

Design and Challenges of Mental Health Assessment Tools Based on Natural Language Interaction

Yixue Cai
Nankai University
College of Software
Tianjin, China
2213221@mail.nankai.edu.cn

Xiyan Su
Nankai University
College of Software
Tianjin, China
2313131@mail.nankai.edu.cn

Dongpeng Yao
Nankai University
Tianjin, China
13651362566@163.com

Rongduo Han
Nankai University
College of Software
Tianjin, China
1120240404@mail.nankai.edu.cn

Nan Gao*
Nankai University
College of Software
Tianjin, China
nan.gao@nankai.edu.cn

Haining Zhang*
Nankai University
College of Software
Tianjin, China
zhanghaining@nankai.edu.cn

Abstract

Mental health assessments are of central importance to individuals' well-being. Conventional assessment methodologies predominantly depend on clinical interviews and standardised self-report questionnaires. Nevertheless, the efficacy of these methodologies is frequently impeded by factors such as subjectivity, recall bias, and accessibility issues. Furthermore, concerns regarding bias and privacy may result in misreporting in data collected through self-reporting in mental health research. The present study examined the design opportunities and challenges inherent in the development of a mental health assessment tool based on natural language interaction with large language models (LLMs). An interactive prototype system was developed using conversational AI for non-invasive mental health assessment, and was evaluated through semi-structured interviews with 11 mental health professionals (six counsellors and five psychiatrists). The analysis identified key design considerations for future development, highlighting how AI-driven adaptive questioning could potentially enhance the reliability of self-reported data while identifying critical challenges, including privacy protection, algorithmic bias, and cross-cultural applicability. This study provides an empirical foundation for mental health technology innovation by demonstrating the potential and limitations of natural language interaction in mental health assessment.

CCS Concepts

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

*Corresponding authors

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UbiComp Companion '25, Espoo, Finland

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1477-1/2025/10
<https://doi.org/10.1145/3714394.3756239>

Keywords

Mental Health Assessment; Natural Language Interaction; Large Language Models; Self-Report Data; Adaptive Questionnaire

ACM Reference Format:

Yixue Cai, Xiyan Su, Dongpeng Yao, Rongduo Han, Nan Gao, and Haining Zhang. 2025. Design and Challenges of Mental Health Assessment Tools Based on Natural Language Interaction. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '25)*, October 12–16, 2025, Espoo, Finland. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3714394.3756239>

1 Introduction

Mental health assessments are imperative for the well-being and social functionality of individuals[16]. Conventional mental health assessment methodologies are mainly based on clinical interviews, professional counseling, and standardised self-report questionnaires. Despite the prevalence of these methods, they are frequently constrained by subjectivity, recall bias, and accessibility barriers. Nevertheless, a substantial body of research has cast doubt on the reliability of self-reporting. For instance, researchers have observed a semantic discrepancy between self-reported and observed emotions[5], as well as low consistency between self-reporting and behavioral measures. Longitudinal studies have further revealed issues such as social desirability bias, recall errors, and declining compliance over time [1, 7, 14, 23, 27]. These challenges are especially salient in the domain of mental health, as social stigmatization and concerns regarding privacy may result in intentional or unintentional misreporting.

Natural language interaction is a promising approach to address these limitations. It can reduce the artificiality of structured questionnaires while capturing richer, contextual information about an individual's mental state. Large language models (LLMs) are suitable for mental health applications where nuanced communication is essential.

This study explores the design of mental health assessment tools based on natural language interaction with LLMs. We have developed an interactive prototype system, the Intelligent Mental Health Tool, which uses conversational AI to conduct non-invasive mental health assessments. The system can generate and adapt self-report

questions based on user feedback, analyse text data to infer emotional states and detect inconsistencies, and enhance psychological log analysis methods. Furthermore, it has been demonstrated that the platform provides users and clinicians with interpretable feedback, thus enhancing transparency and trust.

