From Chatbots to Confidants: A Cross-Cultural Study of LLM Adoption for Emotional Support

Anonymous ACL submission

Abstract

Large Language Models (LLMs) are increasingly used not only for instrumental tasks, but as always-available and non-judgmental confidants for emotional support. Yet what drives adoption and how users perceive emotional support interactions across countries remains unknown. To address this gap, we present the first large-scale cross-cultural study of LLM use for emotional support, surveying 4,641 participants across seven countries (USA, UK, Germany, France, Spain, Italy, and The Netherlands). Our results show that adoption rates vary dramatically across countries (from 20% to 59%). Using mixed models that separate cultural effects from demographic composition, we find that: Being aged 25-44, religious, married, and of higher socioeconomic status are predictors of positive perceptions (trust, usage, perceived benefits), with socioeconomic status being the strongest. English-speaking countries consistently show more positive perceptions than Continental European countries. We further collect a corpus of 731 real multilingual prompts from user interactions, showing that users mainly seek help for loneliness, stress, relationship conflicts, and mental health struggles. Our findings reveal that LLM emotional support use is shaped by a complex sociotechnical landscape and call for a broader research agenda examining how these systems can be developed, deployed, and governed to ensure safe and informed access.

1 Introduction

Large Language Models (LLMs) are a jack of all trades, performing diverse tasks such as question answering and decision support (Chatterji et al., 2025). Beyond these instrumental uses, researchers and designers are increasingly interested in AI systems as potential companions or sources of interpersonal support (e.g. Andersson, 2025; Wu et al., 2025; Savoldi et al., 2025). In particular, conversational interfaces make LLMs accessible for emotionally oriented interactions such as seeking reassurance, discussing personal concerns, or reflecting on experiences[1].

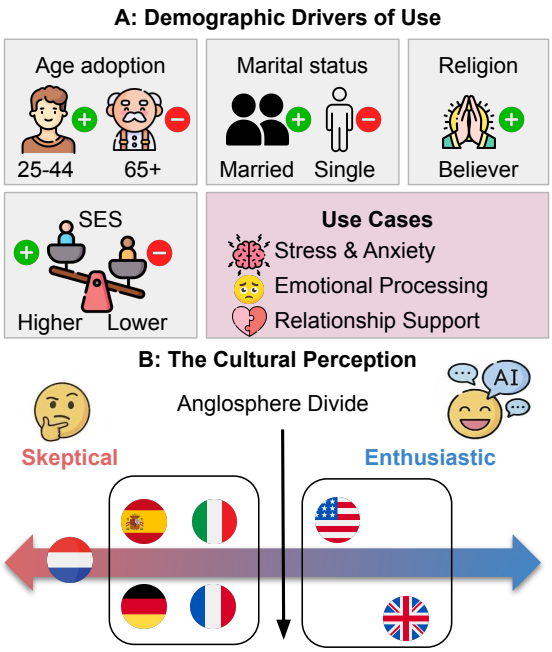


Figure 1: Global and Demographic Landscape in LLM Adoption and Perceptions for Emotional Support (N=4,641). SES: Socio-economic status.

At the same time, many individuals face barriers to accessing traditional mental health support, including cost, stigma, and limited availability of services (Hou et al., 2025; Tyson et al., 2023). As a result, digital technologies are increasingly explored as complementary sources of emotional support (Chaudhry and Debi, 2024; Pandya et al., 2024). Conversational AI systems may lower barriers to seeking help by offering immediate, and always-available interaction (Haensch, 2025; Pandya et al., 2024). Yet, the growing use of LLMs for emotionally sensitive interactions raises questions about

[1] https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025

safety, trust, and appropriate system design (Kaffee et al., 2025).

While these ethical and design implications have received significant attention, empirical evidence remains scarce on how commonly the general public turns to LLMs for emotional support and what factors shape these interactions. In particular, we lack population-level evidence on (1) who adopts LLMs for emotional support, (2) how sociodemographic and cultural backgrounds shape their perceptions of these systems, and (3) what users actually disclose to these systems.

To address this gap, we conduct a survey examining how members of the public report using LLMs for emotional and mental health–related support. We examine whether cultural and sociodemographic factors mediate the use of LLMs for emotional support (Figure 1). Specifically, we address three pivotal research questions (RQs):

(RQ1) Who uses LLMs for emotional support, and what are the main purposes and benefits that motivate people to use them?

(RQ2) How do demographics and cultural background shape user perceptions of these systems?

(RQ3) What do users actually disclose when seeking such support?

Contributions 1) We present the first cross-cultural empirical study of GenAI chatbot use for emotional support, surveying 4,641 participants across seven countries. 2) We provide an extensive sociodemographic and cultural analysis using Cumulative Link Mixed Models across three constructs (Trust and perceived privacy, Perceived benefits, and Usage intention) to disentangle demographic from cultural effects. 3) We collect and analyze 731 real user prompts across six languages, identifying key emotional support topics.

Together, our findings provide the first population-level evidence of how cultural and sociodemographic factors shape human–AI emotional support interactions, highlighting a complex sociotechnical landscape that motivates a broader research agenda to explore how these systems can be developed, deployed, and governed to promote use that is safe and informed.

We will release our survey data upon publication to support future research in this area.

2 Related Work

Interest in public perceptions of conversational AI for wellbeing has grown considerably, prompting surveys, usability studies, and social media analyses, predominantly conducted in English (Haensch, 2025; Chaudhry and Debi, 2024; Cameron et al., 2018). Findings suggest that users value the availability, non-judgmental nature, and accessibility of AI conversational agents, while pointing to shallow and context-insensitive responses as key limitations (Chaudhry and Debi, 2024; Haensch, 2025). Younger adults and those with higher education and AI familiarity tend to show relatively more openness toward AI tools (Tyson et al., 2023). These attitudinal patterns are consistent with the CASA paradigm, which predicts that people apply the same social norms to AI as to humans (Nass and Moon, 2000) and disclose more under conditions of reduced social presence (Papneja and Yadav, 2025). The psychological barriers underlying AI resistance, including perceived opacity, emotionlessness, and rigidity, further shape whether individuals are willing to engage with these systems (De Freitas et al., 2023). Beyond wellbeing, broader surveys of general LLM use show that SES and demographic factors stratify adoption rates and interaction styles, with higher-SES users interacting more abstractly (Bassignana et al., 2025) and younger, more educated users showing higher adoption and literacy (Savoldi et al., 2025).

Yet none of these works study the adoption and user perceptions of LLMs specifically for emotional wellbeing support, and all remain confined to a single country and predominantly Anglophone populations, leaving open whether adoption and perception patterns vary across cultural and sociodemographic contexts. We address this gap by surveying 4,641 participants across seven countries in six languages, collecting prompts from real prior interactions, and providing the first cross-cultural analysis of LLM adoption for emotional support that explicitly separates demographic from cultural effects.

3 Survey Methodology

3.1 Study Design

We conduct a cross-cultural survey to examine demographic and cultural variation in the adoption and perception of LLMs for emotional support. We select seven countries representing the highest global shares of ChatGPT[2] visitors (FirstPageSage, 2026): the United States (17.1%), France (4.3%), Spain (3.7%), the United Kingdom (2.7%), Italy

[2]Currently the most used conversational LLM.

(2.5%), Germany (2.4%), and the Netherlands (1.1%), ensuring our sample reflects the populations most actively engaging with these systems. We collected the data via a Qualtrics panel,[3] using quota sampling to ensure even distribution across countries (N=200 per country) and gender balance within each national cohort.

Before starting the survey, we presented participants with an onboarding screen to obtain informed consent. To establish a shared baseline understanding, we provided examples of "AI agents" (e.g., ChatGPT, Gemini, Claude) and defined "emotional support and mental wellbeing" as using these tools for managing stress, processing emotions, or personal reflection. We informed participants of the survey's anonymous, voluntary nature, the estimated duration, and their right to withdraw at any time. The survey then proceeded in four steps. First, participants completed a demographic questionnaire. Second, a filter question classified them as *users* (having used AI chatbots for emotional support) or *non-users*, enabling demographic comparison between adopters and non-adopters.Third, identified users reported their usage patterns (§3.2) and evaluated their perceptions across several constructs. Finally, the survey concluded with an optional open-ended text where users could share their most recent support-seeking prompts. To maintain data privacy, we instructed participants to exclude any names, personal details, or identifiable information before submitting their text.

Pilot test 20 experts in survey methodology, NLP, and HCI from our target countries reviewed the survey to refine it. The original English survey was automatically translated into Spanish, French, Italian, German, and Dutch via Qualtrics and subsequently reviewed by native-speaking experts from the panel to ensure linguistic and cultural accuracy. Pilot feedback informed refinements to question wording; for instance, absolute statements were softened to better capture user perception (e.g., changing "I can express my feelings" to "I feel that I can express my feelings"). The average completion time of 8-9 minutes was used to determine participant compensation on the platform.

Following the pilot, we deployed the survey and collected 5,319 responses. To ensure data integrity, we applied Qualtrics quality filters to exclude suspected bot activity, duplicate submissions, incomplete or straight-line responses. In addition, we excluded participants with completion times faster than 30% of the median, and those who failed either of two embedded attention checks. The final validated sample consisted of 4,641 participants.

[3]Qualtrics provides reliable access to diverse, cross-cultural respondents and robust built-in data quality checks.

3.2 Questionnaire Design

The survey comprises four sections: demographics, AI usage, evaluation of user perception constructs, and prompt collection.

Demographics Participants reported age, gender, nationality, education, religion, and marital status. We assessed subjective socioeconomic status using the MacArthur Scale of Subjective Social Status (Adler et al., 2000), which captures perceived social standing rather than objective income, making it particularly suited for cross-cultural comparison. We handled all data in compliance with GDPR and ensured that participants were fully anonymized. Further details are provided in Ethical Considerations.

Usage Patterns We asked users about their AI usage patterns, including the AI agents used, frequency of use, primary devices, and their specific emotional support use cases.

3.3 Constructs to Capture User Perceptions

We assess three key areas to capture user perceptions of LLMs for emotional support. To create robust, composite response variables for our Cumulative Link Mixed Model (CLMM) analysis, we merged conceptually related questionnaire subscales into three distinct constructs. Each item was measured on a 5-point Likert scale (from *Strongly disagree* to *Strongly agree*). The full questionnaire, including items drawn from existing validated scales and those developed by the authors, is provided in Table 11 in Appendix G. The three resulting constructs are:

- **Trust and Privacy (TRU):** This construct merges the "Trust" and "Privacy" sub-scales to capture the user's overall confidence in the system. It assesses the agent's perceived credibility, honesty, and transparency, alongside the user's perceived safety regarding the handling of sensitive data, interaction confidentiality, and the clear communication of system limitations (Marimon et al., 2024; Schmidmaier et al., 2024).
- **Perceived Benefits (BENF):** This construct evaluates the perceived advantages of using LLMs for emotional support. It captures practi-

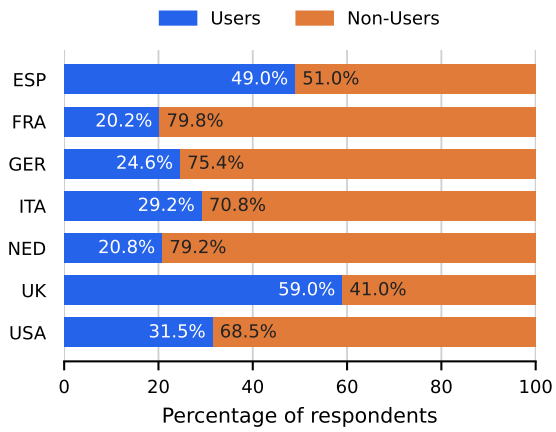


Figure 2: Cross-country comparison of AI chatbot use for emotional support and mental wellbeing.

cal, low-barrier benefits such as 24/7 accessibility and cost-effectiveness, alongside psychological benefits, including the ability to engage in non-judgmental exploration, gain new perspectives, and receive actionable advice.

