Personalizing Human-LLM Interactions through Mixed Profiling

Elias Müller*
human-centered systems lab (h-lab)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
elias.mueller@kit.edu

Kirsten Greiner*
human-centered systems lab (h-lab)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
kirsten.greiner@kit.edu

Adrian Wegener
Karlsruhe Institute of Technology
Karlsruhe, Germany
adrian.wegener@kit.edu

Florian Onur Kuhlmeier
human-centered systems lab (h-lab)
Karlsruhe Institute of Technology
Karlsruhe, Germany
florian.kuhlmeier@kit.edu

Alexander Maedche
human-centered systems lab (h-lab)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
alexander.maedche@kit.edu

Mixed Profiling

... is a user modeling approach that deliberately combines **quantitative psychometric profiling**, based on standardized and stable measurements of user characteristics, with **qualitative profiling**, which captures interpretive understandings of user context.

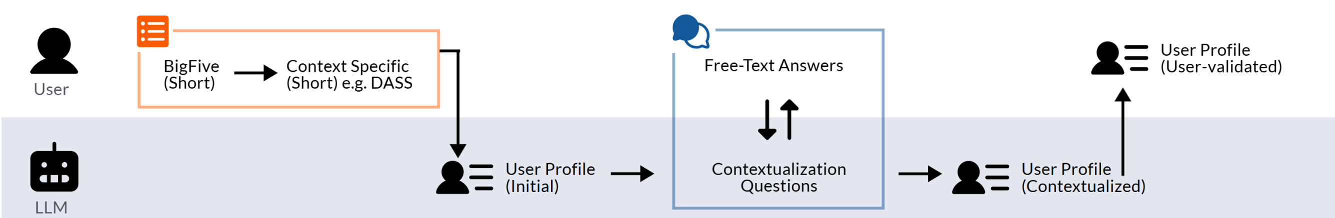

Figure 1: Overview of the *Mixed Profiling* approach

Abstract

While Large Language Models (LLMs) are transitioning into the role of companions, human-LLM interaction remains largely indifferent to the nuanced individualities of the human user. Most current personalization of LLMs relies on behavioral data, while established psychometric methods, such as questionnaires that assess individual traits and states, are rarely leveraged. We propose *Mixed Profiling*, a novel approach that combines standardized short questionnaires with adaptive LLM dialogs to create precise user profiles. We implemented and evaluated the *Mixed Profiling* approach for the mental health context as an example in an online study (N=40). Participants completed questionnaires and participated in an LLM-based chat. They rated the *Mixed Profiling* approach's results as significantly more trustworthy than those of a basic questionnaire-based profile. By bridging explicit psychometrics and data-driven personalization, our work lays the foundation for a more psychologically grounded and transparent personalization in human-LLM interaction.

CCS Concepts

• **Human-centered computing** → **User models**; **User studies**; **Empirical studies in HCI**.

Keywords

Personalization, Profiling, LLM, Behavior Change

*Both authors contributed equally to this research.

CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3799351

ACM Reference Format:
Elias Müller, Kirsten Greiner, Adrian Wegener, Florian Onur Kuhlmeier, and Alexander Maedche. 2026. Personalizing Human-LLM Interactions through Mixed Profiling. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3772363.3799351

1 Introduction

Large Language Models (LLMs) have rapidly evolved from research prototypes to everyday companions, supporting users in domains such as productivity [16], education [36], and mental health [9, 28, 29]. However, while these systems engage millions of users [3], personalization in most LLM interactions relies solely on behavioral data such as chat history or query frequency [17, 33]. Users' individual traits and state characteristics are largely neglected. This stands in contrast to decades of psychological research, which offers validated and reliable methods for capturing individual traits and states, such as personality, depression severity, or mood [10].

Personalization can be grounded in explicit data, as chat-based user inputs represent deliberate self-expression rather than implicitly observed behavioral data [22]. However, by combining established psychometric approaches with personal user input, we argue that personalized user profiles can be more firmly grounded in established psychological foundations.