To evaluate and refine this approach, we conducted in-depth interviews with six professional counsellors and five psychiatrists. These interviews helped us map typical workflows in counselling sessions, determine the current state of AI integration in mental health services, and gather expert perspectives on the challenges and opportunities associated with LLM-based assessment tools. Our research was guided by the following questions:

How to design a mental health assessment tool based on natural language interaction? What are the main challenges it currently faces?

This study offers an empirical foundation for mental health technology innovation and explores the potential of AI-driven adaptive questioning to improve the reliability of self-reported data. The study also highlights key challenges such as privacy protection, bias, and cross-cultural applicability.

The contributions of this paper are as follows.

- Through in-depth interviews with professional counsellors and psychiatrists, we systematically explored the limitations and challenges of self-reported data in mental health assessment and identified specific requirements for LLM-based assessment tools;
- We proposed and evaluated a design methodology for an adaptive questionnaire system based on large language models, which can dynamically adjust questions, integrate multimodal data, and provide interpretable feedback;
- We identified critical challenges in implementing such systems, including privacy concerns, bias mitigation, and cross-cultural adaptation, providing concrete guidance for future research and development in this emerging field.

2 Related Works

2.1 Measurements of Mental Health

Traditional mental assessment methods are often limited by subjectivity, recall bias, and accessibility barriers [24]. However, recent technological advances have enabled new approaches, such as mobile devices and wearable sensors, to passively and continuously collect behavioural and physiological data in real-world settings [8]. For example, Samyoun et al. [22] proposed M3Sense, a framework that uses multimodal sensors to capture user behaviour and impact, while Tlachac et al. [26] demonstrated the use of smartphone logs for mental health screening during the pandemic. In educational contexts, Gao et al. (2020) predicted student engagement using the n-Gage system. These new methods, which combine machine learning and digital data, offer more granular and personalised models of mental health [1, 7, 14, 23, 27], but also raise new challenges around privacy and data interpretation.

2.2 The Challenge of Self-Report Authenticity

A substantial body of research has cast doubt on the reliability of self-report. For instance, Boyd & Andalibi [5] observed a semantic gap between observed emotions and self-reported affect, while Das Swain et al. [9] found low agreement between self-report and behavioural measures. Longitudinal investigations have further revealed issues such as social desirability bias, recall errors, and declining compliance over time [11?, 12]. These challenges are especially evident in mental health settings, where stigma and concerns about privacy can lead to deliberate and inadvertent misreporting.

2.3 AI and LLM-based Solutions

For example, Xu et al. [28] conducted a comprehensive evaluation of large language models on various mental health prediction tasks using online text data, highlighting the effectiveness of instructional fine-tuning in improving model performance. Bouguettaya et al. [4] provided a comprehensive review of AI-driven report-generation tools in mental healthcare. Furthermore, several studies have examined the effectiveness of LLMs as psychological assessors, demonstrating their potential to provide interpretable mental health screening via psychometric practices [13, 15, 18–20]. These findings suggest that AI has the potential to bridge the gap between subjective and objective indicators, enabling more personalized and adaptive assessment, while also raising critical questions related to privacy, bias, and cross-cultural applicability.

3 Methodology

Our research employed a mixed-methods approach consisting of two main phases:

3.1 Prototype Development Phase

We developed a functional prototype of a natural language-based mental health assessment tool to provide a concrete reference point for expert evaluation. The prototype integrated the Symptom Checklist-90 (SCL-90) scale [10] into a conversational format using the Doubao 1.5 Pro large language model [2] on the Coze platform [6]. This phase aimed to transform abstract concepts into a tangible system that experts could evaluate.

3.2 Expert Evaluation Phase

We conducted semi-structured interviews with 11 mental health professionals, comprising six counselors and five psychiatrists. These interviews served multiple purposes: to gain insight into their current workflows and assessment methodologies, to explore their perspectives regarding the challenges associated with self-reported data, to gather feedback on our prototype system, and to identify essential design requirements for future development. Through these conversations with practitioners, we aimed to establish a comprehensive understanding of the clinical context in which our system would operate and the specific needs it should address.