- **Usage Intention and Behavior (USE):** To measure overall system engagement, we merged the "Attitude Toward Using" and "Actual System Use" sub-scales, which are grounded in the Technology Acceptance Model (TAM) (Davis, 1989). This combined construct captures both the user's favorable evaluation of the LLM as a supportive tool and their self-reported frequency of behavioral engagement.

4 Results

We present findings along our three RQs.

4.1 User and Non-User Profiles

To address *RQ1* we first compare demographic profiles of users against non-users. We then examine the users interaction patterns and underlying motivations.

Adoption Rates by Country Figure 2 shows the distribution of users and non-users per country. We observe noticeable differences between the countries. The only country in which users constitute the majority is the United Kingdom (59.0%). Spain comes close with 49.0% users. France, Germany, and the Netherlands report the lowest percentage of users (approximately 20% of participants). Across the seven countries surveyed, an average of 28.9% of participants are users.

Aggregate Demographic Differences To investigate demographic differences between users and non-users, we conducted chi-square tests. Except for gender ($\chi^2(4) = 3.4$, $p_{\text{adj}} = .495$), all demographic variables show significant differences: age ($\chi^2(6) = 915.3$, $p_{\text{adj}} < .001$), education ($\chi^2(7) = 319.0$, $p_{\text{adj}} < .001$), SES ($\chi^2(9) = 212.9$, $p_{\text{adj}} < .001$), marital status ($\chi^2(5) = 103.3$, $p_{\text{adj}} < .001$), and religion ($\chi^2(8) = 72.5$, $p_{\text{adj}} < .001$). Effect sizes range from small to moderate, with age showing the largest difference between the groups.

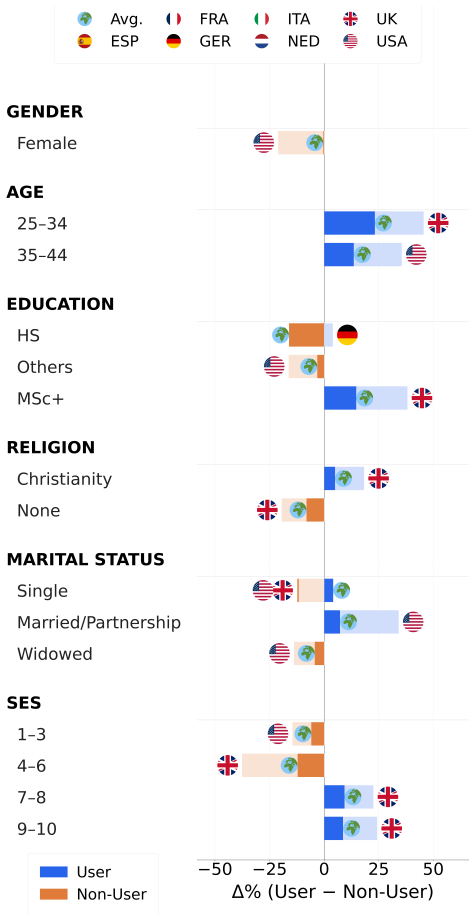


Figure 3: Percentage-point differences in demographic composition between users and non-users. Tinted bars indicate countries deviating >1.5 SD from the global average (marked with the globe emoji).

Demographic Profiles and Country-Level Variation Figure 3 shows the difference in demographic distribution across user groups. Users are predominantly aged 25-44, highly educated (MSc, PhD), report higher socioeconomic status, and are more likely to be married or partnered. Non-users are more likely to be high-school graduates, single or widowed, and non-religious. Notable country-level differences emerge particularly in the UK and the USA. The UK shows the starkest age, education, and SES gaps, with younger and higher-SES

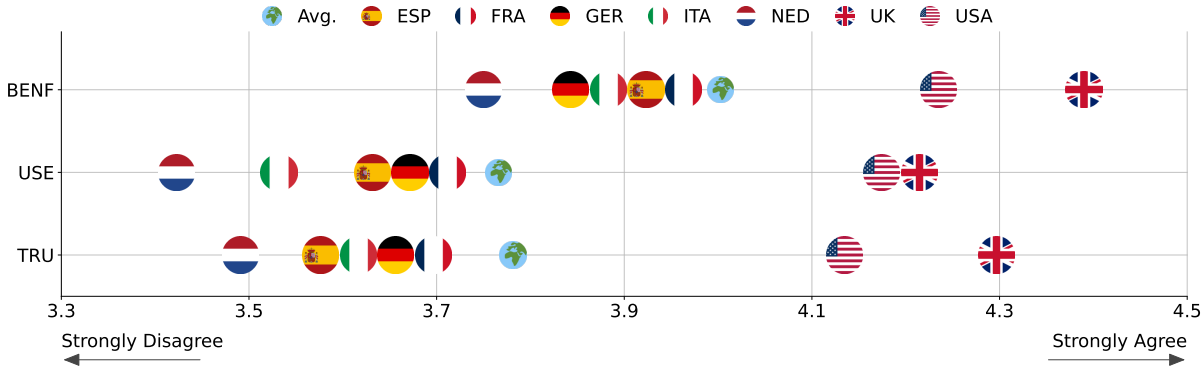


Figure 4: Mean scores per construct: Perceived Benefits (BENF), Usage intention (USE), Trust & Privacy (TRU) across seven countries in a 5-point Likert scale (from Strongly disagree to Strongly agree).

individuals more prevalent among users, who are also more Christian and less non-religious than non-users. In the USA, non-users skew female, are overrepresented in the 35-44 age range, are more likely to be married, and are less likely to have lower SES. Germany is an exception where high-school graduates are more prevalent among users. For full demographic breakdowns, see Table 2 and Table 3 in Appendix B.

Interaction Patterns and Purposes Moving beyond demographics, we examined how users interact with these systems (see Figure 7 in Appendix A). Usage is roughly split between occasional users (34.0%) and frequent, weekly users (31.5%). ChatGPT is the dominant AI agent (79.7%), and smartphones are the preferred interface by a wide margin (88.2%). Users primarily turn to these systems for the following purposes: managing stress and anxiety (49.4%), finding encouragement or motivation (42.5%), processing and reflecting on emotions (41.3%). A notable share also report experimenting out of curiosity (44.9%), suggesting that for many, intentional support-seeking and exploratory use co-occur. Other common uses of personal support include relationship support (30.2%) and companionship to combat loneliness (17.6%).

Perceived Benefits Figure 8 shows that users primarily value accessibility and practical utility. The highest-rated benefits are 24/7 availability (mean score = 4.26) and cost-effectiveness (4.12), indicating that users highly value immediate, low-barrier access to support. Additionally, psychological safety emerged as a critical advantage, with users highly rating the ability to express feelings without fear of human judgment (4.10).

In summary, LLM adoption for emotional support skews toward younger, higher-educated, and higher-SES individuals. Users are primarily motivated by stress management, emotional processing, and encouragement, drawn by the 24/7 availability, cost-effectiveness, and non-judgmental nature of these systems.

4.2 Cross-Country Variations in Perception

To address *RQ2*, we first investigate how these perceptions vary across different cultural backgrounds at the country level.

Figure 4 presents the mean scores per construct across seven countries on a 5-point Likert scale. There is a sharp divide between the US and the UK, and the other countries. The **UK** has the highest mean scores across all three constructs, followed by the **USA**. The Anglosphere countries are above the combined mean, while mainland European countries consistently score lower. Within Europe, the **Netherlands** stands out by scoring lowest across the three constructs. To confirm statistical significance, we apply the Kruskal-Wallis test followed by Dunn's test to identify pairwise differences. The UK and the USA stand apart from all other countries across all constructs, and differ only on Perceived Benefits (BENF). Mainland European countries are largely homogeneous: Germany, France, and Spain show no significant differences within themselves. Italy also does not differ from Spain and Germany. The Netherlands stands out among European countries: it diverges from France on all constructs, and from Germany and Spain on two constructs each. For more details, see Appendix D.

4.3 Mixed Model Analysis of Culture and Demographics

Our findings in §4.2 show that perceptions vary significantly across countries. However, according

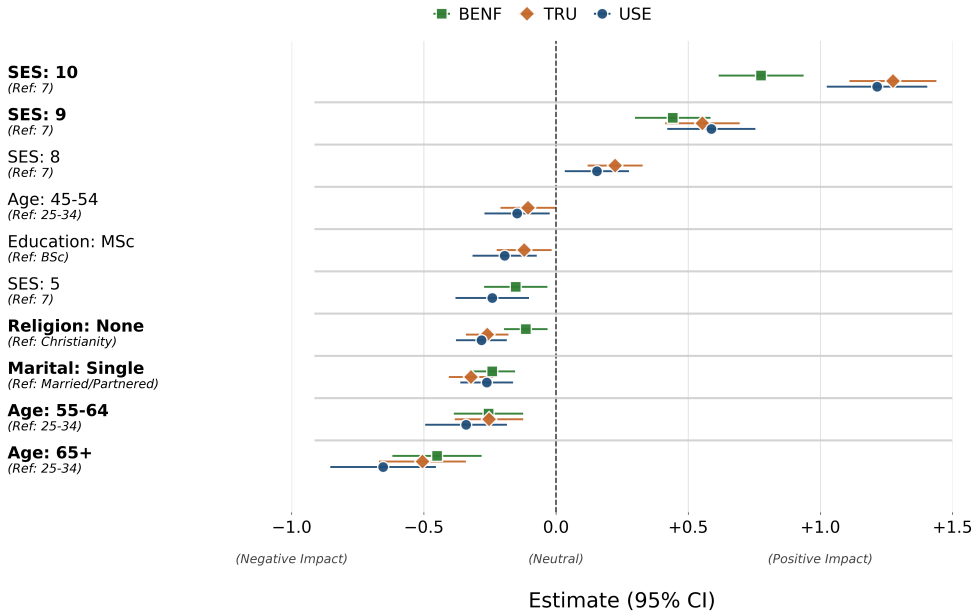


Figure 5: Mixed-model estimates of demographic factors (fixed effects) that are significant in at least 2 of the 3 constructs (BENF, USE, TRU). The ones that are significant in all three are in bold. The positive and negative impacts depend on the reference category.

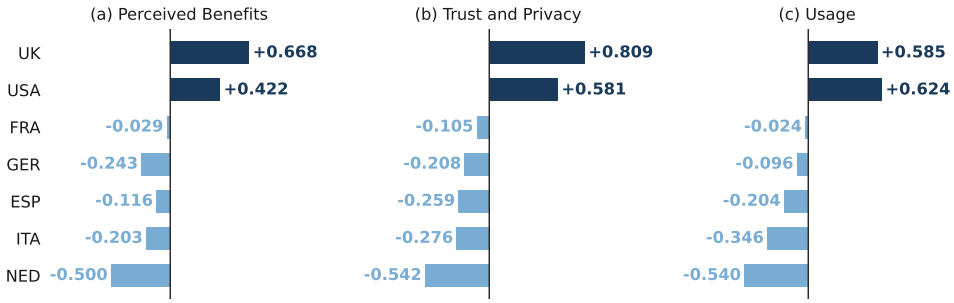


Figure 6: Country intercepts (random effects) between the three models (BENF, USE, TRU). Being from the UK and the USA is associated with positive perceptions, while mainland European countries have more negative baselines.

to §4.1, the demographic composition across countries is not necessarily similar. For example, UK and US participants are predominantly younger, highly educated, report higher SES, are more religious, and have higher marriage rates (see Figure 3). Therefore, in this section, we answer *RQ2* by examining whether observed cross-country differences reflect genuine cultural effects or are an artifact of demographic composition.

