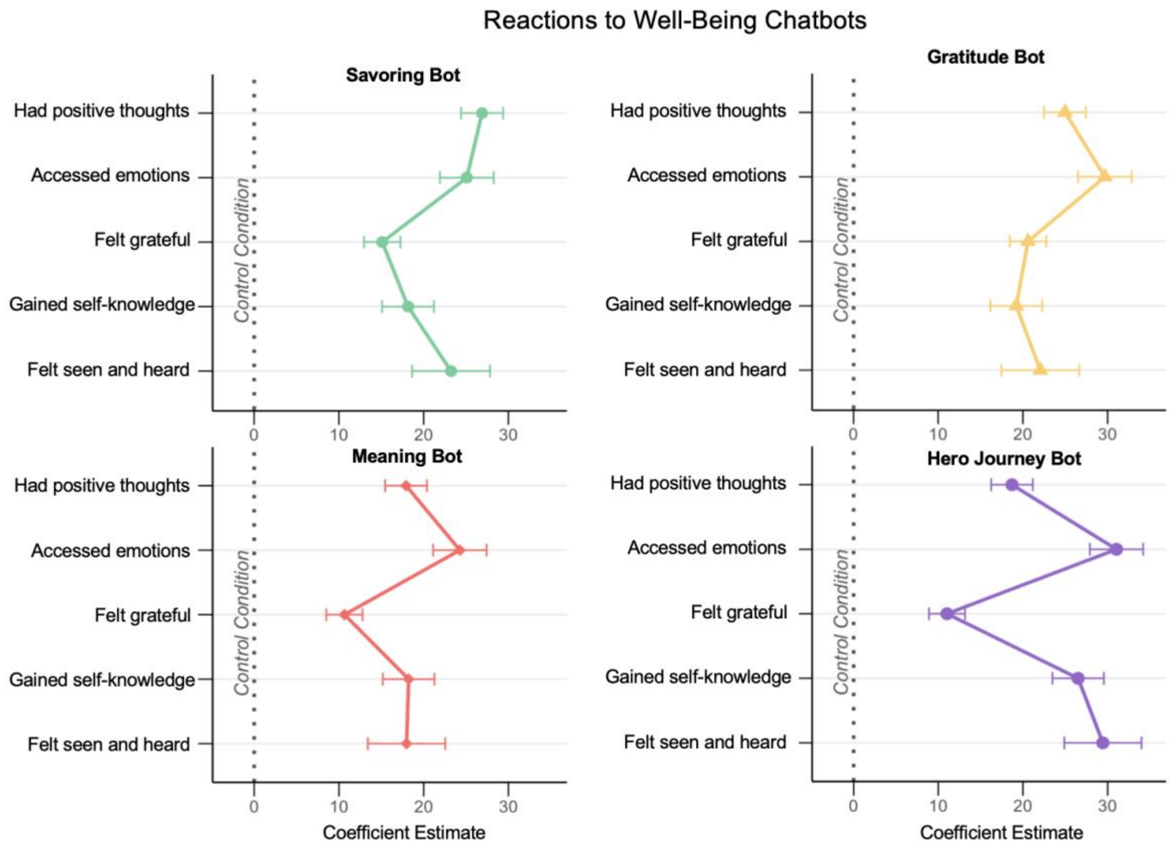


Fig. 4 Unstandardized coefficients from preregistered OLS models estimating effects of savoring (top left), gratitude (top right), meaning (bottom left), and hero's journey (bottom right) chatbot dialogues on conversation reaction outcomes, relative to the control chatbot. Points indicate treatment effects on each outcome. Error bars represent 95% confidence intervals. Unstandardized coefficients correspond to estimated percentage point change on outcomes.

Exploring the Content of Well-Being Chatbot Conversations

We next analyzed the language used by participants in the chatbot interactions, spanning 32,748 conversational turns and 540,315 words. On average, participants engaged in 11.3 conversational turns (*SD* = 4.5), writing 185 words (*SD* = 123). We relied on two methods of language analysis. First, we used established psychological dictionaries (Linguistic Inquiry and Word Count (LIWC); Pennebaker et al., 2022), focusing on pronouns, content themes, and emotion words. Second, as a data-driven, open-vocabulary complement, we modeled 50 Latent Dirichlet Allocation (LDA) topics using the open-source DLATK codebase [64], which determined semantically coherent clusters of words occurring in the participant responses.

We then correlated dictionary and topic frequencies with indicator variables representing each chatbot condition (i.e., one-hot encoded dummy variables for the four well-being chatbots and the control condition). These analyses controlled for pre-treatment levels of affective well-being, age, gender, and the Big Five personality traits, and we adjusted for multiple comparisons using Bonferroni correction [65,66]. Across both dictionary (see SI Appendix, Section 3.7 for all correlations) and topic analyses (see SI Appendix, Section 3.7.2 for all topics), we found face-valid patterns of language use across the chatbot conditions (compared to each other and the control condition) that reflected the themes each chatbot was prompted to focus on, and clear differences in the use of pronouns across conditions (see bolded table cells of pronouns and content themes results in Fig. 5, Top).

Participants' dialogues with the savoring chatbot were characterized by more frequent "we"-pronouns (r=.25, 95% CI=[.21, .29], p<.001, see Fig. 5, top), references to memories (e.g., "remember," "trip," "moment;" r=.39, 95% CI=[.36, .42], p<.001), and positive emotion words (r=.25, 95% CI=[.22, .29], p<.001), reflecting the conditions's focus on recalling a joyful experience. In the gratitude chatbot condition, participants also more often used we-pronouns (r=.20, 95% CI=[.17, .24], p<.001), and showed strong tendencies to use she/he (third-person singular) pronouns (r=.74, 95% CI=[.72, .76], p<.001) and social process terms (e.g., “talk,” “friend,” “kindness”; *r*=.61, 95% CI=[.59, .64], *p*<.001), reflecting the condition’s focus on expressing gratitude for a specific other person.