3.3 Data Analysis

Interview transcripts were analyzed using thematic analysis to identify key patterns and insights. We focused on three main areas: (1)

common mental health issues among college students, (2) professional inquiry approaches and diagnostic criteria, and (3) specific feedback on our prototype system.

4 Prototype Design and Development

4.1 Introduction to the System

To address our research questions, we developed a prototype system using advanced language models, providing interview participants with a concrete demonstration of natural language interaction for mental health assessment. Our system transforms traditional psychological testing into a conversational experience by integrating standardized assessments within natural dialogue.

The prototype is built on the Doubao 1.5 Pro large language model deployed on the Coze platform, selected for its strong performance in Chinese language understanding and contextual conversation management. We implemented the SCL-90 as the underlying assessment instrument for two reasons: it is widely used for mental health screening among college students, and its 90 items can be effectively restructured into conversational content, allowing the system to elicit symptom information through natural dialogue rather than explicit questioning.

The system adaptively navigates through assessment areas based on user responses, focusing on domains where potential concerns are detected while maintaining a supportive conversational flow.

4.2 System Architecture and Workflow

The system architecture consists of four core modules that form a complete assessment loop. The natural language interface converts standardised SCL-90 questions into three types of natural inquiry patterns: situational questions corresponding to interpersonal sensitivity dimensions, experiential questions corresponding to somatic symptom dimensions, and coping questions assessing psychological resilience. The dynamic assessment engine is capable of analysing emotional intensity and symptom keywords in user language in real time, constructing a symptom matrix, and intelligently selecting follow-up questions based on current symptom coverage to evaluate psychological resilience. The context processing mechanism maintains conversational coherence through a three-layer memory architecture to ensure a natural and smooth assessment process. The feedback and intervention module generates three-dimensional feedback based on the results of the assessment and provides the relevant support resources when risk signals are detected.

The system workflow is divided into three stages (Figure 1). The first stage, which takes approximately two minutes, is the relationship-building stage. This stage establishes trust through everyday topics and perform basic determination of the degree of impairment in social functioning. The in-depth assessment phase (approximately five minutes) dynamically activates relevant symptom dimension assessment pathways based on the user's initial feedback. The results delivery stage (approximately one minute) analyses user feedback to generate multi-dimensional assessment results, including emotional state, functional impact and positive resources.

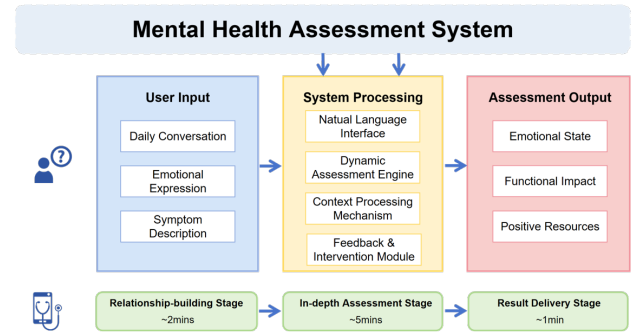


Figure 1: System Workflow

4.3 Large Language Model Configuration

On Coze platform, we made several key configuration changes to the Doubao 1.5 Pro model. These included setting randomness parameter to 0.8 to ensure dialogue variety and naturalness while maintaining response consistency; employing a three-round dialogue memory mechanism to balance dialogue coherence and computational complexity; and setting maximum response length to 4,096 tokens to enable the system to generate detailed evaluation content.

We constructed a multilevel instruction structure to clarify the system's professional identity as a 'college student mental health counsellor', used the first person to conduct naturalistic dialogues for the assessment, and implemented a dynamic assessment system, naturalistic questioning techniques, and a hybrid feedback mechanism. This mechanism describes the three assessment phases and their implementation strategies in detail, incorporating the nine dimensions of the symptoms and their scoring criteria of the SCL-90. These nine dimensions and their scoring criteria are integrated into the dialogue logic to provide the system with a professional assessment basis.