Methodology We use a Cumulative Linear Mixed Model (CLMM) to predict single constructs with demographic variables as fixed effects and country as a random effect. We excluded demographic categories with too few samples (< 50) and collapsed SES scores 1–3 into a single lowest category due to small cell sizes. The most populous category within each variable is selected as the reference group. We report results for three models that estimate demographic and cultural effects on Perceived Benefits (BENF), Trust & Privacy (TRU), and Usage Intention (USE), respectively.

Results The full result tables for all three models are presented in Appendix E. Figure 5 provides a visual outline of the fixed effects that are significant in at least two of the models. We find that SES does not linearly influence perceptions, but rather, being in the highest group (>7) has a noticeable impact. Similarly, being older than 45 lowers perceptions exponentially, and being 65+ is the most negative effect across all three models. However, being younger (18-24) does not consistently have a positive influence.

Religion shows steady findings across models, with religious individuals having more positive per-

ceptions. The differences between Christianity, Islam, and other religions are insignificant compared to the divide between believers and non-believers, suggesting that specific religious affiliation matters less than whether someone holds a religious belief. Similarly, marital status is consistent with married or partnered individuals showing more positive perceptions than singles across all models. However, there isn't a significant difference between married/partnered and divorced individuals. We observe that gender does play a significant role in only one construct (trust and privacy), with a small effect size. The education group also provides inconsistent findings. MSc graduates (compared to BSc) have more negative perceptions in two of the three models. However, the effect sizes are small, and we don't see any other differences between education categories.

The Intraclass Correlation Coefficient (ICC), which quantifies how much of the variance is attributable to random effects, is 4.2% for BENF, 6.1% for TRU, and 5.0% for USE. This suggests that demographics are the primary drivers of variation in perceptions. Likelihood ratio tests comparing the full mixed models against fixed-effects-only specifications confirm that the country-level random effect is statistically significant across all models ($p < .001$), indicating that meaningful between-country variation persists even after accounting for demographic composition. Figure 6 shows that across all three models, the UK and the US have large, positive intercepts. Meanwhile, mainland European countries have negative ones. The Netherlands consistently shows the largest negative effect, in line with previous findings (see Figure 4). Besides that, compared to other countries, France is more neutral, with intercepts closer to zero.

In summary, **both demographics and cultural background significantly shape user perceptions, though demographics account for the majority of the variance**. Demographically, high socioeconomic status is the strongest driver of positive perceptions, followed by being 25-45, religious, and married or partnered. Culturally, even after controlling for demographic differences, a significant geographic divide remains: users in the Anglosphere (the UK and the USA) hold notably positive views toward using LLMs for emotional support, whereas mainland European countries display more negative perceptions, with the Netherlands being the most skeptical.

4.4 Prompt Analysis

Methodology To address *RQ3*, we analyze the prompts shared by participants during the survey. Of the 1,343 users, 54.4% opted to share their most recent emotional support message (n=731). Sharing rates varied considerably across countries: the UK was most willing (67.4%), followed by Italy (62.1%), France (58.9%), the USA (58.2%), and the Netherlands (50.8%), while Spain (46.5%) and Germany (38.5%) shared the least. To identify categories, we cluster the prompts with HDBSCAN (McInnes et al., 2017) and UMAP (McInnes et al., 2020). We embedded the prompts using Sentence-Transformer (Reimers and Gurevych, 2019) and the multilingual M3-Embedding (Chen et al., 2024). Clusters are assigned a category through a combination of GPT-5.2 (OpenAI, 2025) and manual evaluation. After identifying mental wellbeing prompts, we applied the same clustering logic to analyze the underlying topics within this subset.

Results Of the 731 shared prompts, 267 (36.5%) are about mental wellbeing and interpersonal connection. The remainder covered physical health and bodily autonomy (n=93), productivity and knowledge (n=158), and others, such as filler text or refusals (n=213). Within the wellbeing and interpersonal connection clusters, we identified five key topics: general stress and anxiety, relationship conflict, family and caregiving, loneliness, and trauma and mental health. Topic distributions are consistent across countries (see Figure 14). Examples of shared prompts representing these key topics are shown in Table 10 in Appendix F.

5 Discussion

Our results show heterogeneity along geographic and sociodemographic lines: Demographics play a significant role in shaping perceptions, with SES and age group emerging as the most influential determinants. While Anglosphere users present high trust, mainland European users remain skeptical regarding how LLMs handle sensitive information. Perceptions are particularly negative in the Netherlands, consistent with previous findings showing Dutch citizens rank last globally in attitudes toward AI (Autoriteit Persoonsgegevens, 2025). This is due to a complex interaction between mental health stigma, access to care, digital literacy, and privacy concerns.

Regarding *mental health stigma*, "psychological safety" is a key driver of adoption: users readily

disclose sensitive information because they do not fear human judgment, consistent with the model of computer-mediated communication (Joinson, 2001). Users are not just using LLMs as a cheaper substitute for human therapists; they are seeking them out *because* they are not human. However, this perceived safety may be partly illusory: LLMs systematically over-affirm users rather than offering critical engagement (Cheng et al., 2025).

Mental health stigma further explains country-wise differences: Participants in the UK are the most common adopters of LLMs for emotional wellbeing and report more positive perceptions. Notably, adoption rates are much higher than in the US despite similar ratings across constructs, possibly reflecting differing attitudes towards mental health support, as it is more stigmatised in the U.K. (Digiuni et al., 2013; Todd and Shapira, 1974). Another possible explanation is *access*. The UK offers accessible, free-at-the-point-of-use care through the National Health Service (NHS), but with significant wait times. Therefore, Britons may perceive LLMs as beneficial for emotional support. Similarly, mental health stigma in Spain has significantly decreased over the last few years (Varaona et al., 2024), but mental health support access remains low in many European countries (Fiorillo, 2025; Oblak, 2025).

Our results broadly align with general adoption patterns for LLMs: their use is more common among younger and richer people (Bassignana et al., 2025). These groups are generally considered to have higher digital literacy. However, *digital literacy* involves not only the ability to use technologies, but to engage with them critically.[4] Given the uncertain benefits and dangers of relying on such systems in emotionally sensitive contexts, as well as the potential *privacy implications* of sharing personal information with them, it is not clear that this practice necessarily reflects digital literacy. Dutch Data Protection Authority (AP) warned that commercial AI therapy apps frequently harvest sensitive data (Autoriteit Persoonsgegevens, 2025). Furthermore, OpenAI's CEO acknowledged the absence of legal "doctor-patient confidentiality" when using chatbots as therapists (Perez, 2025). While users feel safe, their data remains legally unprotected.

Beyond privacy concerns, emotional interaction with AI systems raises important ethical concerns (Curry and Cercas Curry, 2023; Gabriel et al., 2024). Even among countries with lower adoption rates, 1 in 5 people report getting emotional support from LLMs. However, we currently lack benchmarks and guidelines regarding desirable LLM behaviour in emotional support roles. Existing work raises serious concerns: LLMs consistently prioritize companionship-reinforcing over boundary-maintaining behaviours (Kaffee et al., 2025), foster dependency with documented risks of self-harm in vulnerable users (Chu et al., 2025), and, despite displaying empathy, fail to interpret emotions genuinely and can make value judgements about certain identities (Cuadra et al., 2024). These limitations are compounded by a broader structural gap: emotion modeling in NLP has largely assumed emotions are universally experienced, failing to account for demographic and cultural variation (Plaza-del Arco et al., 2024a), and NLP systems further misascribe emotions based on stereotypes (Plaza-del Arco et al., 2024b,c).

Together, our findings point to a widening gap between the perceived emotional competence of LLMs and their demonstrated limitations in sensitive interactions. As adoption grows, benchmarks, governance frameworks, and AI literacy programmes that help users critically assess these systems become increasingly urgent.

6 Conclusion

We survey the use of LLMs for emotional and mental wellbeing support across seven countries, examining how cultural and sociodemographic factors shape adoption and user perceptions. We find statistically significant differences in both adoption rates and perceptions across all measured constructs. In particular, higher socioeconomic status, younger age, being partnered, and religious belief are consistently associated with more positive perceptions of trust, perceived benefits, and usage intention. We collect 731 prompts from real prior interactions, finding that use centers on stress, anxiety, loneliness, and relationship difficulties. Our cross-cultural analysis reveals a robust Anglosphere–Continental Europe divide that persists even after controlling for demographics. Future work should explore adoption of these tools beyond Western contexts, develop multilingual benchmarks for emotionally sensitive interactions, and establish governance frameworks and AI literacy programmes to ensure safe and informed use.

[4] https://www.cedefop.europa.eu/en/tools/vet-glossary/glossary/digitale-alphabetisierung

Limitations

Several limitations should be acknowledged. First, regarding sample representativeness: demographic composition varied across countries in age, education, and socioeconomic status, and within countries, the demographic profiles of AI users and non-users diverged notably, particularly in the UK and the USA. Additionally, panel respondents may differ from the general population in digital literacy. Second, the user/non-user classification relied on a single self-report filter question, which may be subject to social desirability bias or varying interpretations of what constitutes "emotional support" across languages and cultures. Third, the seven-country sample is limited to Western and Southern European contexts plus the USA, which may limit generalizability to regions where norms around self-disclosure, mental health, and technology use differ. Fourth, Likert-scale self-reports are susceptible to acquiescence bias and cultural response styles, which may partially confound cross-cultural comparisons.

Ethical Considerations

The present study was approved by the Institutional Ethics Board[5]. Research on the use of LLMs for emotional support and mental wellbeing raises ethical considerations related to privacy, mental health sensitivity, and responsible interpretation of user experiences. Because individuals may disclose highly personal information when interacting with AI systems in emotionally vulnerable contexts, careful attention to data collection, participant protection, and reporting practices is required.

Users may discuss mental health struggles, interpersonal conflicts, or other highly personal topics when interacting with LLMs. Collecting descriptions or examples of such interactions therefore poses risks of exposing sensitive information. To mitigate this, we did not collect directly identifying information and asked participants not to include identifying details in open-ended responses or example prompts. Data were analyzed in aggregated form, and any illustrative examples were edited or paraphrased to prevent re-identification. More broadly, our findings highlight a tension between the perceived privacy of LLM interactions and the potential sensitivity of the information users disclose to AI systems.

Participants were informed about the purpose of the study, the types of information collected, including sociodemographic variables and descriptions of LLM use, and how their responses would be used for research and publication. Participation was voluntary, and participants could withdraw before submitting the survey.

Although the study does not target clinical populations, individuals seeking emotional support from LLMs may be experiencing distress or limited access to other forms of support. Survey questions were therefore designed to focus on patterns of use and perceptions (e.g., acceptance of and anthropomorphic attitudes toward LLMs) rather than encouraging detailed disclosure of traumatic experiences. When reporting results, we avoid framing LLMs as substitutes for professional mental health care and instead emphasize that the study captures users' perceptions and behaviors.