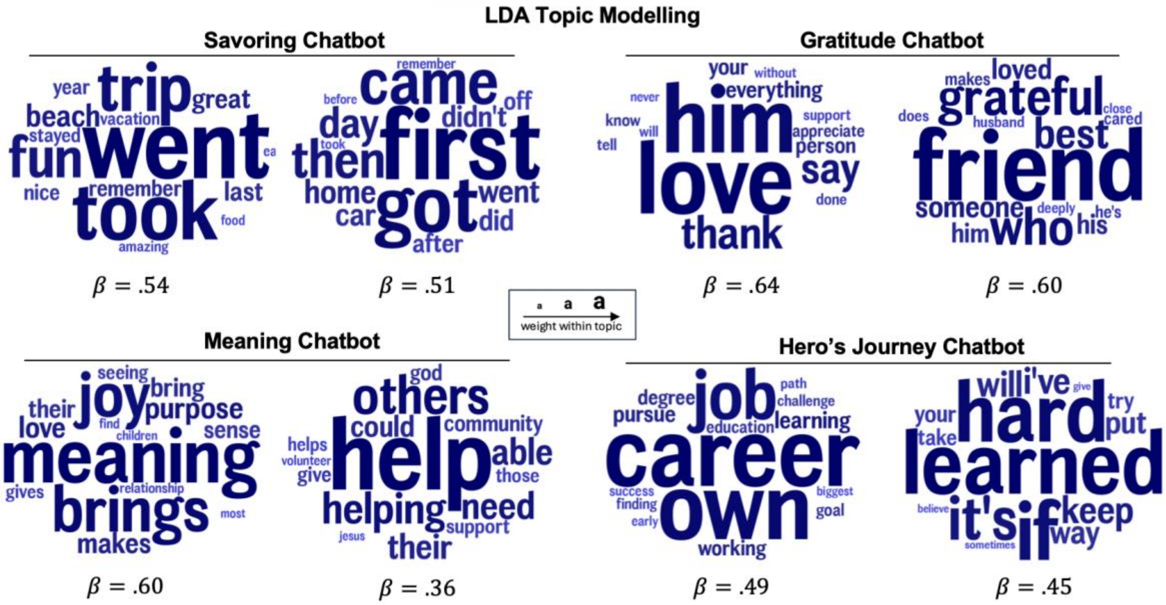
Participants in the meaning chatbot condition were uniquely likely touse “they” pronouns (e.g., “they,” “them,” “their”; *r*=.40, 95% CI=[.36, .43], *p*<.001) and cognitive process words (e.g., “realize”, “think”, “understand”; *r=.25*, 95% CI=[.21, .28], *p*<.001), reflecting the chatbot’s focus on participants’ sources of meaning and purpose. A qualitative review of these sentences containing these word categories suggests that participants often reflected on how their lives positively affect others. Finally, participants in the hero’s journey condition frequently used “I”-pronouns (e.g., “I,” “me,” “my”; *r*=.37, 95% CI=[.35, .41], *p*<.001), referenced cognitive processes (e.g., “know*,”* “decide,” “learn*”;* *r*=.20, 95% CI=[.17, .24], *p*<.001), and a focus on achievement-related language (e.g., “win”, “success”, “goal*;*” *r*=.44, 95% CI=[.41, .47], *p*<.001) This reflects the condition’s focus on past achievements. Notably, unlike the other interventions the hero’s journey conversations were distinguished by more negative emotion words (e.g., “sad,” “hurt,” “angry”; *r*=.20, 95% CI=[.17, .24], *p*<.001), consistent narratives of overcoming adversity.

The topics that distinguished the different conditions similarly reflected face-valid themes for the conditions. (Fig. 5, bottom). Participants interacting with the savoring chatbot interactions referenced memorable experiences, such as meaningful “firsts,” and “trips.”, which could explain why this chatbot increased not only positive emotions, but also meaning as the events are not only positive but also meaningful. In the gratitude condition, participants expressed appreciation for others (“thank,” “love”), and recalled support they had received from friends, family, or mentors. With the meaning chatbot, participants reflected on their “purpose” and “meaning” in life, often focusing on “helping” “others.”

Finally, participants interacting with the hero's journey chatbot referenced "proud" moments and "hard learned" lessons, often involving "overcoming challenges."

Findings from both language analyses closely echoed the mediation results. Across all conditions, emotional expression was central, but more reflective and cognitive processes emerged specifically in the meaning and hero's journey chatbots. In line with the additional outcome analyses, these results show that while each chatbot improved well-being, they did so through distinct conversational pathways, emphasizing different people, themes, and modes of reflection.

Linguistic Inquiry and Word Count (LIWC)					
	Dictionaries	**Savoring Chatbot (β)**	**Gratitude Chatbot (β)**	**Meaning Chatbot (β)**	**Hero's Journey Chatbot (β)**
Pronouns	We	**0.22**	0.16	-0.08	-0.29
	She / He	-0.22	**0.71**	-0.15	-0.34
	They	-0.08	-0.18	**0.36**	-0.10
	I	-0.33	0.07	-0.03	**0.29**
Content Themes	Perceptions	**0.34**	-0.19	-0.01	-0.14
	Social Processess	-0.32	**0.62**	0.03	-0.32
	Cognitive Processess	-0.31	-0.13	**0.23**	0.20
	Achievement	-0.13	-0.20	-0.03	**0.35**
Emotions	Positive Emotions	**0.09**	**0.07**	**0.07**	-0.22
	Negative Emotions	-0.01	-0.05	-0.05	**0.11**



Note. All coefficients have been corrected for multiple comparisons using the Benjamini-Hochberg procedure.