We designed a language conversion mechanism that transforms standardized scale items into natural dialogue expressions, aiming to preserve the essence of the assessment while improving the naturalness of communication. For example, the standard question 'Do you often feel nervous or easily agitated?' is converted into 'How do you react when you are under stress lately?'. This approach reduces the time required for assessment to 5-10 minutes compared to traditional standardized psychological assessments, while potentially increasing user engagement and data authenticity. All procedures follow established ethical codes for psychological assessment.

5 User Interview Progress

In order to achieve the research objectives, we recruited university counsellors and psychiatrists, with whom we conducted semi-structured in-depth interviews.

Table 1 categorises participants by profession and displays their basic and professional information, including gender, specialty field, occupation (Psychological Counselor = Psy, Psychiatrist = Ali), years of experience (years), clinical hours dedicated to mental health cases (hours), and department affiliation (Department). The use of

a hyphen (-) indicates that the information is not applicable or was not provided by the participant.

Table 1: Participant Information

ID	Sex	Professional field	Occ.	Yrs	Hours	Dept.
P1	F	College students' mental health	Psy	5	500	-
P2	F	College students' mental health	Psy	10	300+	-
P3	F	College students' MH, emotional disorders, stress	Psy	6	1500	-
P4	M	College students' MH and stress management	Psy	15	3000	-
P5	F	College students' MH, emotional disorders, stress	Psy	3	1000	-
P6	F	MH, emotional disorders, stress management	Psy	12	5000	-
S1	F	Other (mental disorders)	Ali	10	-	Psy
S2	M	Dysthymic disorder	Ali	10	-	Psy
S3	F	MH, emotional disorders, stress management	Ali	13	-	PC
S4	F	College students' mental health	Ali	16	-	Psy
S5	F	College students' MH, emotional disorders, stress	Ali	6	-	Psy

5.1 Recruitment of Subjects

Participants were recruited through an online questionnaire. They were recruited from tertiary hospitals and screened according to the years of experience and treatment approach. Finally, 11 participants (six counsellors and five psychiatrists) were recruited to conduct online interviews after signing the informed consent form. After the interview, all participants received a remuneration of 100 yuan.

5.2 Interview Question Design

In order to understand the processes and principles of psychological counselling and diagnosis for college students, and to further enhance the adaptive mental health assessment system, two open-ended interview questionnaires have been designed targeting participants' professions, based on the functionalities of the developed prototype and the goals for improvement.

The questionnaires progressed in stages, starting with a basic introduction to understand college students' psychological status and collecting objective information. We then collected information on counselling and diagnostic processes and entry points. Together with the participants, we refined the model and rationale. Finally, we collected feedback and information on professionalisation and improvement. The questionnaire encouraged a comprehensive approach to students' psych problems, to achieve an accurate classification, understand assessment indicators, and collect effective counselling and diagnostic strategies.

5.3 Interview Process

The interviews were scheduled according to each participant's preferred time and online platform, each lasting between 25 and 45 minutes. Participants were then informed about the recording process before the interviews officially began. The interviews focused on the mental health status of college students and were based on the questionnaire questions. The key key questions and related topics were then explored. Finally, researchers provided participants with relevant requirements for the AI usage, offering additional references and targeted feedback for participants. After the interviews were transcribed, the researchers corrected them and summarised the information thoroughly.

5.3.1 Common Types of Mental Disorders Among College Students and Information Required for Diagnosis. Psychiatrists and counsellors experience mild to moderate anxiety and depression among college students. They agree that understanding a client's (A person who seeks professional psychological services (e.g., therapy, assessment), particularly in nonmedical settings like counselling or psychotherapy. This term emphasizes collaboration and autonomy.) complaint and symptoms (severity, duration, etc.) and their circumstances (academic performance, interpersonal relationships, etc.) guides the conversation and observes the client's behaviour. This information is then supplemented by scales to assess the diagnosis.