As the field advances toward more personalized LLM-based interactions, accurately capturing a user's psychological profile without disrupting conversational flow remains a critical challenge [13]. Researchers must therefore develop lightweight profiling methods that balance depth, user burden, and ethical transparency [15]. Traditional psychological questionnaires provide comprehensive insights but are often too static or intrusive for real-world interaction [32]. Conversely, LLM-based profiling purely based on chat content may produce inaccurate or biased profiles due to limited or noisy input. Consequently, developing approaches that integrate rigorous psychometric assessment with seamless, chat-based interactions presents an important research priority. We followed the research question: *"How can a mixed profiling approach combining standardized psychometric questionnaires with open-ended chat be operationalized to create user profiles for personalized LLM interactions?"*

To address this question, we propose a novel *Mixed Profiling* approach that combines standardized, context-specific short-form questionnaires with personalized open-ended LLM-based dialogue. To evaluate our approach, we chose the field of mental health as the context. Accurate user profiling is crucial for personalization in this field, as tailored interventions based on individual psychological characteristics significantly improve therapy outcomes [20]. We aim to refine the user profile with personal information in a natural way and get more precise insights beyond the questionnaire answers.

With this work, we contribute to the **HCI community**:

(1) **conceptually**, by introducing and defining *Mixed Profiling* as a novel approach for constructing personalized user profiles that combine standardized psychometric questionnaires with open-ended personal user input;
(2) **practically**, by providing an open source system implementation, including a structured prompt framework for LLM interactions; and
(3) **empirically**, by providing insights into how users perceive and evaluate the *Mixed Profiling* process and the resulting profiles compared to approaches without open-ended user input.

2 Related Work

User Modeling and Profiling in HCI: User modeling has a long tradition in HCI, where models of users are employed to improve usability, relevance, and contextual responsiveness [12]. Early approaches primarily represented user preferences, goals, and behaviors, often assuming relatively stable user characteristics [12]. More recent work, including guidelines for Human–AI Interaction, emphasizes that user models should evolve over time to reflect changes in user behavior and system use [1]. However, evaluating the effectiveness of such continuously adapting models remains challenging, as user models co-evolve with user acclimation and increasing system familiarity [5]. In practice, many contemporary systems rely on behavioral traces to infer user characteristics implicitly. For example, adaptive user interfaces infer personality traits or preferences from social media activity [8, 31] or phone usage logs [31]. While these approaches scale well, they risk inaccuracies and trust issues [30]. At the same time, users appear willing to share personal and sensitive data, including emotional states [35]. Hybrid user modeling approaches combine implicit unstructured behavioral data with explicit user input, such as questionnaires or preference settings [4, 24]. These approaches acknowledge the limitations of purely implicit inference [30] while retaining scalability, but need some time to acquire initial user data [26]. However, explicit and implicit inputs are often treated as interchangeable signals rather than as qualitatively different forms of knowledge about the user [4, 24]. Moreover, reliable explicit data collection frequently relies on static or intrusive questionnaires that can disrupt interaction and reduce engagement [11]. Prior work suggests that conversational or qualitative questionnaires can be more engaging than static forms [11], yet such approaches are rarely combined with standardized psychometric measures in adaptive, real-time user modeling. Our proposed *Mixed Profiling* closes these research gaps in hybrid user modeling by leveraging the advantages of explicit approaches while addressing the limitations of static and item-based, intrusive, and user-context-free modeling methods.

Personalization of Chatbots: Despite extensive research on user modeling, many chatbots still employ largely "one-size-fits-all" interaction strategies [30]. This lack of personalization is particularly problematic in (behavioral) change domains such as mental health [23] and learning [14], where alignment with individual needs is critical. Prior work highlights the importance of balancing user control over personalization with system transparency [5], as both excessive opacity and excessive configuration complexity can undermine autonomy and trust. Recent systems have explored more explicit personalization options, allowing users to adjust chatbot personas, shared user profiles, visual representations, modalities, or technical parameters such as model temperature [7, 35]. While such flexibility can increase engagement and foster open interaction [35], it often places substantial effort on users or designers to define appropriate settings. As a result, alternative approaches such as community-based personalization have been proposed to reduce individual burden [35]. Conversely, limited opportunities for user input can lead to reduced trust and perceived loss of autonomy [5]. Empirical studies indicate that users can successfully personalize chatbot behavior without expert knowledge [6], and that customized chatbot personas can increase engagement [6]. Other work shows that grounding personalization in concrete examples or contextual information can improve conceptual alignment between users and systems [34]. Nevertheless, how personalization strategies evolve over longer-term use remains insufficiently understood [35]. Across this literature, personalization is typically framed either as user-driven configuration or as system-driven inference from interaction data. No approaches integrate validated psychological constructs into conversational personalization in a way that remains lightweight, transparent, and adaptive.