Fig. 5 Language categories and topics significantly associated with different well-being chatbot conditions, adjusted for pre-treatment levels of affective well-being, age, gender, and the Big Five personality traits, controlling for multiple comparisons (DLATK [64]). *Top*: highest correlations for a given dictionary are bolded, associations not significant at corrected thresholds in gray. *Bottom*: Most associated topics are shown, with size of word indicating its prevalence within each topic.

Motivation to Use Validated Well-Being Chatbots

To assess interest in using well-being chatbots whose effects have been evaluated by researchers, we designed and fielded a survey on a sample (N = 3,056) that is representative of the US population on several demographic benchmarks, including age, gender, ethnicity, political affiliation, and education (see SI Appendix, Section 3.3 for quota details). Results showed that 28% of respondents reported having already used generative AI "to improve their emotional well-being or mental health," and 53% reported being willing to try interacting with an AI-powered chatbot "designed to improve people's emotional well-being" that research had shown to have positive effects. These findings indicate quite substantial rates of Americans using AI chatbots for mental health and emotional support, and even greater levels of interest in using well-being chatbots, validated as generally helpful by experimental research (see SI Appendix, Section 3.8.3 for demographics that predict highest motivation).

Returning to our experiment, we tested whether interacting with well-being chatbots (averaged across all chatbot conditions; see SI Appendix, Section 3.8.1 for results by chatbot) increased participants' willingness to use well-being chatbots in the future and also their willingness to see a therapist. Participants in all chatbot conditions reported greater willingness to engage with well-being chatbots, with an average increase of 7.55pp relative to the control (b=7.55, 95% CI [6.10, 9.00], $t(2920)$=10.22, p<.001, d=0.23). All interventions except the meaning chatbot also increased participants' willingness to consider traditional therapy, with an average increase of 2.78pp (b=2.78, 95% CI [1.59, 3.96], $t(2920)$=4.61, p<.001, d=0.08). These effects held for people who never went to therapy before and also for those who had low levels of motivation for AI chatbots at baseline (see

SI Appendix Section 3.8.2 for details).Thus, interacting with well-being chatbots not only boosted openness to future chatbot use but, in most cases, also increased willingness to pursue therapy with a human professional. The latter finding offers evidence against the concern that conversations with well-being chatbots such as these would serve as a substitute for people engaging with traditional therapy, though future research should continue to evaluate this potential adverse effect.

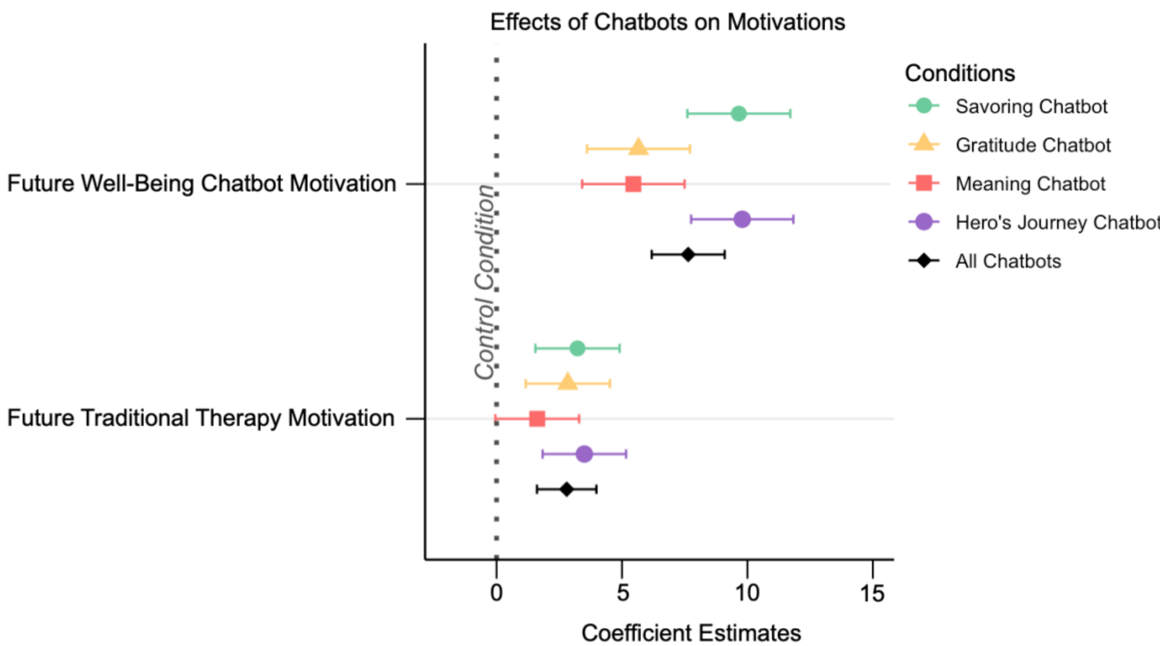


Fig. 6. Unstandardized coefficients from preregistered OLS models estimating effects of savoring, gratitude, meaning, and hero's journey chatbot dialogues, and the average for all chatbot dialogues, on reported willingness to engage with a well-being chatbot in the future and and to go to a therapist, relative to the control chatbot. Points indicate treatment effects on each outcome. Error bars represent 95% confidence intervals. Unstandardized coefficients correspond to estimated percentage point change on outcomes.

DISCUSSION

We found that engaging with well-being chatbots that were designed to employ interventions based on prior psychological research significantly improved affective well-being, meaning in life, life satisfaction, and reduced state anxiety and depression. These effects generalized to individuals with lower baseline well-being and elevated anxiety or depressed mood. Finally, we found that such interventions are already in widespread use, and that engaging with them further increased participants' willingness to seek both chatbot-based and traditional therapeutic support.

