Medicine Plus 2 (2025) 100085

Contents lists available at ScienceDirect

Medicine Plus

journal homepage: www.sciencedirect.com/journal/medicine-plus

Perspective

Foundation models for digital mental health: igniting the dawn

Kun Qian[a,b,*], Haojie Zhang[a,b], Xin Jing[c], Bin Hu[a,b,*], Yoshiharu Yamamoto[d], Björn W. Schuller[c]

[a] *Key Laboratory of Brain Health Intelligent Evaluation and Intervention, Ministry of Education (Beijing Institute of Technology), Beijing 100081, China*

[b] *School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China*

[c] *Chair of Health Informatics, Technische Universität München, Munich, 80333, Germany*

[d] *Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Tokyo, 113-0033, Japan*

In the past decade, information technologies (e.g., artificial intelligence (AI), big data, wearables) have deeply influenced the field of mental health. As a typical breaking-through idea, computational psychophysiology (CPP) has changed the paradigm of mental healthcare from traditional "symptom description-driven" to "data-driven".[1] In this rapidly evolving landscape, AI-powered tools like ChatGPT are further revolutionizing the field, marking a significant milestone in the AI era. This technological leap is closely tied to the emergence of foundation models (FMs), which have set a new benchmark in AI development. Foundation Models represent a significant advancement by offering large, pre-trained models that can be fine-tuned for a wide variety of tasks. Their ability to process diverse data types—ranging from text and images to video and structured data—enables them to solve complex problems across numerous domains, including mental health.[2] Built on transformer architectures and trained on massive datasets in an unsupervised manner, these models generate generalizable features that make them highly adaptable for downstream tasks. An overview of how FMs contribute across the key stages of digital mental health—ranging from data acquisition to adaptive feedback—is illustrated in Fig. 1.

Recent advancements in AI-enabled systems have significantly impacted mental health, including emotion recognition and personalized interventions. A notable application of FMs is the development of AI-powered chatbots[3], like GPT-4, for psychiatric assessments and automated patient communication, enabling non-invasive mental health screening and transforming how healthcare providers interact with patients and make clinical decisions. These models leverage clinical records, sensor data, and biometric readings to predict depression, anxiety, and stress, offering real-time monitoring and dynamic treatment adjustments. FMs are also used in emotion recognition systems that analyze multimodal inputs,

* Corresponding authors.
E-mail addresses: qian@bit.edu.cn (K. Qian), bh@bit.edu.cn (B. Hu).

https://doi.org/10.1016/j.medp.2025.100085
Received 18 March 2025; Received in revised form 15 April 2025; Accepted 23 April 2025
Available online 29 April 2025
2950-3477/

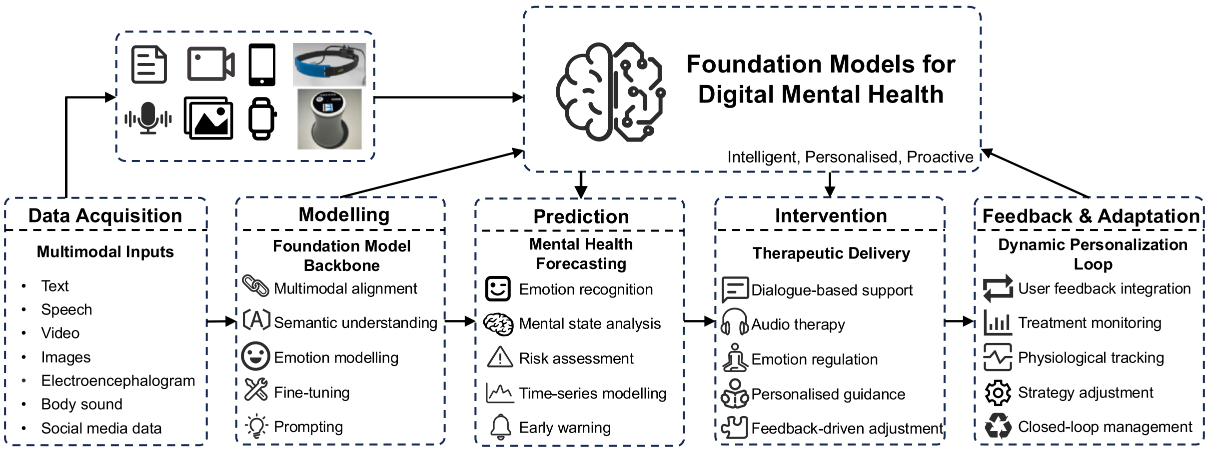

Fig. 1. A conceptual framework of foundation models for digital mental health.

including text, speech, and physiological signals, to detect emotional distress and track conditions such as autism spectrum disorder (ASD) and social anxiety[4]. Language plays a crucial role in mental health assessment, with text analysis helping to identify cognitive distortions and depressive symptoms, such as increased use of negative words, while speech analysis captures emotional changes through variations in speech rate, tone, and pauses, offering deeper insights into psychological states. Additionally, audio-based interventions, driven by FMs, generate calming content for emotional regulation and stress relief, proving particularly effective in self-regulation and emotional well-being for individuals avoiding in-person therapy.