Contribution in Relation to Related Work: Prior research on user modeling and chatbot personalization lacks a framework that unites quantitative, psychometrically grounded user profiles with qualitative, conversational refinement. Our *Mixed Profiling* approach responds to this need by offering a scalable and transparent path toward psychologically grounded personalization in human–AI interaction.

3 Introducing *Mixed Profiling*: Positioning, Process and Implementation

Positioning: In Figure 1, we introduced *Mixed Profiling* as a user modeling approach that deliberately combines quantitative psychometric profiling, based on standardized and stable measurements of user characteristics, with qualitative profiling, which captures interpretive understandings of user context. Rather than treating these approaches as interchangeable or additive, *Mixed Profiling* positions quantitative measures as anchors whose meaning and use are shaped through evolving qualitative context. Table 1 summarizes how *Mixed Profiling* is positioned in relation to quantitative and qualitative profiling.

Process: Our *Mixed Profiling* approach for personalizing human-LLM interaction follows the process shown in Figure 1. We instantiate the process using a specific example in the context of building a user profile for a mental health chatbot. Specifically, we want to capture personality using the Big Five Inventory (BFI) questionnaire[25] as an individual trait and depression severity using the Depression, Anxiety and Stress Scale (DASS) questionnaire [19] as an individual state. The corresponding process looks as follows:

(1) Let the user fill out an individual trait questionnaire: In this first step, we capture personality as an important individual trait for human-LLM interaction in mental health using the established 10-item BFI questionnaire [25].
(2) Let the user fill out an individual state questionnaire: In the second step, we capture an individual state as input for the LLM interaction. In our example, we chose the 21-item DASS questionnaire [19].
(3) Based on the user responses in the questionnaires, an initial user profile is created by the LLM in the role of a psychologist and therapist with the task to generate a user profile for a mental health chatbot. All questions with answers from the questionnaires are embedded. The profile should include, for each dimension of the questionnaires, a paragraph; in our example, 8 (5 from Big5 and 3 from DASS-21).
(4) Let the user perform an LLM-based dialogue for profile refinement: For this purpose, the created profile is added as a system prompt, and the task is to conduct an interaction with the LLM. The LLM will ask questions about the dimensions that need clarification. It works through the dimensions one by one, without getting stuck on any one dimension if several need clarification.
(5) Profile contextualization: To this end, the LLM receives the profile from the questionnaires and the dialogue, including all questions and answers. Furthermore, the profile is to be expanded with the new information. In the case of an information conflict, the LLM was instructed to prefer the information from the dialogue. The existing structure and dimensions remain unchanged.

Technical Implementation: We leveraged Otree [2] to implement the system. For LLM integration, we built on the GitHub repository of "Otree GPT." [1] We used GPT-OSS [21] hosted on servers at the research facility as the LLM. We prompted the system as following: (1) role description, in our case a mixture of psychologist and therapist, (2) then the specific task, such as creating a profile or reconciling a profile with follow-up questions, (3) the information, such as questionnaire answers, existing profile, or questions and answers, (4) a brief description of the procedure and thoughts that the LLM should do step by step, and (5) what the output should look like and limitations for the output, such as the profile should contain all dimensions of the questionnaires, or messages should end with a question. The code of the artefact is available on GitHub.

4 Evaluation Method

The study uses an online experiment to evaluate the proposed *Mixed Profiling* approach in a controlled setting as a proof-of-concept. Two different measurements were chosen for the evaluation: (1) Participants evaluated the profile before and after the LLM dialogue according one User Experience Questionnaire (UEQ) criteria (trustworthiness of content) [27] to assess the profiles individually, and

[1] https://github.com/clintmckenna/oTree_gpt [18]

Table 1: Positioning *Mixed Profiling* in relation to quantitative and qualitative profiling.

	Quantitative Profiling	Qualitative Profiling	Mixed Profiling
Epistemology	Measurements	Interpretation	Measured anchors, interpreted context
Knowledge Orientation	Nomothetic (Population)	Idiographic (Individual)	Nomothetic, idiographic enriched
Temporalities	More sparse, stable	More contextual, dynamic	Stable core and evolving context
Adaption Level	Coarse, predefined	Fine-grained, flexible	Layered, some flexibility
Data Structure	High structure, standardised	Low structure, latent	Standardised and latent data
Legibility	System-facing	User-facing	Shared, combined, and user validated

(2) participants were able to manually refine the contents of the profile after the LLM interview to observe qualitatively what had been adjusted. The study was approved by our institutional ethics review board.