Across all conditions, participants reported greater emotional access and a stronger sense of being seen and heard, two factors previously identified as central to psychological well-being [41]. Beyond these shared effects, the gratitude chatbot elicited stronger feelings of gratitude, whereas the hero's journey chatbot was especially effective in fostering self-knowledge. The content analysis of participants' language use in the chatbot dialogues indicated that participants interacting with the savoring and gratitude chatbots used more language reflecting positive emotion, while those interacting with the meaning and hero's journey chatbots used more language reflecting cognitive insight or social connection. Importantly, the interventions tested here likely represent a subset of the psychological strategies that could be deployed by AI chatbots, underscoring the potential for further innovation.

Access to professional mental health support remains a challenge [67]. As a result, growing numbers are turning to generative AI-powered alternatives to human therapy [2,3]. In a nationally representative U.S. sample, we found that 51% of Americans were willing to try validated well-being chatbots, and 28% had already used generative AI for emotional support. The latter indicates a striking proportion of Americans have already turned to AI for mental health support, with more certain to follow in the years to come. This widespread engagement makes evaluation work in this space critical.

The use of four distinct chatbots, grounded in prior research, within a single preregistered experiment using a consistent design and shared outcome measures helps to address a key gap in the well-being literature [15]. Our study helps set realistic expectations for who is likely to benefit, on what aspects of psychological well-being, and how much. Evaluation can also inform engagement. For instance, among the well-being chatbots designed and evaluated here, individuals seeking greater depth of self-insight, and who are comfortable reflecting on prior adversity, may benefit most by engaging with the hero's journey bot. Those specifically seeking a more positive mood, however, might benefit more from the savoring chatbot. As evidence accumulates, AI chatbot found to offer unique effects may offer people options for dialogues they can engage with that fit their emotional needs, enabling more personalized and effective support.

Limitations and Future Research

One limitation is that our interventions focused on short-term subjective well-being outcomes. The chatbots were designed to enhance momentary emotional states, such as affective well-being, rather than to produce lasting change via a single interaction. Given the small to moderate improvements observed, we determined the likelihood of durable effects weeks later was too low to justify additional waves of data collection [68]. Future work, however, should identify ways to affect durable improvements in psychological well-being.

Interventions that target more stable cognitions - such as beliefs about the self, relationships, or society - may be more likely to produce long-term change. Additionally, chatbots that successfully guide people to behave differently in the world, ideally helping people to enact existing motivations and/or forming new and healthier habits, are also promising and should be carefully studied.

A related limitation is the one-time nature of the intervention. Just as traditional therapy typically requires multiple sessions to produce lasting effects [69,70], it is unlikely that a single, short chatbot dialogue would yield enduring benefits. Future research should explore whether repeated or sustained engagement with well-being chatbots can sustain the immediate gains we observed here. Encouragingly, we found interacting with a well-being chatbot increased participants' interest in using such tools again and in seeking support from a human therapist. This suggests that well-being chatbots could foster durable improvements not only through their immediate effects, but also by increasing users' motivation to engage with other well-being-enhancing resources.

The chatbots in this study were designed to implement specific, empirically validated strategies from the empirical psychology literature using a predetermined conversational format guided by carefully constructed prompts. This structure limited the chatbot's freedom in dialogues, ensuring that observed effects could be attributed to the strategy each chatbot was designed to employ. However, this controlled setup also limits generalizability of our findings. Our results speak only to the effects of AI dialogues as we structured them, and within the U.S. population. They can not speak to the use of AI chatbots for mental health support in an informal or ad hoc way, the evaluation of which requires direct study. Indeed, the informal use of AI chatbots for mental health purposes raises significant concerns. Anecdotal reports show unstructured engagement with generative AI chatbots can distort users’ perceptions of reality, sometimes drawing them into conspiratorial or mystical belief systems [71]. Moreover, recent research has shown that commercially available LLMs can produce responses that are inaccurate, unsafe, or inconsistent with clinical guidelines [72], underscoring the need for expert oversight, transparent evaluation standards, and rigorous evaluation.

Conclusion

In sum, our findings demonstrate that generative AI chatbots grounded in evidence-based interventions can meaningfully improve psychological well-being across a range of outcomes. These tools not only show broad effectiveness and specificity in their mechanisms, but also hold strong promise for real-world adoption, with many already using AI chatbots in an informal way for mental health support. As generative AI continues to advance, AI well-being interventions could provide a scalable and accessible complement to traditional mental health resources.

MAIN REFERENCES

1. Demszky, D. *et al.* Using large language models in psychology. *Nat. Rev. Psychol.* **2**, 688–701 (2023).
2. Stade, E. C. *et al.* Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *Npj Ment. Health Res.* **3**, 1–12 (2024).
3. Zao-Sanders, M. How People Are Really Using Gen AI in 2025. *Harv. Bus. Rev.* (2025).
4. Beatty, C., Malik, T., Meheli, S. & Sinha, C. Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study. *Front. Digit. Health* **4**, (2022).
5. Heinz, M. V. *et al.* Randomized Trial of a Generative AI Chatbot for Mental Health Treatment. *NEJM AI* **2**, AIoa2400802 (2025).
6. Leo, A. J. *et al.* Digital Mental Health Intervention Plus Usual Care Compared With Usual Care Only and Usual Care Plus In-Person Psychological Counseling for Orthopedic Patients With Symptoms of Depression or Anxiety: Cohort Study. *JMIR Form. Res.* **6**, e36203 (2022).
7. Hua, Y. *et al.* A scoping review of large language models for generative tasks in mental health care. *Npj Digit. Med.* **8**, 1–10 (2025).
8. Ong, J. C. L. *et al.* Medical Ethics of Large Language Models in Medicine. *NEJM AI* **1**, AIra2400038 (2024).
9. Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R. & Yang, D. The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being. Preprint at

https://doi.org/10.48550/arXiv.2506.12605 (2025).