To minimize ambiguity and avoid potentially leading or stigmatizing questions about mental health, the survey relied on established measurement scales for constructs such as technology acceptance and anthropomorphic perceptions of AI. Using validated instruments helps ensure that questions are interpretable, comparable to prior research, and less likely to introduce unintended framing effects when studying sensitive behaviors related to emotional support.

Participants received compensation aligned with the norms of the recruitment platform and the estimated survey duration.

[5] Institution name withheld for blind review

References

Nancy E Adler, Elissa S Epel, Grace Castellazzo, and Jeannette R Ickovics. 2000. Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, white women. *Health psychology*, 19(6):586.

Marta Andersson. 2025. Companionship in code: Ai's role in the future of human connection. *Humanities and Social Sciences Communications*, 12(1):1–7.

Autoriteit Persoonsgegevens. 2025. AI & Algorithmic Risks Report Netherlands (ARR). Technical report, Dutch Data Protection Authority.

Elisa Bassignana, Amanda Cercas Curry, and Dirk Hovy. 2025. The AI gap: How socioeconomic status affects language technology interactions. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 18647–18664, Vienna, Austria. Association for Computational Linguistics.

Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2018. Assessing the usability of a chatbot for mental health care. In *International conference on internet science*, pages 121–132. Springer.

Aaron Chatterji, Thomas Cunningham, David J. Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. How people use chatgpt. Working Paper 34255, National Bureau of Economic Research.

Beenish Moalla Chaudhry and Hamid Reza Debi. 2024. User perceptions and experiences of an AI-driven conversational agent for mental health support. *mHealth*, 10:22.

Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. M3-embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2318–2335, Bangkok, Thailand. Association for Computational Linguistics.

Myra Cheng, Sunny Yu, Cinoo Lee, Pranav Khadpe, Lujain Ibrahim, and Dan Jurafsky. 2025. Elephant: Measuring and understanding social sycophancy in llms. *arXiv preprint arXiv:2505.13995*.

Minh Duc Chu, Patrick Gerard, Kshitij Pawar, Charles Bickham, and Kristina Lerman. 2025. Illusions of intimacy: Emotional attachment and emerging psychological risks in human-ai relationships. *arXiv preprint arXiv:2505.11649*.

Andrea Cuadra, Maria Wang, Lynn Andrea Stein, Malte F. Jung, Nicola Dell, Deborah Estrin, and James A. Landay. 2024. 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*, CHI '24, New York, NY, USA. Association for Computing Machinery.

Alba Curry and Amanda Cercas Curry. 2023. Computer says "no": The case against empathetic conversational AI. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8123–8130, Toronto, Canada. Association for Computational Linguistics.

Fred D. Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 13(3):319–340.

Julian De Freitas, Stuti Agarwal, Bernd Schmitt, and Nick Haslam. 2023. Psychological factors underlying attitudes toward ai tools. *Nature Human Behaviour*, 7(11):1845–1854.

Malena Digiuni, Fergal W Jones, and Paul M Camic. 2013. Perceived social stigma and attitudes towards seeking therapy in training: A cross-national study. *Psychotherapy*, 50(2):213.

Andrea Fiorillo. 2025. A roadmap for better and personalized mental health care in europe: the priorities of the european psychiatric association. *European Psychiatry*, 68(1):e60.

FirstPageSage. 2026. Chatgpt usage statistics: February 2026. Accessed: 2026-02-26.

Iason Gabriel, Arianna Manzini, Geoff Keeling, Lisa Anne Hendricks, Verena Rieser, Hasan Iqbal, Nenad Tomašev, Ira Ktena, Zachary Kenton, Mikel Rodriguez, and 1 others. 2024. The ethics of advanced ai assistants. *arXiv preprint arXiv:2404.16244*.

Anna-Carolina Haensch. 2025. "it listens better than my therapist": Exploring social media discourse on llms as mental health tool. *arXiv preprint arXiv:2504.12337*.

Yu Hou, Hal Daumé Iii, and Rachel Rudinger. 2025. Language models predict empathy gaps between social in-groups and out-groups. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 12288–12304, Albuquerque, New Mexico. Association for Computational Linguistics.

Adam N Joinson. 2001. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European journal of social psychology*, 31(2):177–192.

Lucie-Aimée Kaffee, Giada Pistilli, and Yacine Jernite. 2025. Intima: a benchmark for human-ai companionship behavior. *arXiv preprint arXiv:2508.09998*.

Frederic Marimon, Natalia Amat-Lefort, Marta Mas-Machuca, and Anna Akhmedova. 2024. GPT-QUAL [dataset and questionnaire].

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205.

Leland McInnes, John Healy, and James Melville. 2020. Umap: Uniform manifold approximation and projection for dimension reduction. *Preprint*, arXiv:1802.03426.

Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1):81–103.

Lara Oblak. 2025. Public mental health stigma and suicide rates across europe. *Frontiers in public health*, 13:1554072.

OpenAI. 2025. Introducing GPT-5.2. `https://openai.com/index/introducing-gpt-5-2/`. Accessed: 2026-03-04.

Aditya Pandya, Param Lodha, and Amit Ganatra. 2024. Is ChatGPT ready to change mental healthcare? Challenges and considerations: a reality-check. *Frontiers in Human Dynamics*, 5:1289255.

Hashai Papneja and Nikhil Yadav. 2025. Self-disclosure to conversational ai: a literature review, emergent framework, and directions for future research. *Personal and ubiquitous computing*, 29(2):119–151.

Sarah Perez. 2025. Sam altman warns there's no legal confidentiality when using chatgpt as a therapist.

Flor Miriam Plaza-del Arco, Alba A. Cercas Curry, Amanda Cercas Curry, and Dirk Hovy. 2024a. Emotion analysis in NLP: Trends, gaps and roadmap for future directions. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5696–5710, Torino, Italia. ELRA and ICCL.

Flor Miriam Plaza-del Arco, Amanda Cercas Curry, Alba Curry, Gavin Abercrombie, and Dirk Hovy. 2024b. Angry men, sad women: Large language models reflect gendered stereotypes in emotion attribution. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7682–7696, Bangkok, Thailand. Association for Computational Linguistics.

Flor Miriam Plaza-del Arco, Amanda Cercas Curry, Susanna Paoli, Alba Cercas Curry, and Dirk Hovy. 2024c. Divine LLaMAs: Bias, stereotypes, stigmatization, and emotion representation of religion in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4346–4366, Miami, Florida, USA. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Beatrice Savoldi, Giuseppe Attanasio, Olga Gorodetskaya, Marta Marchiori Manerba, Elisa Bassignana, Silvia Casola, Matteo Negri, Tommaso Caselli, Luisa Bentivogli, Alan Ramponi, and 1 others. 2025. Generative ai practices, literacy, and divides: An empirical analysis in the italian context. *arXiv preprint arXiv:2512.03671*.

Matthias Schmidmaier, Jan Rupp, Darina Cvetanova, and Sven Mayer. 2024. Perceived empathy of technology scale (PETS): Measuring empathy of systems toward the user. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–18.

Judy L Todd and Ariella Shapira. 1974. Us and british self-disclosure, anxiety, empathy, and attitudes to psychotherapy. *Journal of Cross-Cultural Psychology*, 5(3):364–369.

Alec Tyson, Giancarlo Pasquini, Alison Spencer, and Cary Funk. 2023. 60% of Americans would be uncomfortable with provider relying on AI in their own health care. Technical report, Pew Research Center.

Andrea Varaona, Rosa M Molina-Ruiz, Luis Gutiérrez-Rojas, Maria Perez-Páramo, Guillermo Lahera, Carolina Donat-Vargas, and Miguel Angel Alvarez-Mon. 2024. Snapshot of knowledge and stigma toward mental health disorders and treatment in spain. *Frontiers in Psychology*, 15:1372955.

Shenghan Wu, Yimo Zhu, Wynne Hsu, Mong-Li Lee, and Yang Deng. 2025. From personas to talks: Revisiting the impact of personas on LLM-synthesized emotional support conversations. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 5439–5453, Suzhou, China. Association for Computational Linguistics.

A User Interaction Patterns and Perceived Benefits

Figure 7 presents an overview of user interaction patterns with AI agents for mental wellbeing and emotional support, including the main usage purposes.

Figure 8 shows the key benefits of seeking support from AI agents, ranked by their average scores provided by users (from 1 to 5). In Figure 9, we present the average scores for different Perceived Benefits across countries.

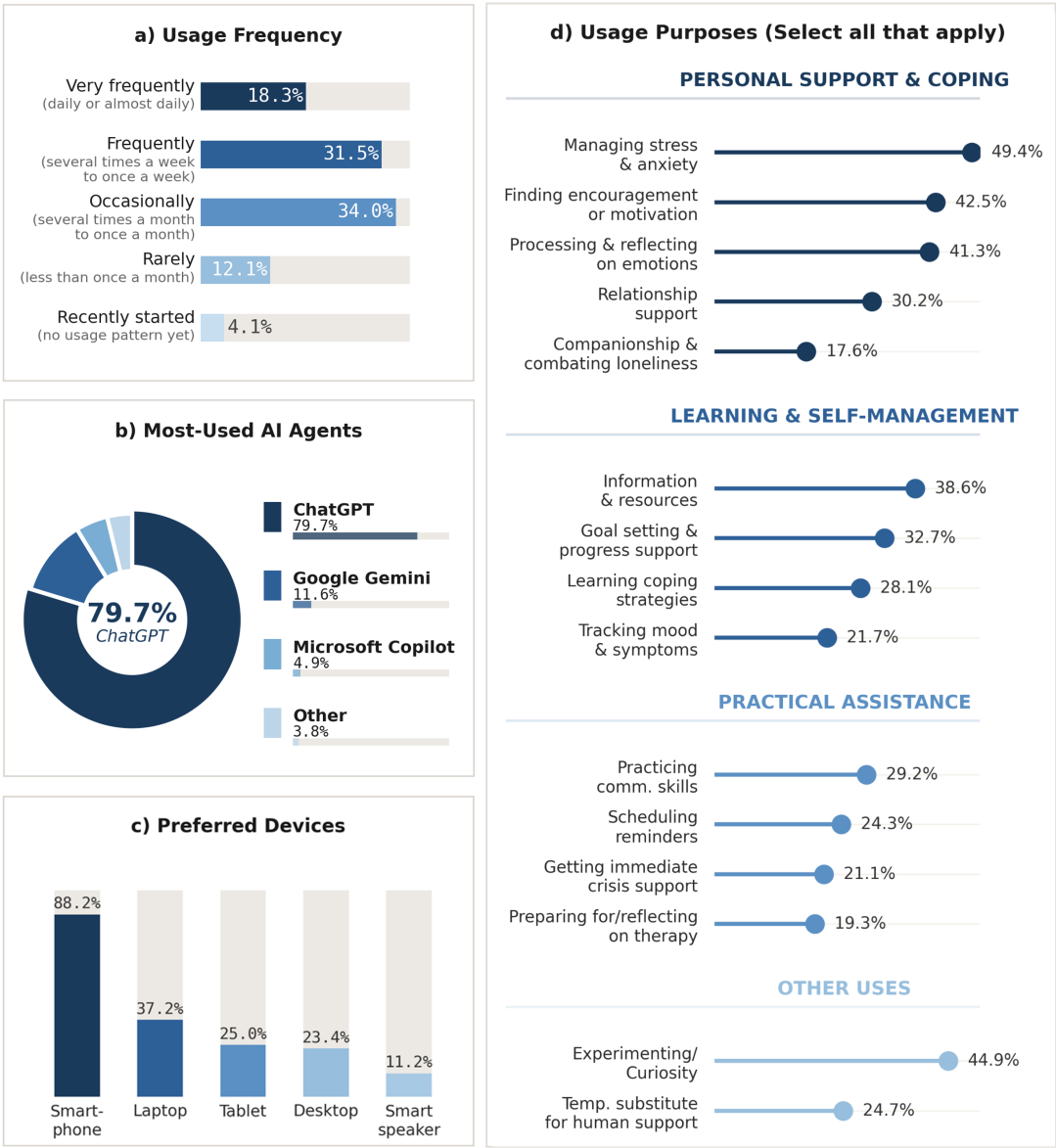


Figure 7: Overview of user interaction patterns with AI agents for mental wellbeing and emotional support: distribution by (a) Frequency of use, (b) Most-used agents, (c) Preferred devices, and (d) Grouped functional purposes.