Study Procedure: For this purpose, the study was divided into two parts. The participants followed the process described above. At the beginning, the short forms of the BFI and DASS were administered, and the order of the questionnaires and questions was randomized for each participant. After the profile was created in the background, a chat dialogue opened, and eight questions were asked. After the new profile was created, the participants evaluated both profiles, with the order randomized. The participants rated all 8 dimensions on a 4-item scale, the trustworthiness of content criteria of UEQ: useful, plausible, trustworthy, and accurate. Furthermore, each dimension was also explained so that the participants were supported in understanding the terminology. After both profiles were evaluated, participants could refine and validate all 8 profile dimensions following the LLM interaction. For this purpose, a text field was pre-filled with the text for each dimension.

Participants: We recruited 40 participants via Prolific. The median study time was 25.5 min and participants received 6 £. Participants were recruited from Europe only, and fluency in English was required. 23 identified as women and 17 as men. The average age was 33.45 with a standard deviation of 10.5.

5 Results

Quantitative Evaluation of Profile Evaluation Assessment: The four items useful, plausible, trustworthy, and accurate of the UEQ were used to evaluate each dimension of the two profiles. The UEQ (scale from −3 to 3) was 1.97 ($SD = 1.01$) for dialogues and 1.65 ($SD = 1.22$) for the questionnaire. The main analysis was conducted on ratings averaged across items using a linear mixed-effects model with method and dimension as fixed effects and participant as a random effect. The model revealed a significant main effect of dialogue vs. questionnaire-based profile, with dialogue-based profiles receiving higher trustworthiness ratings than questionnaire-based profiles ($\beta = 0.32$ and $p = 0.001$). This effect was consistent across all eight questionnaire dimensions. To assess robustness, we fitted an additional mixed-effects model at the item level, including item as a fixed effect. Results were consistent with the aggregated analysis, indicating that the observed difference was not driven by any specific item. Because the interaction between method and dimension did not improve model fit, results are reported from the more parsimonious additive model. No significant differences were found for 3-ways interaction model, including item level, but the comparison between the two profile versions at dimension and item level is, on average, higher for mixed profiles in all comparisons, by an average of 0.32.

Time of Interaction for the steps: For the BFI, participants need a median of 39.24 s (mean: 59.19 s, SD: 65.48 s). The median time taken to complete the DASS is 76.25 s (mean: 90.78 s, SD: 43.58 s). The median time taken to complete the chat section is 522.14 s (mean: 573.29 s, SD: 228.73 s). The evaluation of questionnaire-based profil with UEQ has a median completion time of 208.07 s (mean: 227.17 s, SD: 116.18 s). The median duration of the evaluation of the profile based on the questionnaire profile and LLM dialogue is 206.45 s (mean: 217.52 s, SD: 121.91 s). The median duration of the refinement step is 106.26 s (mean: 138.84 s, SD: 134.99 s).

Profile Refinements: A total of 20 users made refinements, making between 1 and 5 changes over all dimensions. In terms of dimensions, extraversion and depression were adjusted most frequently, while stress was not adjusted at all. The refinements relate to inserting spaces (30 times), adding text (19 times), deleting text (12 times), and changing text (9 times). The most common refinement was adding examples of statements, such as *"Also some alone time walking works."* or *"Yet the user has a tendency to continuously think about the project or work until finished correctly."* The second most common refinement was adjusting the emphasis of statements in the profile, such as changing *"occasional"* to *"very mild"* or *"[modest to] moderate"* to *"[modest to] high."* The third most common refinement was deleting concrete examples mentioned in the profile, such as *"racing heart"* or *"paced breathing."*

6 Discussion

We propose a *Mixed Profiling* approach for LLM-based interaction. This method builds on and combines existing user modeling approaches, enabling us to use standardized, well-founded questionnaires with open-ended questions to gain more personalized insights, thereby increasing trustworthiness. Existing approaches have obtained this information implicitly from chat history, but need time to gain user data. However, this time is relevant in various contexts and can increase the dropout rate. We argue that our approach could increase engagement, as a faster, more accurate profile can enable targeted LLM-based interactions.