10. Diener, E. Subjective well-being. *Psychol. Bull.* **95**, 542–575 (1984).
11. van Agteren, J. *et al.* A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nat. Hum. Behav.* **5**, 631–652 (2021).
12. Henry, M. *et al.* The Meaning-Making intervention (MMi) appears to increase meaning in life in advanced ovarian cancer: a randomized controlled pilot study. *Psychooncology.* **19**, 1340–1347 (2010).
13. Rogers, B. A. *et al.* Seeing your life story as a Hero's Journey increases meaning in life. *J. Pers. Soc. Psychol.* **125**, 752–778 (2023).
14. Carr, A. *et al.* Effectiveness of positive psychology interventions: a systematic review and meta-analysis. *J. Posit. Psychol.* **16**, 749–769 (2021).
15. Folk, D. & Dunn, E. A systematic review of the strength of evidence for the most commonly recommended happiness strategies in mainstream media. *Nat. Hum. Behav.* **7**, 1697–1707 (2023).
16. Weiss, L. A., Westerhof, G. J. & Bohlmeijer, E. T. Can We Increase Psychological Well-Being? The Effects of Interventions on Psychological Well-Being: A Meta-Analysis of Randomized Controlled Trials. *PLOS ONE* **11**, e0158092 (2016).
17. Moore, J. *et al.* Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. in *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* 599–627 (Association for Computing Machinery, New York, NY, USA, 2025). doi:10.1145/3715275.3732039.
18. Henderson, L. & Knight, T. Integrating the hedonic and eudaimonic perspectives to more comprehensively understand wellbeing and pathways to wellbeing. (2012)

doi:10.5502/ijw.v2i3.3 %5D.

19. Disabato, D. J., Goodman, F. R., Kashdan, T. B., Short, J. L. & Jarden, A. Different types of well-being? A cross-cultural examination of hedonic and eudaimonic well-being. *Psychol. Assess.* **28**, 471–482 (2016).
20. Kapteyn, A., Lee, J., Tassot, C., Vonkova, H. & Zamarro, G. Dimensions of Subjective Well-Being. *Soc. Indic. Res.* **123**, 625–660 (2015).
21. DeNeve, K. M. & Cooper, H. The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychol. Bull.* **124**, 197–229 (1998).
22. Diener, E., Suh, E. M., Lucas, R. E. & Smith, H. L. Subjective well-being: Three decades of progress. *Psychol. Bull.* **125**, 276–302 (1999).
23. Lyubomirsky, S. & Layous, K. Well-being. (2025).
24. Bolier, L. *et al.* Positive psychology interventions: a meta-analysis of randomized controlled studies. *BMC Public Health* **13**, 119 (2013).
25. Sin, N. L. & Lyubomirsky, S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. *J. Clin. Psychol.* **65**, 467–487 (2009).
26. van Agteren, J. *et al.* A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nat. Hum. Behav.* **5**, 631–652 (2021).
27. Lyubomirsky, S., Sousa, L. & Dickerhoof, R. The costs and benefits of writing, talking, and thinking about life's triumphs and defeats. *J. Pers. Soc. Psychol.* **90**, 692–708 (2006).
28. Koo, M., Algoe, S. B., Wilson, T. D. & Gilbert, D. T. It's a wonderful life: Mentally subtracting positive events improves people's affective states, contrary to their affective

forecasts. *J. Pers. Soc. Psychol.* **95**, 1217–1224 (2008).

29. Kurtz, J. L. Looking to the Future to Appreciate the Present: The Benefits of Perceived Temporal Scarcity. *Psychol. Sci.* **19**, 1238–1241 (2008).
30. Wood, A. M., Linley, P. A., Maltby, J., Kashdan, T. B. & Hurling, R. Using personal and psychological strengths leads to increases in well-being over time: A longitudinal study and the development of the strengths use questionnaire. *Personal. Individ. Differ.* **50**, 15–19 (2011).
31. Heekerens, J. B. & Eid, M. Inducing positive affect and positive future expectations using the best-possible-self intervention: A systematic review and meta-analysis. *J. Posit. Psychol.* **16**, 322–347 (2021).
32. Quoidbach, J., Wood ,Alex M. & and Hansenne, M. Back to the future: the effect of daily practice of mental time travel into the future on happiness and anxiety. *J. Posit. Psychol.* **4**, 349–355 (2009).
33. Sedlmeier, P. *et al.* The psychological effects of meditation: A meta-analysis. *Psychol. Bull.* **138**, 1139–1171 (2012).
34. Capaldi, C. A., Passmore, H.-A., Nisbet, E. K., Zelenski, J. M. & Dopko, R. L. Flourishing in nature: A review of the benefits of connecting with nature and its application as a wellbeing intervention. *Int. J. Wellbeing* **5**, (2015).
35. Regan, A., Radošić, N. & Lyubomirsky, S. Experimental effects of social behavior on well-being. *Trends Cogn. Sci.* **26**, 987–998 (2022).
36. Walton, G. M., Cohen, G. L., Cwir, D. & Spencer, S. J. Mere belonging: The power of social connections. *J. Pers. Soc. Psychol.* **102**, 513–532 (2012).
37. Fava, G. A. & Ruini, C. Development and characteristics of a well-being enhancing

psychotherapeutic strategy: well-being therapy. *J. Behav. Ther. Exp. Psychiatry* **34**, 45–63 (2003).