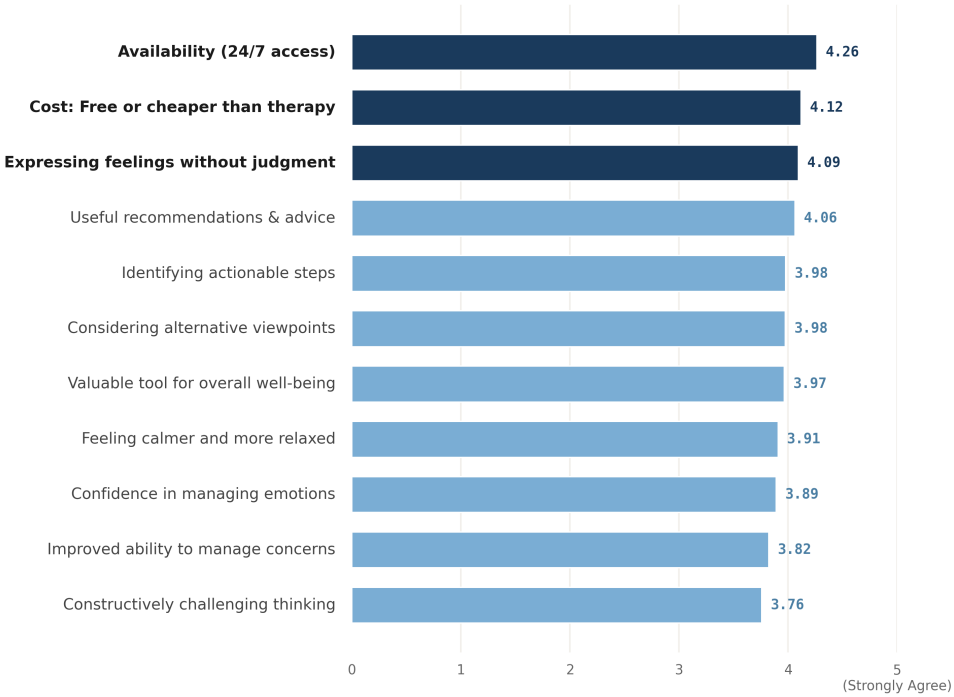


Figure 8: Top-rated Perceived Benefits of AI conversational agents for mental wellbeing and emotional support. Global average scores (Scale: 1-5). Top 3 benefits highlighted.

Figure 9: Perceived Benefits across countries. Average scores (Scale: 1-5). Countries ordered by overall enthusiasm (Left to Right).

B User and Non-User Demographic Breakdown

This section provides the total sample sizes of users and non-users across countries (Table 1), a full demographic breakdown of both user groups, all countries combined (Table 2), and a demographic breakdown of user groups per country (Table 3).

Country	Users	Non Users	Total	% User
ESP	200	208	408	49.0%
FRA	190	750	940	20.2%
GER	192	590	782	24.6%
ITA	198	480	678	29.2%
NED	195	741	936	20.8%
UK	184	128	312	59.0%
USA	184	401	585	31.5%
Total	**1,343**	**3,298**	**4,641**	**28.9%**

Table 1: Adoption rates of AI conversational agents for mental wellbeing and emotional support across seven countries.

	Users 1,343	Non-Users 3,298	Total 4,641
Gender (%)			
Men	660 (49.1)	1,598 (48.5)	2,258 (48.7)
Women	**681 (50.7)**	**1,687 (51.2)**	**2,368 (51.0)**
Non-binary/Other	2 (0.2)	13 (0.3)	15 (0.3)
Age (%)			
Under 18	2 (0.1)	22 (0.7)	24 (0.5)
18–24	109 (8.1)	78 (2.4)	187 (4.0)
25–34	**409 (30.5)**	237 (7.2)	646 (13.9)
35–44	370 (27.6)	459 (13.9)	829 (17.9)
45–54	249 (18.5)	697 (21.1)	946 (20.4)
55–64	133 (9.9)	808 (24.5)	941 (20.3)
65+	71 (5.3)	**997 (30.2)**	**1,068 (23.0)**
Education (%)			
Primary school	17 (1.3)	177 (5.4)	194 (4.2)
HS	277 (20.7)	**1,201 (36.8)**	**1,478 (32.1)**
Some college, no degree	83 (6.2)	230 (7.1)	313 (6.8)
Associate degree	138 (10.3)	480 (14.7)	618 (13.4)
BSc	**308 (23.0)**	465 (14.3)	773 (16.8)
MSc	263 (19.6)	290 (8.9)	553 (12.0)
PhD	78 (5.8)	60 (1.8)	138 (3.0)
Professional degree	175 (13.1)	357 (11.0)	532 (11.6)
Religion (%)			
Christianity	**797 (59.3)**	**1,791 (54.3)**	**2,588 (55.8)**
Islam	83 (6.2)	78 (2.4)	161 (3.5)
No religion	373 (27.8)	1,184 (35.9)	1,557 (33.6)
Other	52 (3.9)	119 (3.6)	171 (3.7)
Prefer not to say	38 (2.8)	126 (3.8)	164 (3.5)
Marital Status (%)			
Single	385 (28.7)	807 (24.5)	1,192 (25.7)
Married/Partnership	**864 (64.3)**	**1,881 (57.0)**	**2,745 (59.1)**
Divorced	62 (4.6)	342 (10.4)	404 (8.7)
Separated	9 (0.7)	60 (1.8)	69 (1.5)
Widowed	15 (1.1)	179 (5.4)	194 (4.2)
Prefer not to say	8 (0.6)	29 (0.9)	37 (0.8)
SES (%)			
1 (Worst off)	11 (0.8)	66 (2.0)	77 (1.7)
2	17 (1.3)	96 (2.9)	113 (2.4)
3	48 (3.6)	220 (6.7)	268 (5.8)
4	114 (8.5)	340 (10.3)	454 (9.8)
5	181 (13.5)	658 (20.0)	839 (18.1)
6	240 (17.9)	**717 (21.7)**	957 (20.6)
7	**306 (22.8)**	688 (20.9)	**994 (21.4)**
8	239 (17.8)	341 (10.3)	580 (12.5)
9	101 (7.5)	69 (2.1)	170 (3.7)
10 (Best off)	86 (6.4)	103 (3.1)	189 (4.1)

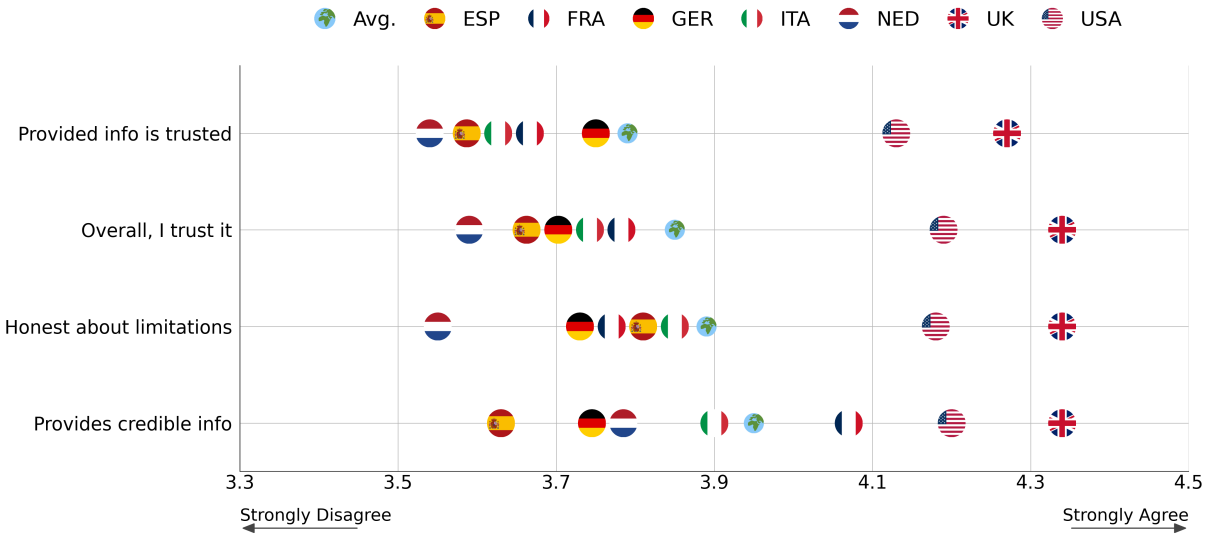
Table 2: Breakdown of demographic differences between users and non-users. The most common groups are boldfaced.