The difference lies in the fact that counsellors' assessments are primarily based on psychological assessment criteria, while psychiatrists' judgments are formed based on diagnostic systems. For example, P2 states: "After thoroughly understanding the impact of the issues the client is facing, such as duration, severity, and the extent of impact on social functioning, we combine their subjective experiences to assess whether the issue falls under general psychological problems or severe psychological problems."; S2 states, "We assess whether clients meet standardized diagnostic criteria (such as ICD-10), integrating hospital systems to construct core symptom clusters based on initial presentations. These clusters are progressively refined and differentiated to prioritize symptom onset, ultimately forming a diagnosis when meeting criteria for course and severity. Diagnostic priority is given to the distress level of symptoms."

5.3.2 Inquiry Approach and Assessment Diagnostic Criteria. Psychological counsellors typically adopt an inquiry approach centered on open communication. After clarifying the primary complaint, they primarily listen and provide emotional support, conducting superficial emotional cognition, behavioral assessment, mental state assessment, and crisis assessment through language. After jointly confirming the counseling direction with the client, they gradually implement intervention measures based on the information obtained. Their assessment criteria primarily include the SAS, SDS, SCL-90, and DSM-V.

"By analyzing potential issues or scenarios, I help students organize or explore what they've said. I may share my perspective with them and invite them to discuss whether we can delve deeper into that aspect together."(P5)

Psychiatrists, after understanding the chief complaint, focus on confirming basic information and related symptoms. They construct a symptom cluster based on the symptoms described by the client, determine the focus, and then conduct relevant inquiries. Finally, they use open-ended questions to further understand the symptoms and determine the treatment plan. Their diagnostic criteria primarily include ICD-10, state-related knowledge, disease-related knowledge, and DSM-5.

"During the first consultation, we inquire about relevant circumstances, such as where they are currently attending school, their current environment, their childhood state, and their parents' parenting style. Then, we discuss issues which cause them confusion. We typically ask them to provide a specific scenario, and then

we delve into the details of the event, using this as a starting point.”(S5)

It is noteworthy that during actual interviews, researchers once asked participants about their “questioning process.” However, most counselors and psychiatrists indicated that there is no fixed questioning process. When interacting with clients, they follow the principle of “following the client’s train of thought,” adjusting subsequent questions based on the client’s responses and their own professional orientation. They aim to guide clients toward further directed interpretations rather than posing isolated, unrelated, or closed-ended questions solely for information gathering. This response emphasises the distinct nature of addressing psychological symptoms, thereby significantly enhancing the research methodology. This is the rationale behind the selection of this title.

5.3.3 Feedback on Prototype. After standardized AI usage, counselors and doctors provided targeted feedback. All participants affirmed the AI’s basic functions but also identified areas requiring improvement or enhancement.

Psychological counselors emphasized empathy and guidance, highlighting issues with the AI’s empathy functionality and the coherence of its responses. The AI demonstrated preliminary empathy during use, establishing a foundation for comfort, but it still faced challenges such as rigid response language and insufficient recognition capabilities, which caused frustration for users.

For example, P1 state *“During communication, it showed concern for my role as an anxious student, and the opening remarks were also natural.”* Similarly, P3 noted that the AI’s responses helped users understand their own state and provided relevant professional knowledge, but also mentioned issues with its ability to connect questions: *“During a single-round interaction, it showed concern for my well-being, which made me feel at ease. However, when I entered simple, conversational responses, it immediately displayed ‘Unrecognized’ and offered technical jargon, demanding a detailed description of the situation to complete the query. Yet, if I told it I didn’t want to answer, it would switch to another topic and continue the conversation—which proved quite useful.”*

Psychiatrists emphasize accuracy and comprehensiveness, highlighting AI’s ability to capture and collect information. The AI adheres to psychological expertise and scale-based methodologies to ensure comprehensive information gathering. However, it exhibits issues with inaccurate responses and weak logical reasoning during data input, resulting in reduced consistency during usage.