38. McClain, O. S. and C. 34% of U.S. adults have used ChatGPT, about double the share in 2023. *Pew Research Center* https://www.pewresearch.org/short-reads/2025/06/25/34-of-us-adults-have-used-chatgpt-about-double-the-share-in-2023/ (2025).
39. Morrone, M. Teens flock to companion bots despite risks. *Axios* https://www.axios.com/2025/07/16/ai-bot-companions-teens-common-sense-media (2025).
40. Rubin, M. *et al.* Comparing the value of perceived human versus AI-generated empathy. *Nat. Hum. Behav.* 1–15 (2025) doi:10.1038/s41562-025-02247-w.
41. Yin, Y., Jia, N. & Wakslak, C. J. AI can help people feel heard, but an AI label diminishes this impact. *Proc. Natl. Acad. Sci.* **121**, e2319112121 (2024).
42. Heffner, J. *et al.* Increasing happiness through conversations with artificial intelligence. Preprint at https://doi.org/10.48550/arXiv.2504.02091 (2025).
43. Ovsyannikova, D., de Mello, V. O. & Inzlicht, M. Third-party evaluators perceive AI as more compassionate than expert humans. *Commun. Psychol.* **3**, 4 (2025).
44. McBain, R. K. *et al.* Competency of Large Language Models in Evaluating Appropriate Responses to Suicidal Ideation: Comparative Study. *J. Med. Internet Res.* **27**, e67891 (2025).
45. Oğuz-Uğuralp, Z., Uğuralp, A. K., Puntoni, S. & De Freitas, J. Why Most Resist AI Companions. SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.5097445 (2025).
46. Rubin, M. *et al.* Comparing the value of perceived human versus AI-generated empathy. *Nat. Hum. Behav.* 1–15 (2025) doi:10.1038/s41562-025-02247-w.

47. Cuadra, A. *et al.* The Illusion of Empathy? Notes on Displays of Emotion in Human-Computer Interaction. in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* 1–18 (Association for Computing Machinery, New York, NY, USA, 2024). doi:10.1145/3613904.3642336.

48. Baumel, A., Muench, F., Edan, S. & Kane, J. M. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. *J. Med. Internet Res.* **21**, e14567 (2019).

49. Landry, A. P. & Halperin, E. Intergroup psychological interventions: The motivational challenge. *Am. Psychol.* **80**, 206–219 (2025).

50. Torous, J., Nicholas, J., Larsen, M. E., Firth, J. & Christensen, H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid. Based Ment. Health* **21**, 116–119 (2018).

51. Inzlicht, M., Cameron, C. D., D'Cruz, J. & Bloom, P. In praise of empathic AI. *Trends Cogn. Sci.* **28**, 89–91 (2024).

52. Fang, C. M. *et al.* How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Randomized Controlled Study. Preprint at https://doi.org/10.48550/arXiv.2503.17473 (2025).

53. Bryant, J., Carveth, R. A. & Brown, D. Television Viewing and Anxiety: An Experimental Examination. *J. Commun.* **31**, 106–119 (1981).

54. Hurley, D. B. & Kwon, P. Savoring Helps Most When You Have Little: Interaction Between Savoring the Moment and Uplifts on Positive Affect and Satisfaction with Life. *J. Happiness Stud.* **14**, 1261–1271 (2013).

55. Algoe, S. B. Find, Remind, and Bind: The Functions of Gratitude in Everyday

Relationships. *Soc. Personal. Psychol. Compass* **6**, 455–469 (2012).

56. Steger, M. F., Kashdan, T. B. & Oishi, S. Being good by doing good: Daily eudaimonic activity and well-being. *J. Res. Personal.* **42**, 22–42 (2008).
57. Campbell, J. *The Hero with a Thousand Faces*. vol. 17 (New World Library, 1949).
58. Pennebaker, J. W. & Beall, S. K. Confronting a traumatic event: Toward an understanding of inhibition and disease. *J. Abnorm. Psychol.* **95**, 274–281 (1986).
59. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–1070 (1988).
60. Costin, V. & Vignoles, V. L. Meaning is about mattering: Evaluating coherence, purpose, and existential mattering as precursors of meaning in life judgments. *J. Pers. Soc. Psychol.* **118**, 864–884 (2020).
61. Diener, E., Emmons, R., Larsen, R. & Griffin, S. The Satisfaction With Life Scale. *J. Pers. Assess.* **49**, 71–5 (1985).
62. Spielberger, C. D., Gonzalez-Reigosa, F., Martinez-Urrutia, A., Natalicio, L. F. S. & Natalicio, D. S. The State-Trait Anxiety Inventory. *Rev. Interam. Psicol. J. Psychol.* **5**, (2017).
63. Lovibond, P. F. & Lovibond, S. H. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav. Res. Ther.* **33**, 335–343 (1995).
64. Schwartz, H. A. *et al.* DLATK: Differential Language Analysis ToolKit. in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (eds Specia, L., Post, M. & Paul, M.) 55–60 (Association for

Computational Linguistics, Copenhagen, Denmark, 2017). doi:10.18653/v1/D17-2010.

65. Eichstaedt, J. C. *et al.* Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychol. Methods* **26**, 398–427 (2021).

66. Schwartz, H. A. *et al.* Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLOS ONE* **8**, e73791 (2013).

67. Vigo, D. V. *et al.* Effective Treatment for Mental and Substance Use Disorders in 21 Countries. *JAMA Psychiatry* **82**, 347–357 (2025).

68. He, Y. *et al.* Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials. *J. Med. Internet Res.* **25**, e43862 (2023).

69. Bruijniks, S. J. E. *et al.* The effects of once- versus twice-weekly sessions on psychotherapy outcomes in depressed patients. *Br. J. Psychiatry* **216**, 222–230 (2020).