	ESP			FRA			ITA			GER			NED			UK			USA		
	U 200	NU 208	T 408	U 190	NU 750	T 940	U 198	NU 480	T 678	U 192	NU 590	T 782	U 195	NU 741	T 936	U 184	NU 128	T 312	U 184	NU 401	T 585
Gender (%)																					
Men	99	79	178	91	373	464	**99**	**269**	**368**	94	**314**	**408**	95	**384**	**479**	90	**67**	**157**	**92**	112	204
Women	**100**	**127**	**227**	**98**	**377**	**475**	**99**	210	309	**98**	273	371	**100**	354	454	**94**	61	155	**92**	**285**	**377**
Non-binary/Other	1	2	3	1	0	1	0	1	1	0	3	3	0	3	3	0	0	0	0	4	4
Age (%)																					
Under 18	0	3	3	1	6	7	0	2	2	1	3	4	0	5	5	0	1	1	0	2	2
18–24	22	5	27	21	27	48	16	7	23	22	14	36	16	11	27	8	0	8	4	14	18
25–34	53	21	74	**60**	53	113	**55**	35	90	39	30	69	44	41	85	**117**	23	**140**	41	34	75
35–44	42	31	73	45	93	138	43	78	121	**51**	77	128	**58**	114	172	44	19	63	**87**	47	134
45–54	**65**	**69**	**134**	33	178	211	48	134	**182**	28	102	130	38	135	173	10	16	26	27	63	90
55–64	16	57	73	17	**202**	**219**	27	**137**	164	34	160	194	22	165	187	4	25	29	13	62	75
65+	2	22	24	13	191	204	9	87	96	17	**204**	**221**	17	**270**	**287**	1	**44**	45	12	**179**	**191**
Education (%)																					
Primary school	2	10	12	5	57	62	3	35	38	4	50	54	2	13	15	0	2	2	1	10	11
HS	18	24	42	**61**	**351**	**412**	**68**	**263**	**331**	**59**	155	214	36	236	272	9	**37**	46	26	**135**	**161**
Some college	8	9	17	11	41	52	25	36	61	15	23	38	2	13	15	9	23	32	13	85	98
Associate degree	31	41	72	39	125	164	9	15	24	1	4	5	**48**	**248**	**296**	0	2	2	10	45	55
BSc	59	**52**	111	35	84	119	38	40	78	30	47	77	**48**	126	174	46	33	**79**	52	83	135
MSc	12	10	22	30	61	91	32	43	75	27	64	91	41	67	108	**58**	13	71	**63**	32	95
PhD	8	10	18	2	11	13	4	7	11	4	6	10	5	10	15	45	10	55	10	6	16
Professional degree	**62**	50	**112**	6	10	16	19	39	58	50	**224**	**274**	13	22	35	16	7	23	9	5	14
Religion (%)																					
Christianity	**124**	**129**	**253**	**87**	**368**	**455**	**142**	**367**	**509**	**100**	**286**	**386**	67	287	354	**153**	**83**	**236**	**124**	**271**	**395**
Islam	2	1	3	24	35	59	2	2	4	12	13	25	15	19	34	8	1	9	20	7	27
No religion	65	68	133	68	294	362	41	91	132	66	260	326	**100**	**370**	**470**	16	36	52	17	65	82
Other	4	7	11	4	13	17	6	12	18	7	15	22	9	24	33	1	4	5	21	44	65
Prefer not to say	5	3	8	7	40	47	7	8	15	7	16	23	4	41	45	6	4	10	2	14	16
Marital Status (%)																					
Single	55	50	105	69	172	241	82	152	234	66	144	210	67	157	224	22	30	52	24	102	126
Married/Partnership	**130**	**123**	**253**	**114**	**467**	**581**	**102**	**263**	**365**	**103**	**320**	**423**	**110**	**447**	**557**	**160**	**82**	**242**	**145**	**179**	**324**
Divorced	11	26	37	6	61	67	6	37	43	11	79	90	16	86	102	2	8	10	10	45	55
Separated	2	3	5	1	21	22	0	9	9	3	11	14	1	5	6	0	2	2	2	9	11
Widowed	2	4	6	0	26	26	2	14	16	7	32	39	1	36	37	0	5	5	3	62	65
Prefer not to say	0	2	2	0	3	3	6	5	11	2	4	6	0	10	10	0	1	1	0	4	4
SES (%)																					
1 (Worst off)	1	1	2	1	13	14	1	10	11	3	17	20	2	6	8	1	1	2	2	18	20
2	1	1	2	6	28	34	3	3	6	2	21	23	3	21	24	1	4	5	1	18	19
3	5	10	15	16	51	67	4	30	34	12	45	57	3	31	34	1	9	10	7	44	51
4	22	12	34	31	114	145	22	45	67	12	76	88	11	44	55	5	15	20	11	34	45
5	30	44	74	35	**205**	**240**	39	98	137	28	**126**	154	16	81	97	12	**32**	44	21	**72**	93
6	**53**	**57**	**110**	**46**	166	212	**44**	**144**	**188**	**39**	118	**157**	23	146	169	13	22	35	22	64	86
7	46	45	91	32	105	137	40	98	138	**39**	105	144	**71**	**245**	**316**	**45**	21	**66**	33	69	**102**
8	26	24	50	10	48	58	33	29	62	34	55	89	54	136	190	44	12	56	**38**	37	75
9	12	6	18	5	6	11	6	9	15	14	12	26	8	19	27	29	3	32	27	14	41
10 (Best off)	4	8	12	8	14	22	6	14	20	9	15	24	4	12	16	33	9	42	22	31	53

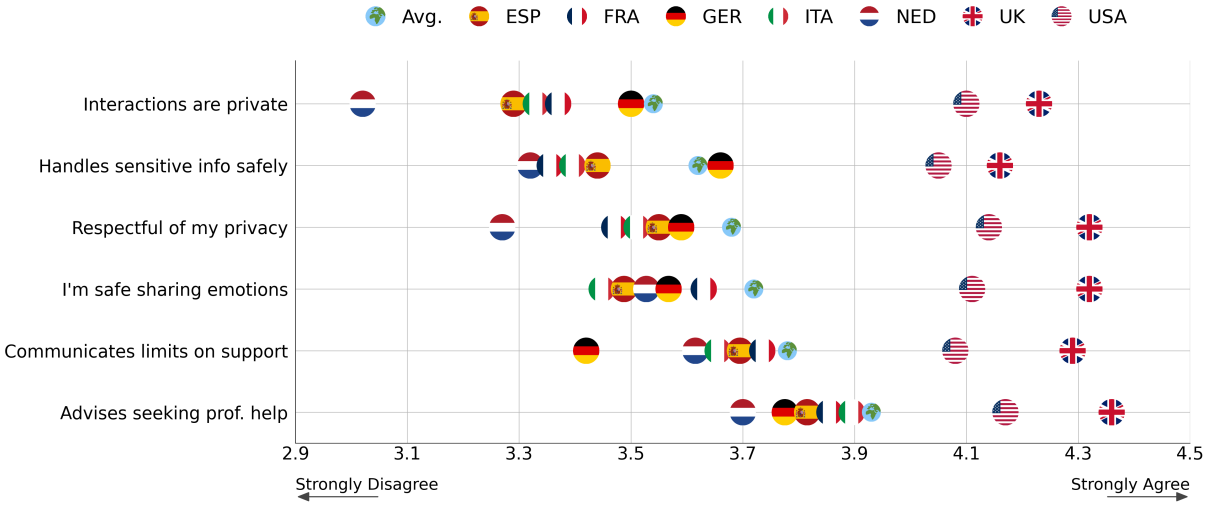
Table 3: Demographic breakdown by country for users (U), non-users (NU), and total (T). The most common group within each demographic category per country is boldfaced.

C Geographic Differences in Trust & Privacy Perceptions

Figure 10 presents mean agreement scores for trust and privacy perceptions of mental wellbeing chatbots across seven countries, with items sorted by ascending global average. In both panels, the UK and USA consistently score highest, clustering toward the 'Strongly Agree' end of the scale, while continental European countries tend to rate lower. Trust perceptions show relatively tight cross-country clustering on items like 'Provided info is trusted,' but greater spread on 'Provides credible info' and 'Honest about limitations.' Privacy perceptions exhibit a similar UK/USA–continental Europe divide, with the largest cross-country variation appearing on items related to emotional safety and professional help advice. Across both constructs, the Netherlands usually has the most negative perceptions, distinguishing itself from other European countries. Dutch participants especially rate LLMs' honesty about limitations and interaction privacy low, showing skepticism.



(a) Trust perceptions



(b) Privacy perceptions

Figure 10: Trust and privacy perceptions by country. Items sorted by global average (ascending). Scores represent mean agreement on a 1–5 Likert scale.

	BENF			USE			TRU		
Country	$\bar{x}$	$\tilde{x}$	σ	$\bar{x}$	$\tilde{x}$	σ	$\bar{x}$	$\tilde{x}$	σ
Combined	4.00	4.00	0.62	3.77	3.86	0.80	3.78	3.80	0.72
ESP	3.91	4.00	0.69	3.64	3.71	0.85	3.60	3.70	0.84
FRA	3.95	4.00	0.64	3.69	3.86	0.85	3.67	3.80	0.74
ITA	3.89	3.90	0.55	3.53	3.57	0.73	3.63	3.70	0.62
GER	3.87	4.00	0.67	3.68	3.86	0.81	3.65	3.80	0.71
NED	3.75	3.80	0.56	3.42	3.43	0.79	3.49	3.60	0.59
UK	**4.39**	**4.40**	0.44	**4.22**	**4.29**	0.61	**4.30**	**4.40**	0.53
USA	4.24	4.30	0.46	4.17	4.21	0.53	4.14	4.25	0.59

Table 4: Descriptive statistics for each construct by country. $\bar{x}$ = mean; $\tilde{x}$ = median; σ = standard deviation. The highest mean and median scores are bold-faced.

D Descriptive Statistics for Constructs

Descriptive statistics for all constructs per country are shown in Table 4. The demographic breakdown of constructs is represented in Table 5. The full statistical test results mentioned in §4.2 are displayed in Table 6, and visualized in Figure 11.

Looking at the tables, the UK has the highest mean and median scores in all constructs, with the US following closely. The table and the figure confirm that both the UK and the USA statistically differ from the rest of the countries across all constructs. Within, they only differ in BENF. It is evident that the Anglosphere forms a separate cluster. Following this, the remaining countries form their own cluster: Germany does not statistically differ from any mainland European country except the Netherlands. Italy aligns with the Netherlands, Spain, and Germany in all constructs and differs from France in one construct. Spain differs from the Netherlands in two constructs. This pattern confirms that the Netherlands distinguishes itself from the others.

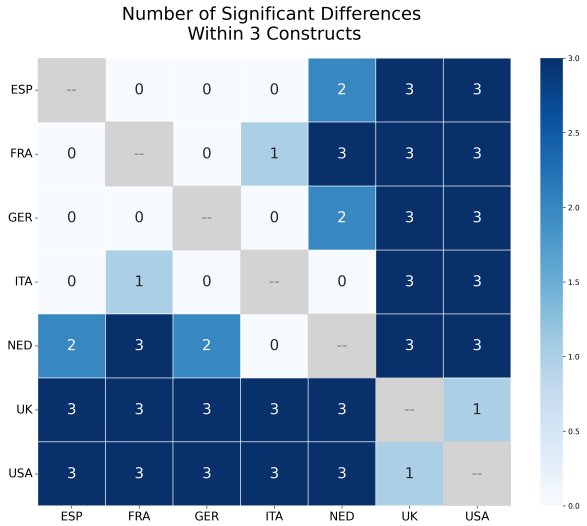


Figure 11: Pairwise country comparisons showing the number of constructs (out of 3) that differ significantly between each pair. The UK and the USA clearly deviate from the mainland European countries and form their own cluster.