S1 mentioned the issue of questioning approach: *“It’s a good listening partner. However, when I mentioned that I’ve been feeling down lately and have some troubles, it kept asking me about the disease-related aspects and didn’t address my emotions.”*

6 Discussion

Our research aimed to develop and evaluate an LLM-based mental health assessment tool that overcomes limitations of traditional questionnaires through imperceptible assessment in natural conversation. This section discusses our findings, limitations, and future directions in relation to our core contribution: creating an intelligent agent that can conduct psychological assessments through natural dialogue while maintaining clinical validity.

6.1 Summary of Contributions

The primary contribution of this work is the development and initial validation of an intelligent mental health assessment system based on the Doubao 1.5 Pro large language model. Unlike traditional questionnaire-based assessments that often feel clinical and impersonal, our system transforms standardized SCL-90 assessment items into natural conversational elements. For example, rather than directly asking “Do you often feel nervous or easily agitated?”, our system might inquire, “How do you react when you are under stress lately?” This approach maintains assessment validity while significantly improving the user experience.

Our semi-structured interviews with 11 participants (six counselors and five psychiatrists) validated this approach, with participants noting that natural language interaction could potentially increase user engagement and data authenticity compared to traditional assessment methods. However, the interviews also revealed three areas in need of improvement before such systems can be used in clinical settings: the ability to show empathy, the accuracy of assessments, and crisis intervention protocols.

6.2 Areas for Improvement

Empathy. Research indicates that individuals with psychological symptoms or disorders often require communication from an equal stance with appropriate emotional support rather than excessive sympathy [25]. This finding was corroborated by the results of the clinician interviews conducted for this study. Some participants emphasised that users should “feel that their inner thoughts are understood and be given the courage to overcome negative emotions” (S2) when interacting with the assessment system.

The current prototype demonstrates limitations in recognizing and appropriately responding to emotional cues, particularly when users express distress indirectly. To address this limitation, we propose enhancing our system with improved emotional detection capabilities and response templates derived from therapeutic communication practices [3]. Specifically, the system should initiate conversations from everyday topics, prioritize emotional comfort, and transition to supportive statements when detecting sustained emotional distress, such as *“I notice this seems difficult to talk about. Perhaps we could take a moment to focus on something more comfortable before returning to this topic if you’re willing.”*

Intelligent assessment. The accuracy of psychological assessment depends significantly on the interactivity of language and comprehensiveness of information collection [21]. Our current prototype faces challenges in recognizing certain response patterns and maintaining conversational coherence—limitations identified by multiple clinicians in our interviews.

To improve assessment accuracy, we will enhance the system’s natural language understanding capabilities to better interpret colloquial expressions of psychological distress. For instance, when a user states, “I just don’t feel like getting out of bed anymore,” the system should recognize this as potentially indicating depression rather than physical fatigue. Additionally, we will implement more sophisticated conversational strategies to encourage elaboration without appearing intrusive, such as using open-ended prompts like *“Would you like to share how that experience affected you?”* rather than direct assessment questions. These improvements directly

address clinicians' concerns about the system's ability to gather clinically relevant information through natural conversation.

Crisis intervention. Effective crisis intervention capabilities are essential for any mental health assessment tool, as highlighted by both existing literature [17] and our interview participants. Current prototype lacks mechanisms for identifying and responding to indicators of acute psychological distress or suicidal ideation.

Based on clinician feedback, we propose implementing a specialized crisis detection module that can recognize linguistic patterns associated with imminent risk. When concerning language is detected, the system will implement a tiered response protocol: first providing supportive statements, then offering crisis resources, and finally activating an alert to the user's pre-designated support person in severe cases (with prior consent). This approach balances immediate safety concerns with user autonomy and privacy—a balance that multiple clinicians in our study identified as crucial for technology-mediated mental health assessment.