We implemented the proposed *Mixed Profiling* method and conducted an online study to evaluate it. Thereby, two different user profiles were generated. The first uses information from the questionnaires, and the second extends the first with our *Mixed Profiling* method, including the LLM dialogue. Our results indicate that profiles generated using the *Mixed Profiling* approach exhibit significantly greater trustworthiness across all dimensions of BFI and DASS. Trustworthiness compromises the items' usefulness, plausibility, trustworthiness, and accuracy. For analyzing the interaction at the item level, the number of participants was too low. Nevertheless, the descriptive analysis showed that participants, on average, rated the mixed profile higher across all dimensions and items. These findings provide initial validation for *Mixed Profiling* as a trustworthy user profile approach for LLM-base interaction. However, the ratings are based solely on users' subjective views, not on experts' views, such as those of therapists. Half of the participants chose to further refine certain dimensions of their profiles without any additional incentive. Moreover, the average time spent on the refinement page indicates that participants meaningfully engaged with the provided content, although the high standard deviation points to variability in engagement levels. Participants refined their profile in several ways, including adding examples, removing specific examples by the LLM, or adjusting points of content. This suggests that *Mixed Profiling* may foster a sense of involvement between users and the system.

LLMs carry the risk of hallucinating, which is a risk that needs to be taken into account, especially in the context of mental health. In our study, the profiles could contain false information. However, in our investigation of refinements made by the users, we did not find that participants made changes due to hallucinations, but further investigation could be conducted to determine the extent to which participants experienced hallucinations.

Overall, we believe that the *Mixed Profiling* approach represents an innovative and, at the same time, practical process for user profiling. In addition, we argue that our findings from the online study are transferable to the real world, as we motivated our test subjects extrinsically, but users in the real world have an interest in making the process as good as possible to personalize their LLM-based interaction and enhance the trustworthiness when using the profile.

This work makes several contributions to the HCI community. First, it *conceptualizes and introduces Mixed Profiling* as a method that combines classic user modeling via item-based questionnaires with contemporary LLM-based dialogues to build a more personalized user profile. Second, we implemented an artefact including a structure for an LLM-Prompt used in the proposed method. Finally, we contribute empirical findings from an online study using this artefact.

This work comes with the following limitations. First, the approach depends on our prompt design choices. Different prompt formulations may lead to different outcomes. Secondly, the evaluation study used a small European sample and no information on prior experience, limiting the generalizability of the findings. Thirdly, a within-subjects design was chosen for the proof-of-concept evaluation; however, to gain greater, more specific insight, more treatments would need to be evaluated in a between-subjects design. Finally, participants were recruited via Prolific and were motivated via fixed incentives. The findings are not necessarily transferable to the real world, where users may have different motivations and interact differently.

7 Future Work

Building on the study's results, we outline directions for future research. First, a systematic comparison across different dialogue strategies and conditions, such as Chat-only, in a between-subjects study design could provide insights into their effects on the profile's satisfaction and trustworthiness, and the between-subjects design could also eliminate various order effects. Second, remapping the final user profile to the original quantitative measures could provide insights into information retention and identify opportunities to improve it. Third, field studies, combined with the application of the profile, are needed to investigate how *Mixed Profiling* performs, compared to other strategies, in the real world. In particular, it would be interesting to observe whether such a profile not only personalize the interaction but also reduces the risk of hallucinations, as the profile provides more guidelines for the LLM. Finally, future research should examine whether the method is applicable in other contexts, such as learning styles or emotion regulation. **Overall, *Mixed Profiling* invites psychologically informed personalization across a range of HCI applications.**

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) -GRK2739/2-Project Nr. 447089431 - Research Training Group: KD2School - Designing Biosignal-Adaptive Systems for Decision-Making Processes.

References

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. doi:10.1145/3290605.3300233

[2] Daniel L. Chen, Martin Schonger, and Chris Wickens. 2016. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9 (March 2016), 88–97. doi:10.1016/j.jbef.2015.12.001

[3] Filippo Chiarello, Vito Giordano, Irene Spada, Simone Barandoni, and Gualtiero Fantoni. 2024. Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation* 133 (May 2024), 103002. doi:10.1016/j.technovation.2024.103002

[4] Marina Farid, Rania Elgohary, Ibrahim Moawad, and Mohamed Roushdy. 2018. User Profiling Approaches, Modeling, and Personalization. *SSRN Electronic Journal* (2018). doi:10.2139/ssrn.3389811