70. Lambert, M. J. Effectiveness of Psychotherapeutic Treatment. *Reson. – E-J. Für Biopsychosoziale Dialoge Psychosom. Med. Psychother. Superv. Berat.* **3**, 87–100 (2015).

71. Hill, K. They Asked an A.I. Chatbot Questions. The Answers Sent Them Spiraling. *The New York Times* (2025).

72. Modi, N. D. *et al.* Assessing the System-Instruction Vulnerabilities of Large Language Models to Malicious Conversion Into Health Disinformation Chatbots. *Ann. Intern. Med.* (2025) doi:10.7326/ANNALS-24-03933.

METHODS AND MATERIALS

We preregistered the study's design, outcome measures, hypotheses, and primary analyses regarding the potential effects of chatbot interactions on affective well-being (positive relative to negative emotions), meaning in life, and life satisfaction. We preregistered use of one-tailed tests, applying a significance threshold of $p < .05$. Full preregistration is available at https://osf.io/w4efh. Beyond these preregistered confirmatory analyses, additional analyses should be viewed as exploratory, though wherever possible we minimize researcher degrees of freedom by employing the same preregistered analytic strategy, though using two-tailed tests to reflect their exploratory nature.

Research materials, language data summaries and survey data, and analysis code are available on the Open Science Framework at https://osf.io/g59fm/. All studies were approved by the Institutional Review Board at Stanford University (Protocol ID: 75275). Study 1 was pre-registered, whereas the survey conducted in Study 2 was not pre-registered prior to data collection.

Experimental Chatbot Study

Participants. Based on our power analysis, our pre-registered sample size was 3,000 (see SI Appendix, Section 0.2 for full details) with 500 participants per treatment condition and 1,000 participants for the control condition, adjusting multiple comparisons [73]. We ultimately recruited 3,209 participants with balanced gender representation through the participant platform provider Bovitz. After excluding participants who failed both attention checks, the final sample included 2,936 individuals (1,468 female, 1463 male, 5 others; age:

M = 37.97, *SD* = 13.02, range = 18-83, non-white = 31.6%, with a college degree = 55.40, median income = $60k - $69k). Participants were compensated $3 for completing the 20-minute task.

Procedures. First, participants were shown an information sheet describing the study and proceeded only if they consented to participate. The study consisted of three parts and was conducted in multiple batches of no more than 50 participants at a time to ensure that our chatbot platform could manage all conversations without delays.

The first part of the study consisted of a baseline survey, in which participants provided demographic information and completed a personality questionnaire, as well as pre-intervention assessments of affective well-being, meaning in life, prior experience with therapy, and motivation to consider therapy or try well-being chatbots, using shortened versions of the scales. This survey took an average of 8.81 minutes (*SD* = 8.95 minutes).

The second part of the study involved the chatbot interaction. Participants were informed that the conversation would last approximately 8–10 minutes, though they were only required to remain in the conversation for a minimum of 3 minutes before being allowed to proceed. While there was no minimum word requirement, participants who contributed fewer than 50 words received a pop-up message noting that they had written less than 50 words and encouraging them to add more. However, they were free to continue to the next stage regardless of their response length. Each chatbot followed a pre-registered script tailored to its intervention type (see SI Appendix, Section 0.5 for full prompts). On average, these interactions included 14.66 conversational turns from the user and lasted 9.57 minutes (*SD* = 7.09 minutes, see SI Appendix Section 3.1 for a condition-specific breakdown). In the control condition, the chatbot asked neutral questions about the layout and contents of their bedroom [58]. As in the intervention conditions, the control chatbot provided a summary and downloadable PDF at the end of the interaction.

Finally, participants completed the post-intervention survey. This survey included the full versions of the measures assessing affective well-being, meaning in life, prior experience with therapy, and motivation to engage with well-being chatbots. Additionally, we included scales to assess potential mediating factors that might influence the effectiveness of the interventions, such as whether participants felt heard or gained new insights about themselves. Participants were also asked again about their willingness to consider therapy or use similar chatbots in the future, along with questions evaluating how pleasant and how difficult they found the chatbot experience to be. The completion time of the post-survey was 7.84 minutes (SD = 8.95 minutes).

Measures. We assessed five well-being outcomes: affective well-being, meaning in life, life satisfaction, depression, and anxiety. All measures were administered before and after the intervention to capture change over time. Affective well-being was measured using the Scale of Positive and Negative Experience (SPANE [74]). Participants rated the extent to which they were experiencing six positive affective states (e.g., "pleasant," "joyful") and six negative states (e.g., "anxious," "sad") right now. Items were rated on a 101-point scale ranging from 0 ("Not at all") to 100 ("Extremely"). A composite score was computed by subtracting the average of the negative items from the average of the positive items. Meaning in life was measured using the composite score of the meaning in life scale [60]. Life satisfaction was assessed using the Satisfaction With Life Scale [61]. To assess state depression (i.e., how depressed participants felt in the moment), we adapted four items from the Depression, Anxiety, and Stress Scale (DASS-21 [63]). Rather than reporting symptoms over the past week, participants were asked how they felt "right now" (e.g., *"I felt that I had nothing to look forward to"*; see SI Appendix Section 3.6 for full item list). State anxiety was measured using items from the State-Trait Anxiety Inventory (STAI [62]; e.g., *"I feel nervous"*). Consistent with effects on depression, all well-being chatbots significantly reduced reported state anxiety. For meaning in life, life satisfaction, depression, and anxiety, items were rated on a 101-point scale ranging from 0 ("Strongly disagree") to 100 ("Strongly agree"), with the prompt: "Please indicate the extent to which you agree or disagree with the following statements RIGHT NOW."