6.3 Limitations and Future Work

Despite the evidence from our research that demonstrates the potential of mental health assessments based on large language models, there are several key limitations that must be addressed.

Firstly, it is important to note that conversational rephrasing has the potential to alter the psychometric properties of the original assessment tool. This, in turn, may necessitate the undertaking of comparative validation studies in the future. Secondly, the system's adaptive approach to user responses challenges the integrity of the assessment, necessitating exploration of methods to balance natural conversation with psychometric integrity.

Presently, the prototype is undergoing expert interview testing, with more rigorous validation to be conducted during the forthcoming clinical trial evaluation phase. Concurrently, we will proceed with the exploration of multimodal assessment methods with a view to enhancing accessibility for diverse populations.

In summary, the findings of the present study demonstrate that mental health assessment using natural conversation based on large language models is technically feasible and has potential value in improving user engagement and data authenticity. However, before such systems can be deployed responsibly in clinical settings, there is a need for significant improvements in empathy, assessment accuracy, and crisis intervention capabilities. These findings provide a solid foundation for the further development of mental health assessment methods that can enhance the assessment experience while maintaining clinical efficacy.

7 Conclusion

This study demonstrates the feasibility and potential benefits of a mental health assessment system based on natural language interaction. The integration of AI-driven technology with psychological assessment principles has been demonstrated to facilitate a more conversational and adaptive assessment process. A thematic analysis of semi-structured interviews with 11 participants (six counsellors and five psychiatrists) was conducted, leading to the identification of three fundamental design principles. The findings of this study indicate that AI-driven adaptive questioning provides a

more comfortable and flexible assessment experience while demonstrating good effectiveness in improving psychological logs and assisting in the construction of psychological models. Given the individual willingness to use AI systems, despite privacy concerns, the information obtained through this method appears highly credible, highlighting its theoretical potential for practical psychological assessment applications. While further validation is required, this work may result in the development of novel assessment methods and the optimisation of counselling and diagnostic processes between mental health professionals and their clients.

Acknowledgments

This work is supported by the Natural Science Foundation of China (Grant No. 62302252).

References

- [1] Daniel A Adler, Vincent W-S Tseng, Gengmo Qi, and et al. 2021. Identifying mobile sensing indicators of stress-resilience. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5, 2 (2021), 1–32.
- [2] BAAI. 2023. Doubao 1.5 Pro: A Large Language Model for Conversational AI. <https://www.baai.ac.cn/doubao>.
- [3] Eric P. S. Baumer, Sherri Jean Katz, Jill E. Freeman, and et al. 2012. Prescriptive Persuasion and Open-Ended Social Awareness: Expanding the Design Space of Mobile Health. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (2012).
- [4] Ayoub Bouguettaya, Victoria Team, Elizabeth M Stuart, and Elias Aboujaoude. 2025. AI-driven report-generation tools in mental healthcare: A review of commercial tools. *General Hospital Psychiatry* (2025).
- [5] Karen L Boyd and Nazanin Andalibi. 2023. Automated emotion recognition in the workplace: How proposed technologies reveal potential futures of work. *Proceedings of the ACM on human-computer interaction* 7, CSCW1 (2023), 1–37.
- [6] ByteDance. 2023. Coze: A Platform for Building and Deploying Conversational AI Applications. <https://www.coze.com>
- [7] Meghna Roy Chowdhury, Wei Xuan, Shreyas Sen, and et al. 2025. Predicting and Understanding College Student Mental Health with Interpretable Machine Learning. *arXiv preprint arXiv:2503.08002* (2025).
- [8] Victor P Cornet and Richard J Holden. 2018. Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of biomedical informatics* 77 (2018), 120–132.
- [9] Vedant Das Swain, Victor Chen, Shrija Mishra, and et al. 2022. Semantic gap in predicting mental wellbeing through passive sensing. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–16.
- [10] Leonard R. Derogatis, Ronald S. Lipman, and Lino Covi. 1973. SCL-90: An outpatient psychiatric rating scale—preliminary report. *Psychopharmacology Bulletin* 9, 1 (1973), 13–28.
- [11] Nan Gao, Mohammad Saiedur Rahaman, Wei Shao, and Flora D Salim. 2021. Investigating the reliability of self-report data in the wild: The quest for ground truth. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 237–242.
- [12] Nan Gao, Wei Shao, Mohammad Saiedur Rahaman, and Flora D Salim. 2020. n-gage: Predicting in-class emotional, behavioural and cognitive engagement in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–26.
- [13] Tin Lai, Yukun Shi, Zicong Du, and et al. 2023. Psy-llm: Scaling up global mental health psychological services with ai-based large language models. *arXiv preprint arXiv:2307.11991* (2023).
- [14] Boning Li and Akane Sano. 2020. Extraction and interpretation of deep autoencoder-based temporal features from wearables for forecasting personalized mood, health, and stress. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–26.
- [15] Julia Ohse, Bakir Hadžić, Parvez Mohammed, and et al. 2024. GPT-4 shows potential for identifying social anxiety from clinical interview data. *Scientific Reports* 14, 1 (2024), 1–12.
- [16] World Health Organization et al. 2013. Comprehensive mental health action plan 2013–2020. *Geneva: World Health Organization* 48 (2013).
- [17] Sachin R. Pendse, Faisal M. Lalani, Munmun De Choudhury, and et al. 2020. “Like Shock Absorbers”: Understanding the Human Infrastructures of Technology-Mediated Mental Health Support. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020).
- [18] Ahmad Radwan, Mohannad Amarnah, Hussam Alawneh, and et al. 2024. Predictive analytics in mental health leveraging LLM embeddings and machine learning