[5] Libby Ferland and Wilma Koutstaal. 2020. How's Your Day Look? The (Un)Expected Sociolinguistic Effects of User Modeling in a Conversational Agent. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–8. doi:10.1145/3334480.3375227

[6] Juhye Ha, Hyeon Jeon, Daeun Han, Jinwook Seo, and Changhoon Oh. 2024. CloChat: Understanding How People Customize, Interact, and Experience Personas in Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–24. doi:10.1145/3613904.3642472

[7] Xu Han, Michelle Zhou, Matthew J. Turner, and Tom Yeh. 2021. Designing Effective Interview Chatbots: Automatic Chatbot Profiling and Design Suggestion Generation for Chatbot Debugging. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. doi:10.1145/3411764.3445569

[8] Andrea L. Hartzler, Bridget Weis, Carly Cahill, Wanda Pratt, Albert Park, Uba Backonja, and David W. McDonald. 2016. Design and Usability of Interactive User Profiles for Online Health Communities. *ACM Transactions on Computer-Human Interaction* 23, 3 (July 2016), 1–33. doi:10.1145/2903718

[9] Michael V. Heinz, Daniel M. Mackin, Brianna M. Trudeau, Sukanya Bhattacharya, Yinzhou Wang, Haley A. Banta, Abi D. Jewett, Abigail J. Salzhauer, Tess Z. Griffin, and Nicholas C. Jacobson. 2025. Randomized Trial of a Generative AI Chatbot for Mental Health Treatment. *NEJM AI* 2, 4 (March 2025). doi:10.1056/AIoa2400802

[10] Fabian Hutmacher and Markus Appel. 2023. The Psychology of Personalization in Digital Environments: From Motivation to Well-Being – A Theoretical Integration. *Review of General Psychology* 27, 1 (March 2023), 26–40. doi:10.1177/10892680221105663

[11] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. doi:10.1145/3290605.3300316

[12] Alfred Kobsa. 2001. Generic User Modeling Systems. *User Modeling and User-Adapted Interaction* 11, 1-2 (March 2001), 49–63. doi:10.1023/A:1011187500863

[13] Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The Personalization of Conversational Agents in Health Care: Systematic Review. *Journal of Medical Internet Research* 21, 11 (Nov. 2019), e15360. doi:10.2196/15360

[14] Joanne Leong, Pat Pataranutaporn, Valdemar Danry, Florian Perteneder, Yaoli Mao, and Pattie Maes. 2024. Putting Things into Context: Generative AI-Enabled Context Personalization for Vocabulary Learning Improves Learning Motivation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–15. doi:10.1145/3613904.3642393

[15] Jiahao Liu, Xueshuo Yan, Dongsheng Li, Guangping Zhang, Hansu Gu, Peng Zhang, Tun Lu, Li Shang, and Ning Gu. 2025. Improving LLM-powered Recommendations with Personalized Information. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Padua Italy, 2560–2565. doi:10.1145/3726302.3730211

[16] Reuben Luera, Ryan Rossi, Franck Dernoncourt, Alexa Siu, Sungchul Kim, Tong Yu, Ruiyi Zhang, Xiang Chen, Nedim Lipka, Zhehao Zhang, Seon Gyeom Kim, and Tak Yeon Lee. 2025. Personalizing Data Delivery: Investigating User Characteristics and Enhancing LLM Predictions. In *Companion Proceedings of the ACM on Web Conference 2025*. ACM, Sydney NSW Australia, 1167–1171. doi:10.1145/3701716.3715452

[17] Simon Lupart, Daniël Van Dijk, Eric Langezaal, Ian Van Dort, and Mohammad Aliannejadi. 2025. Investigating LLM Variability in Personalized Conversational

Information Retrieval. In *Proceedings of the 2025 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*. ACM, Xi'an China, 353–363. doi:10.1145/3767695.3769502

[18] Clint McKenna. 2023. oTree GPT. https://github.com/clintmckenna/oTree_gpt original-date: 2023-03-29T09:09:24Z.

[19] P. Nilges and C. Essau. 2021. DASS. Depressions-Angst-Stress-Skalen - deutschsprachige Kurzfassung. doi:10.23668/PSYCHARCHIVES.4579

[20] Arthur Nye, Jaime Delgadillo, and Michael Barkham. 2023. Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology* 91, 7 (July 2023), 389–397. doi:10.1037/ccp0000820

[21] OpenAI. 2025. gpt-oss-120b & gpt-oss-20b Model Card. https://arxiv.org/abs/2508.10925 _eprint: 2508.10925.