As additional measures of outcomes that might be affected by the chatbot interaction, which potentially mediate the effect, we assessed seven different outcomes: Having positive thoughts during the interaction, having deeper access to their own emotions, how grateful they felt during the intervention, how much they learned about themselves (self-knowledge), and how much they felt seen and heard. These were assessed using 101-point scales using questions like: How much do you agree with the following statements? During the exercise I delved deeper into my emotions than I usually do" from not at all to extremely.

We also examined participants' prior experience with and motivation toward therapy and well-being chatbots, as well as whether these motivations changed following the chatbot interaction. To do this, participants were asked whether they had previously engaged in therapy, how likely they were to seek therapy in the future, and how motivated they were to try well-being chatbots. Motivation to try chatbots was assessed using a three-item composite, including questions such as, "How likely are you to try out such chatbots?" This allowed us to explore both potential moderation by prior therapy experience and shifts in motivation resulting from the intervention.

As covariates, we included key demographic variables such as age and gender. For exploratory heterogeneity analyses, we additionally assessed political affiliation, education level, ethnicity, socioeconomic status, and employment status. To account for individual personality differences, we administered a shortened version of the Big Five personality questionnaire [75].

Representative Survey Study

Participants. We recruited 3,061 participants in the United States through the platform provider Bovitz in three waves, aiming for a nationally representative sample (see SI Appendix, Section 3.3 for how closely the realized sample matched the quota targets). The first wave was collected from May 28th to June 6th, the second from July 7th to 23rd, and the last from August 4th to 19th. Participants who failed more than two attention checks were excluded, resulting in a final sample of 3,056 individuals (1556 female, 1465 male, other = 35; age: M = 48.20, SD = 17.22, range = 18–91, non-white = 35.57%, with a college degree = 30.1%). The survey was part of a larger project and had a median completion time of 17 minutes. Participants received $2 as compensation for their time.

Procedure. After providing informed consent, they completed demographic questions including age, gender, education, and race. Toward the end of the survey, participants were asked about their willingness to use AI well-being chatbots and whether they had previously used such tools.

Measures. Participants indicated their willingness to try an AI-based well-being chatbot using a 101-point slider scale (0 = Definitely would not, 100 = Definitely would). The question described a hypothetical chatbot shown in research to improve happiness and meaning in life. To assess prior use, participants also responded to a yes/no item asking whether they had ever used an AI chatbot (e.g., ChatGPT, Headspace AI) for emotional well-being or mental health.

Analysis Approach

Our preregistered models testing our hypotheses regarding the effects of well-being chatbots on our primary measures of psychological well-being include several covariates: pre-treatment measures of each outcome variable, pre-treatment levels of affective well-being, age, gender, and the Big Five personality traits. In exploratory analyses examining additional dependent variables (e.g., anxiety, depressed mood, mediators such as feeling seen and heard, and motivation to engage with chatbots or traditional therapy), potentially effected by the interactions with any of the well-being chatbots (relative to the control chatbot), we applied the same analytic approach but assessed statistical significance using two-tailed tests. Note that all observed treatment effects were robust to the removal of control variables for pre-treament levels of affective well-being, age, gender, and the Big Five personality traits. See SI Appendix, Section 2.2 for tests of assumptions for statistical tests)

ACKNOWLEDGMENTS

We are grateful for the valuable feedback on the development of the chatbots and the manuscript provided by Joao Sedoc (New York University), Lyle Ungar (University of Pennsylvania) and the World Happiness Project.

We also acknowledge the use of generative AI in the preparation, execution, and revision of this project. Specifically, we used ChatGPT versions 3.5 and 4o to refine the chatbot instructions and to operate the chatbots during the intervention. Additionally, GPT-4o was used to assist in the preparation of visualizations (using ggplot2 in R) and to support the revision and editing of the manuscript language.

Author Contributions: J.P.S.: designed research, performed research, analyzed data, contributed analytic tools, wrote the paper. A.S.: performed research, contributed analytic tools. S.L.: wrote the paper. J.C.E: designed research, performed research, contributed new analytic tools, analyzed data, and wrote the paper. R.W.: designed research, performed research, contributed new analytic tools, analyzed data, and wrote the paper.

Competing Interest Statement: The authors had no conflict of interest to declare.

FUNDING

This research was supported by Stanford's Human-Centered Artificial Intelligence (HAI) Seed Grant (Grant No. 347207-01).

METHODS AND MATERIAL REFERENCES

1. Dunnett, C. W. A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *J. Am. Stat. Assoc.* **50**, 1096–1121 (1955).
2. Pennebaker, J. W. & Beall, S. K. Confronting a traumatic event: Toward an understanding of inhibition and disease. *J. Abnorm. Psychol.* **95**, 274–281 (1986).
3. Diener, E. *et al.* New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. *Soc. Indic. Res.* **97**, 143–156 (2010).
4. Costin, V. & Vignoles, V. L. Meaning is about mattering: Evaluating coherence, purpose, and existential mattering as precursors of meaning in life judgments. *J. Pers. Soc. Psychol.* **118**, 864–884 (2020).
5. Diener, E., Emmons, R., Larsen, R. & Griffin, S. The Satisfaction With Life Scale. *J. Pers. Assess.* **49**, 71–5 (1985).
6. Lovibond, P. F. & Lovibond, S. H. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav. Res. Ther.* **33**, 335–343 (1995).
7. Spielberger, C. D., Gonzalez-Reigosa, F., Martinez-Urrutia, A., Natalicio, L. F. S. & Natalicio, D. S. The State-Trait Anxiety Inventory. *Rev. Interam. Psicol. J. Psychol.* **5**, (2017).
8. Gosling, S. D., Rentfrow, P. J. & Swann, W. B. A very brief measure of the Big-Five personality domains. *J. Res. Personal.* **37**, 504–528 (2003).