- models for social media analysis. *International Journal of Web Services Research* 21, 1 (2024), 1–22.
- [19] Federico Ravenda, Seyed Ali Bahrainian, Andrea Raballo, and et al. 2025. Are LLMs effective psychological assessors? Leveraging adaptive RAG for interpretable mental health screening through psychometric practice. *arXiv preprint arXiv:2501.00982* (2025).
 - [20] Gony Rosenman, Lior Wolf, and Talma Hendler. 2024. LLM Questionnaire Completion for Automatic Psychiatric Assessment. *arXiv preprint arXiv:2406.06636* (2024).
 - [21] C. Rosé, YC. Wang, Y. Cui, and et al. 2008. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning* (2008).
 - [22] Sirat Samyoun, Md Mofijul Islam, Tariq Iqbal, and John Stankovic. 2022. M3sense: Affect-agnostic multitask representation learning using multimodal wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–32.
 - [23] Daun Shin, Hyoseung Kim, Seunghwan Lee, and et al. 2024. Using large language models to detect depression from user-generated diary text data as a novel approach in digital mental health screening. *Journal of Medical Internet Research* 26 (2024), e54617.
 - [24] Arthur A Stone, Saul Shiffman, Joseph E Schwartz, and et al. 2003. Patient compliance with paper and electronic diaries. *Controlled clinical trials* 24, 2 (2003), 182–199.
 - [25] Kenneth R Koedinger Thiemo Wambsganss, Matthias Soellner and Jan Marco Leimeister. 2022. Adaptive Empathy Learning Support in Peer Review Scenarios.
 - [26] ML Tlachac, Ricardo Flores, Miranda Reisch, and et al. 2022. DepreST-CAT: Retrospective smartphone call and text logs collected during the covid-19 pandemic to screen for mental illnesses. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–32.
 - [27] Xuhai Xu, Purna Chikersal, Janine M Dutcher, and et al. 2021. Leveraging collaborative-filtering for personalized behavior modeling: a case study of depression detection among college students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.
 - [28] Xuhai Xu, Bingsheng Yao, Yuanzhe Dong, and et al. 2024. Mental-llm: Leveraging large language models for mental health prediction via online text data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 1 (2024), 1–32.