[22] Yingzhe Peng, Xiaoting Qin, Zhiyang Zhang, Jue Zhang, Qingwei Lin, Xu Yang, Dongmei Zhang, Saravan Rajmohan, and Qi Zhang. 2025. Navigating the Unknown: A Chat-Based Collaborative Interface for Personalized Exploratory Tasks. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. ACM, Cagliari Italy, 1048–1063. doi:10.1145/3708359.3712093

[23] Adam Poulsen, Ian B Hickie, Zsofi de Haan, William Capon, Haley M LaMonica, Ebenezer Eyeson-Annan, Jalal Radwan, Nia Barnabie, Elizabeth M Scott, and Frank Iorfino. 2025. Co-designing a conversational generative artificial intelligence system for youth mental health. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3706599.3719696

[24] Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. 2024. User Modeling and User Profiling: A Comprehensive Survey. doi:10.48550/arXiv.2402.09660

[25] Beatrice Rammstedt, Christoph J. Kemper, Mira Céline Klein, Constanze Beierlein, and Anastassiya Kovaleva. 2013. A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10). *methods* data (2013), 17 Pages. doi:10.12758/MDA.2013.013 Artwork Size: 17 Pages Medium: application/pdf.

[26] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. 2002. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '02)*. Association for Computing Machinery, New York, NY, USA, 253–260. doi:10.1145/564376.564421

[27] Martin Schrepp and Jörg Thomaschewski. 2019. Design and Validation of a Framework for the Creation of User Experience Questionnaires. *International Journal of Interactive Multimedia and Artificial Intelligence* 5, 7 (2019), 88. doi:10.9781/ijimai.2019.06.006

[28] Ashish Sharma, Kevin Rushton, Inna Wanyin Lin, Theresa Nguyen, and Tim Althoff. 2024. Facilitating Self-Guided Mental Health Interventions Through Human-Language Model Interaction: A Case Study of Cognitive Restructuring. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–29. doi:10.1145/3613904.3642761

[29] Steven Siddals, John Torous, and Astrid Coxon. 2024. "It happened to be the perfect thing": experiences of generative AI chatbots for mental health. *npj Mental Health Research* 3, 1 (Oct. 2024), 48. doi:10.1038/s44184-024-00097-4

[30] Dominik Siemon, Rangina Ahmad, Henrik Harms, and Triparna De Vreede. 2022. Requirements and Solution Approaches to Personality-Adaptive Conversational Agents in Mental Health Care. *Sustainability* 14, 7 (March 2022), 3832. doi:10.3390/su14073832

[31] Marko Tkalcic and Li Chen. 2015. Personality and Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, 715–739. doi:10.1007/978-1-4899-7637-6_21

[32] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Verena Winterhalter, Markus Bühner, and Heinrich Hussmann. 2020. Developing a Personality Model for Speech-based Conversational Agents Using the Psycholexical Approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. doi:10.1145/3313831.3376210

[33] Zhenyu Xu, Hailin Xu, Zhouyang Lu, Yingying Zhao, Rui Zhu, Yujiang Wang, Mingzhi Dong, Yuhu Chang, Qin Lv, Robert P. Dick, Fan Yang, Tun Lu, Ning Gu, and Li Shang. 2024. Can Large Language Models Be Good Companions?: An LLM-Based Eyewear System with Conversational Common Ground. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (May 2024), 1–41. doi:10.1145/3659600

[34] Shengchen Zhang, Weiwei Guo, and Xiaohua Sun. 2025. Align with Me, Not TO Me: How People Perceive Concept Alignment with LLM-Powered Conversational Agents. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–10. doi:10.1145/3706599.3720126

[35] Xi Zheng, Zhuoyang Li, Xinning Gui, and Yuhan Luo. 2025. Customizing Emotional Support: How Do Individuals Construct and Interact With LLM-Powered Chatbots. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–20. doi:10.1145/3706598.3713453

[36] Yong Zheng, Xiaosong Lu, Xintong Yu, and Vinay Kambhampati. 2025. Personalizing Educational Responses with LLMs: The Influence of Knowledge, Interests, and Preferences. In *Proceedings of the 26th ACM Annual Conference on Cybersecurity & Information Technology Education*. ACM, Sacramento , California , USA, 141–147. doi:10.1145/3769694.3771123