	BENF			USE			TRU		
	$\bar{x}$	$\tilde{x}$	σ	$\bar{x}$	$\tilde{x}$	σ	$\bar{x}$	$\tilde{x}$	σ
Gender									
Female	**4.00**	**4.00**	0.61	3.73	**3.86**	0.82	3.74	3.80	0.74
Male	3.99	**4.00**	0.62	**3.79**	**3.86**	0.77	**3.81**	**3.90**	0.71
Age									
18–24	3.89	4.00	0.65	3.57	3.71	0.85	3.55	3.70	0.81
25–34	**4.07**	**4.10**	0.60	**3.87**	**4.00**	0.76	**3.87**	**3.90**	0.70
35–44	4.03	**4.10**	0.59	3.85	**4.00**	0.75	3.86	**3.90**	0.70
45–54	3.97	4.00	0.63	3.67	3.86	0.81	3.69	3.80	0.77
55–64	3.88	4.00	0.63	3.61	3.71	0.84	3.67	3.80	0.67
65+	3.82	3.80	0.64	3.48	3.57	0.87	3.59	3.70	0.61
Education									
Primary school	4.04	3.90	0.62	3.78	4.00	0.91	3.67	3.50	0.64
High school	3.91	4.00	0.62	3.67	3.71	0.79	3.71	3.80	0.66
Some college	3.98	4.00	0.64	3.58	3.71	0.86	3.70	3.80	0.73
Associate degree	3.85	3.85	0.64	3.57	3.71	0.87	3.58	3.70	0.81
BSc	4.03	4.10	0.57	3.80	3.86	0.76	3.80	3.90	0.73
MSc	4.06	4.10	0.62	3.82	4.00	0.82	3.88	3.90	0.75
PhD	**4.29**	**4.40**	0.51	**4.17**	**4.29**	0.63	**4.13**	**4.20**	0.68
Professional degree	3.93	4.00	0.64	3.76	3.86	0.74	3.71	3.80	0.67
Religion									
Christianity	4.06	**4.10**	0.60	3.85	**4.00**	0.78	3.87	3.90	0.72
Islam	**4.11**	**4.10**	0.50	**3.99**	**4.00**	0.70	**3.94**	**4.00**	0.68
No religion	3.85	3.90	0.65	3.50	3.57	0.83	3.54	3.60	0.71
Other	4.00	**4.10**	0.55	3.82	3.93	0.78	3.83	3.90	0.70
Prefer not to say	3.85	3.95	0.58	3.67	3.71	0.65	3.65	3.70	0.66
Marital Status									
Single	3.86	3.90	0.62	3.57	3.71	0.84	3.57	3.60	0.72
Married/Partnership	**4.06**	**4.10**	0.60	**3.86**	**4.00**	0.77	**3.89**	**4.00**	0.71
Divorced	3.99	4.05	0.54	3.68	3.86	0.72	3.66	3.80	0.51
Separated	3.34	3.60	0.57	3.29	3.43	0.74	3.00	3.60	0.90
Widowed	3.90	3.90	0.64	3.83	3.86	0.67	3.61	3.80	0.63
Prefer not to say	3.51	3.35	0.62	2.98	3.07	1.02	3.11	3.10	0.36
SES									
1 (Worst off)	4.12	4.30	0.72	3.78	4.00	1.05	3.93	4.00	0.79
2	3.74	3.90	0.85	3.59	4.00	0.96	3.49	3.70	0.94
3	3.90	3.90	0.71	3.66	3.71	0.88	3.54	3.60	0.78
4	3.92	4.00	0.66	3.66	3.71	0.83	3.61	3.65	0.75
5	3.87	3.90	0.65	3.54	3.57	0.79	3.60	3.70	0.77
6	3.99	4.00	0.57	3.70	3.71	0.76	3.72	3.80	0.63
7	3.94	4.00	0.57	3.66	3.86	0.79	3.69	3.75	0.69
8	3.97	4.00	0.62	3.79	3.86	0.78	3.85	3.90	0.68
9	4.28	4.30	0.41	4.16	4.14	0.56	4.15	4.20	0.50
10 (Best off)	**4.36**	**4.50**	0.62	**4.39**	**4.57**	0.63	**4.32**	**4.60**	0.78

Table 5: Descriptive statistics for each construct by demographic group. $\bar{x}$ = mean; $\tilde{x}$ = median; σ = standard deviation. The highest mean and median within each demographic category are bold-faced. Groups with $n < 5$ are omitted.

	BENF		USE		TRU	
	U	p_{adj}	U	p_{adj}	U	p_{adj}
UK–USA	**20035**	**.021**	18304	.878	19644	.069
UK–ESP	**26294**	**< .001**	**26142**	**< .001**	**27886**	**< .001**
UK–FRA	**25017**	**< .001**	**24341**	**< .001**	**26482**	**< .001**
UK–GER	**26482**	**< .001**	**25086**	**< .001**	**27288**	**< .001**
UK–ITA	**28062**	**< .001**	**28246**	**< .001**	**29109**	**< .001**
UK–NED	**29501**	**< .001**	**28706**	**< .001**	**30539**	**< .001**
USA–UK	**20035**	**.021**	18304	.878	19644	.069
USA–ESP	**23560**	**< .001**	**25414**	**< .001**	**25780**	**< .001**
USA–FRA	**22198**	**< .001**	**23558**	**< .001**	**24339**	**< .001**
USA–GER	**23852**	**< .001**	**24370**	**< .001**	**25166**	**< .001**
USA–ITA	**25302**	**< .001**	**27782**	**< .001**	**27068**	**< .001**
USA–NED	**27024**	**< .001**	**28290**	**< .001**	**28636**	**< .001**
ESP–UK	**26294**	**< .001**	**26142**	**< .001**	**27886**	**< .001**
ESP–USA	**23560**	**< .001**	**25414**	**< .001**	**25780**	**< .001**
ESP–FRA	18415	1.000	18276	1.000	18501	1.000
ESP–GER	20028	1.000	18797	1.000	19050	1.000
ESP–ITA	21062	1.000	22104	.266	20434	1.000
ESP–NED	**22992**	**.021**	**22960**	**.020**	22245	.123
FRA–UK	**25017**	**< .001**	**24341**	**< .001**	**26482**	**< .001**
FRA–USA	**22198**	**< .001**	**23558**	**< .001**	**24339**	**< .001**
FRA–ESP	18415	1.000	18276	1.000	18501	1.000
FRA–GER	16734	.812	17868	1.000	17962	1.000
FRA–ITA	20878	.364	**21834**	**.049**	19950	1.000
FRA–NED	**14247**	**< .001**	**14464**	**.002**	**15278**	**.032**
GER–UK	**26482**	**< .001**	**25086**	**< .001**	**27288**	**< .001**
GER–USA	**23852**	**< .001**	**24370**	**< .001**	**25166**	**< .001**
GER–ESP	20028	1.000	18797	1.000	19050	1.000
GER–FRA	16734	.812	17868	1.000	17962	1.000
GER–ITA	19466	1.000	21743	.097	19988	1.000
GER–NED	16000	.106	**14962**	**.006**	**15535**	**.037**
ITA–UK	**28062**	**< .001**	**28246**	**< .001**	**29109**	**< .001**
ITA–USA	**25302**	**< .001**	**27782**	**< .001**	**27068**	**< .001**
ITA–ESP	21062	1.000	22104	.266	20434	1.000
ITA–FRA	20878	.364	**21834**	**.049**	19950	1.000
ITA–GER	19466	1.000	21743	.097	19988	1.000
ITA–NED	16832	.194	18012	1.000	16968	.263
NED–UK	**29501**	**< .001**	**28706**	**< .001**	**30539**	**< .001**
NED–USA	**27024**	**< .001**	**28290**	**< .001**	**28636**	**< .001**
NED–ESP	**22992**	**.021**	**22960**	**.020**	22245	.123
NED–FRA	**14247**	**< .001**	**14464**	**.002**	**15278**	**.032**
NED–GER	16000	.106	**14962**	**.006**	**15535**	**.037**
NED–ITA	16832	.194	18012	1.000	16968	.263

Table 6: Pairwise Mann–Whitney U Statistics and Holm-Adjusted p-Values Across Constructs. Bold values indicate statistical significance after Holm correction ($p_{\text{adj}} < .05$).

E Mixed Model Analysis of Perceived Benefits, Trust, and Usage

We provide the full results table for the CLMMs in this section. Figure 12 shows the fixed effect estimates by age group, and Figure 13 shows the SES estimates across models; the demographics that had the largest negative and positive estimates, respectively.

The usage intention results (Table 8) closely mirror trust and privacy. Age again emerges as a consistent predictor: middle-aged adults (25–44) hold the most positive attitudes, while older adults (55+) show the largest negative effects. Unlike trust and privacy, however, Gender is not significant for usage intention. Among education levels, "Some college, no degree" is a significant negative predictor alongside MSc, while High School does not reach significance. Religion and marital status replicate the previous findings: non-religious respondents report lower usage intention, and single individuals score lower than those who were married or partnered, though divorce was not significant. SES remains the most impactful fixed effect, with the highest group (10) again showing the largest positive effect. The random effects tell a similar story: the conditional ICC is 0.050, indicating that only 5.0% of residual variance is between countries, and the likelihood ratio test again confirms that the country-level effect is significant ($\chi^2(1) = 294.77$, $p < .001$). Country rankings are nearly identical, with Anglophone countries (UK, USA) showing positive intercepts and the Netherlands showing the most negative, reinforcing the cross-construct consistency of cultural differences.

The perceived benefits results (Table 9) largely reinforce the patterns observed for trust and usage. Older adults (55+) again report significantly lower scores. Gender is non-significant, consistent with usage. Religion maintains the same direction across all three constructs: non-religious respondents report lower perceived benefits compared to Christians, while Muslims and other religions do not differ from the reference group. Being single remains a robust negative predictor across all constructs, underscoring the consistent association between marital status and chatbot attitudes. SES remains the most impactful fixed effect: the highest group (10) has the largest positive effect, and groups 9 and above are consistently significant across trust, usage intention, and perceived benefits, confirming that higher socioeconomic status is reliably associated with more favorable chatbot perceptions. The country-level random effects tell a familiar story: the conditional ICC is 0.042, and the likelihood ratio test confirms that between-country variance remains significant ($\chi^2(1) = 330.41$, $p < .001$). The country ranking is virtually identical to trust and usage intention, with the UK and the USA showing positive intercepts and the Netherlands the most negative, further reinforcing the robustness of Anglophone-Continental European attitudinal differences across constructs. Notably, education is the one domain where Perceived Benefits diverged, with no significant effects.

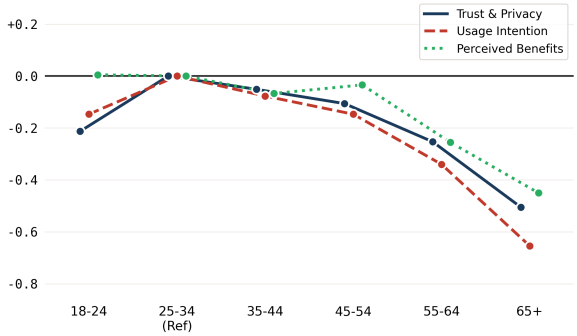


Figure 12: Older generations show lower Trust, Perceived Benefits, and Usage Intention. Fixed effect estimates by age group (Ref = 25-34 years).

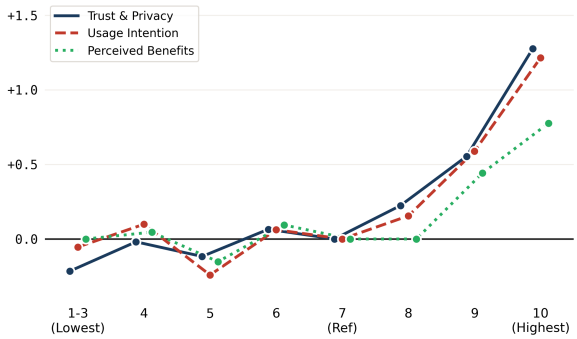


Figure 13: Higher SES drives Trust, Perceived Benefits, and Usage Intention. Fixed effect estimates by Socioeconomic Status (Ref = SES 7).

	Estimate	SE	z	p
Fixed Effects				
Gender (Ref: Male)				
Female	**-0.117**	**0.034**	**-3.403**	**< .001**
Age (Ref: 25-34 years)				
18-24 years	**-0.213**	**0.073**	**-2.925**	**.003**
35-44 years	-0.051	0.046	-1.101	.271
45-54 years	**-0.106**	**0.053**	**-1.994**	**.046**
55-64 years	**-0.253**	**0.066**	**-3.818**	**< .001**
65+ years	**-0.505**	**0.084**	**-6.016**	**< .001**
Education (Ref: BSc)				
HS	**0.144**	**0.055**	**2.616**	**.009**
AD	0.056	0.066	0.841	.400
Some Col.	-0.103	0.078	-1.315	.188
Prof. Deg.	-0.067	0.060	-1.130	.259
MSc	**-0.120**	**0.053**	**-2.271**	**.023**
PhD	0.048	0.083	0.578	.563
Religion (Ref: Christianity)				
None	**-0.260**	**0.041**	**-6.303**	**< .001**
Islam	0.135	0.073	1.842	.066
Other	0.062	0.090	0.686	.493
Marital status (Ref: Married/Partnership)				
Single	**-0.321**	**0.043**	**-7.509**	**< .001**
Divorced	**-0.241**	**0.081**	**-2.962**	**.003**
SES (Ref: 7)				
1–3 (Lowest)	**-0.215**	**0.088**	**-2.442**	**.015**
4	-0.019	0.070	-0.277	.782
5	-0.117	0.060	-1.949	.051
6	0.065	0.054	1.198	.231
8	**0.224**	**0.053**	**4.254**	**< .001**
9	**0.554**	**0.072**	**7.657**	**< .001**
10 (Best off)	**1.275**	**0.084**	**15.239**	**< .001**
Random Effects				
Country	$\sigma^2 = 0.214$, SD $= 0.463$			
Netherlands	-0.542			
Italy	-0.276			
Spain	-0.259			
Germany	-0.208			
France	-0.105			
United States	0.581			
United Kingdom	0.809			

HS = High School, AD = Associate Degree, Some Col. = Some college, no degree, Prof. Deg. = Professional Degree, None = No religion.

Table 7: Cumulative Link Mixed Model Results for Trust and Privacy.

	Estimate	SE	z	p
Fixed Effects				
Gender (Ref: Male)				
Female	-0.036	0.041	-0.879	.380
Age (Ref: 25-34 years)				
18-24 years	-0.147	0.086	-1.710	.087
35-44 years	-0.077	0.054	-1.421	.155
45-54 years	**-0.147**	**0.063**	**-2.330**	**.020**
55-64 years	**-0.340**	**0.079**	**-4.280**	**< .001**
65+ years	**-0.654**	**0.102**	**-6.394**	**< .001**
Education (Ref: BSc)				
HS	0.099	0.066	1.498	.134
AD	0.015	0.079	0.191	.849
Some Col.	**-0.281**	**0.094**	**-2.989**	**.003**
Prof. Deg.	0.031	0.071	0.434	.664
MSc	**-0.194**	**0.062**	**-3.133**	**.002**
PhD	0.181	0.097	1.860	.063
Religion (Ref: Christianity)				
None	**-0.282**	**0.049**	**-5.727**	**< .001**
Islam	0.092	0.086	1.067	.286
Other	-0.012	0.106	-0.116	.907
Marital status (Ref: Married/Partnership)				
Single	**-0.262**	**0.051**	**-5.134**	**< .001**
Divorced	-0.140	0.098	-1.424	.154
SES (Ref: 7)				
1–3 (Lowest)	-0.054	0.106	-0.509	.611
4	0.099	0.084	1.193	.233
5	**-0.241**	**0.071**	**-3.411**	**< .001**
6	0.062	0.064	0.976	.329
8	**0.155**	**0.062**	**2.478**	**.013**
9	**0.588**	**0.085**	**6.877**	**< .001**
10 (Best off)	**1.215**	**0.097**	**12.477**	**< .001**
Random Effects				
Country	$\sigma^2 = 0.173$, SD $= 0.416$			
Netherlands	-0.540			
Italy	-0.346			
Spain	-0.204			
Germany	-0.096			
France	-0.024			
United Kingdom	0.585			
United States	0.624			

HS = High School; AD = Associate Degree; Some Col. = Some college, no degree; Prof. Deg. = Professional Degree; None = No religion.

Table 8: Cumulative Link Mixed Model Results for Usage Intention.

	Estimate	SE	z	p
Fixed Effects				
Gender (Ref: Male)				
Female	0.032	0.035	0.913	.361
Age (Ref: 25-34 years)				
18-24 years	0.005	0.073	0.075	.941
35-44 years	-0.067	0.047	-1.440	.150
45-54 years	-0.034	0.054	-0.636	.525
55-64 years	**-0.255**	**0.067**	**-3.790**	**$< .001$**
65+ years	**-0.450**	**0.086**	**-5.228**	**$< .001$**
Education (Ref: BSc)				
HS	0.017	0.056	0.299	.765
AD	-0.027	0.067	-0.402	.688
Some Col.	-0.025	0.080	-0.309	.757
Prof. Deg.	-0.093	0.061	-1.530	.126
MSc	-0.066	0.053	-1.243	.214
PhD	0.138	0.082	1.685	.092
Religion (Ref: Christianity)				
None	**-0.114**	**0.042**	**-2.717**	**.007**
Islam	-0.037	0.072	-0.514	.607
Other	-0.030	0.090	-0.338	.735
Marital status (Ref: Married/Partnership)				
Single	**-0.241**	**0.044**	**-5.542**	**$< .001$**
Divorced	0.009	0.083	0.111	.912
SES (Ref: 7)				
1–3 (Lowest)	0.000	0.090	0.002	.998
4	0.046	0.071	0.649	.516
5	**-0.152**	**0.061**	**-2.504**	**.012**
6	0.094	0.055	1.722	.085
8	0.000	0.053	-0.005	.996
9	**0.442**	**0.073**	**6.090**	**$< .001$**
10 (Best off)	**0.776**	**0.082**	**9.486**	**$< .001$**
Random Effects				
Country	$\sigma^2 = 0.143$, SD $= 0.379$			
Netherlands	-0.500			
Germany	-0.243			
Italy	-0.203			
Spain	-0.116			
France	-0.029			
United States	0.422			
United Kingdom	0.668			

HS = High School; AD = Associate Degree; Some Col. = Some college, no degree; Prof. Deg. = Professional Degree; None = No religion.

Table 9: Cumulative Link Mixed Model Results for Perceived Benefits.

F Prompt Samples and Country Distributions

Table 10 shows examples of shared prompts representing the following five key topics: general stress and anxiety, relationship conflict, family and caregiving, loneliness, and trauma and mental health.

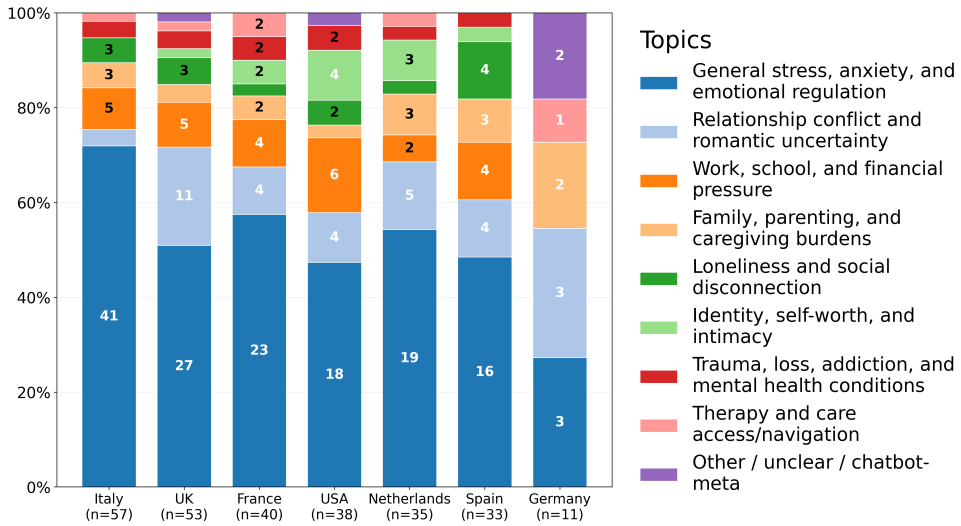


Figure 14: Topic distribution of shared prompts related to mental wellbeing across countries.

Topic	UK	USA	Netherlands
General stress & anxiety	"I'm feeling down today and need some motivational advice or coping strategies to improve my mood."	"I'm feeling stressed and overwhelmed, can you give me some tips to relax and feel more in control?"	"Welke tips en trucs zijn er om goed om te gaan met stress?" [What tips are there to deal with stress?]
Relationship conflict	"How to handle arguments at home."	"How to react to the guy I am dating not texting me for a couple of days?"	"Hoe omgaan met wispelturige partner." [How to deal with an unpredictable partner.]
Family & caregiving	"Can you help me with ways I can handle a tiring sibling?"	"What to do when a child upsets me."	"Hoe moet ik omgaan met het feit dat mijn dochters geen contact willen hebben met mij?" [How do I cope with my daughters cutting contact with me?]
Loneliness	"Most times I feel like I don't belong. Like people always feel off. What could possibly be wrong?"	"How to deal with stress and anxiety alone?"	"Wat te doen tegen eenzaam voelen." [What to do about feeling lonely.]
Trauma & mental health	"How to manage stress and depression and overcome an abusive partner."	"How to deal with a terminal diagnosis."	"Verwerking rouw dood opa." [Processing grief after my grandfather's death.]

Table 10: Representative shared user prompts across five themes for the UK (country with highest TRU, BENF, and USE construct scores), USA (second highest), and the Netherlands (lowest). Dutch prompts are shown in the original language with translations in brackets. Prompts lightly edited for length and anonymization.

G Questionnaire

Construct	Questionnaire Item	Source
Trust		
	I trust the information or support this chatbot provides.	(Marimon et al., 2024)
	The chatbot provides credible information.	(Marimon et al., 2024)
	I trust that this chatbot is honest and transparent about what it can and cannot do.	Authors
	Overall, I trust the chatbot.	(Schmidmaier et al., 2024)
Privacy		
	The chatbot clearly communicates its limitations in providing emotional support.	Authors
	I trust the chatbot to handle sensitive information safely.	(Marimon et al., 2024)
	I feel confident that my interactions with the chatbot are private.	(Marimon et al., 2024)
	Overall, I believe this chatbot is safe and respectful of my privacy.	Authors
	I feel safe sharing my emotions with the chatbot.	(Marimon et al., 2024)
	The chatbot provides advice on when to seek professional assistance or personal support from others.	Authors
Perceived Benefits		
	Clarity & Self-Reflection: The chatbot helps me feel more confident in managing emotional difficulties.	Authors
	Challenge & New Perspectives: The chatbot challenges my thinking in a constructive way. The chatbot helps me consider alternative viewpoints on my concerns.	Authors
	Recommendations & Advice: The chatbot provides recommendations or advice that are useful for my specific situations.	Authors
	Actionability: The chatbot helps me identify actionable steps I can take to improve my mental wellbeing.	Authors
	Accessibility: The chatbot's availability (e.g., 24/7 access) is a significant benefit for my wellbeing support.	Authors
	Non-judgmental Exploration: I feel that I can express my feelings openly without fear of judgment.	Authors
	Ability: Using this chatbot has improved my ability to manage my wellbeing concerns.	Authors
	Cost: The fact that this chatbot is free (or less expensive than traditional therapy) is a major benefit for me.	Authors
	General Usefulness: Overall, I find the chatbot to be a valuable tool for supporting my mental wellbeing. After using the chatbot, I feel calmer or more relaxed.	Authors
Attitude Toward Using		
	Using the chatbot for emotional or mental wellbeing support is a good idea.	(Davis, 1989)
	The chatbot helps me feel emotionally understood.	(Davis, 1989)
	I feel motivated to continue using the chatbot regularly.	Authors
	Overall, my attitude toward using the chatbot is favorable.	(Davis, 1989)
Actual System Use		
	I use the chatbot frequently for wellbeing/emotional support.	(Davis, 1989)
	My interactions with the chatbot for mental wellbeing or emotional support are a regular part of my routine.	(Davis, 1989)
	When I need mental wellbeing or emotional support, the chatbot is usually one of the first resources I turn to.	(Davis, 1989)

Table 11: Survey Constructs, Items, and Sources. Authors: Items proposed by authors.