Furthermore, FMs enable continuous long-term monitoring of patients' mental well-being through various data streams, such as electroencephalography signals, which reflect brain activity. By analyzing these signals, FMs can detect mental disorders such as depression, anxiety, and epilepsy, providing clinicians with valuable insights into the progression of a patient's condition over time and enabling predictions before clinical symptoms of mental health events manifest[5]. FMs also enhance personalized mental health interventions by tailoring treatments based on a patient's history, symptoms, and emotional data, offering adaptive, precise therapeutic strategies. Predictive mental health systems, which integrate wearable devices, social media analysis, and natural language processing (NLP),[6] enable preventive care before symptoms escalate.

Brain-inspired models are further advancing emotional and cognitive function modeling in mental health, which are being used to simulate emotional learning processes and to model the neural mechanisms involved in mental health conditions. These brain-inspired AI models aim to replicate how the human brain processes emotions, learning from patterns of behavior and emotional responses, which can then be applied to deep brain stimulation or neurofeedback techniques to aid in treating conditions such as depression or anxiety. These models, along with brain-computer interfaces (BCIs),[7] offer advanced neurofeedback mechanisms for managing conditions like anxiety and post-traumatic stress disorder (PTSD), offering real-time mood regulation.[8] By combining FMs and brain-inspired models, the future of mental health interventions promises personalized, effective, and timely care, improving both diagnosis and treatment.

The dynamic evolution of mental health disorders, including their onset and progression, is difficult to predict. FMs are emerging as promising tools for early prediction and prevention of mental disorders. Although not yet widely used to model the precise temporal dynamics of these conditions, FMs' predictive capabilities are rapidly advancing. By analysing vast datasets of clinical records, biometric data, and behavioral patterns, FMs can identify early signs of emotional distress or cognitive changes that precede conditions like depression, anxiety, and bipolar disorder[9]. Additionally, FMs are being leveraged for time-series analysis and dynamic modeling to track mental health trajectories and identify early indicators, such as emotion regulation and cognitive dysfunctions, triggering timely interventions.[10] By combining real-time multimodal data from wearables, social media, and digital health platforms, FMs hold the potential to revolutionize mental health prediction, diagnosis, and prevention, enabling personalized healthcare and early intervention.

The exploration of the human mind mechanisms through computational and AI models has gained significant traction in recent years, though the use of FMs to directly understand and replicate human cognition remains in its early stage. FMs have shown promise in simulating emotional processing and decision-making, but fully modeling the complexity of the human mind remains a challenge. Currently, research on human cognition relies heavily on neuroscience-based models, such as neural networks and cognitive neuroscience frameworks, which utilize tools like functional magnetic resonance imaging (fMRI), electroencephalography, and magnetoencephalography to study brain activity during cognitive tasks.[11] Additionally, dynamic systems theory and computational psychiatry model brain circuits to understand conditions like depression and schizophrenia, enabling early diagnostic insights.[12] Psychological frameworks such as cognitive behavioral therapy (CBT), reinforcement learning (RL), and symbolic AI enhance understanding of cognitive distortions and emotional regulation. Advances in affective computing and emotion recognition also help model emotional states using facial expressions, speech, and physiological data. Though FMs are not yet central to cognitive modeling, integrating neuroscience, psychological theories, and FMs holds the potential to create personalized, predictive models for improved mental health diagnostics and interventions[13].

Despite these promising advancements, integrating FMs into mental health practices raises critical ethical considerations, including concerns regarding privacy, algorithmic bias, and clinical accountability[14]. The sensitive nature of mental health data, involving multimodal information such as biometrics, speech, and emotional expressions, necessitates robust privacy protection like differential privacy and federated learning to ensure patient confidentiality and data security[15]. Moreover, algorithmic bias, stemming from training datasets lacking diversity in terms of ethnicity, gender, age, or socioeconomic status, poses risks of systematic discrimination, potentially leading to inequitable care outcomes.[16] Thus, proactive measures—such as comprehensive bias audits, representative dataset curation, and transparent, explainable modeling—are essential for equitable model performance across diverse populations. Additionally, clinical accountability remains paramount; the deployment of AI-powered tools should adopt a human-in-the-loop approach, empowering healthcare providers to oversee, interpret, and validate AI-generated insights, thereby safeguarding patient welfare and aligning AI usage with established clinical standards and ethical guidelines.

These advancements in FMs represent a pivotal shift in how we approach mental health, enabling not only more precise diagnostics but also personalized and adaptive interventions. By integrating FM tools with traditional therapeutic methods—while remaining mindful of privacy, fairness, and clinical accountability—we are poised to create a future where mental health care is more proactive, efficient, and tailored to individual needs.

CRediT authorship contribution statement

Kun Qian: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition. **Haojie Zhang:** Writing – original draft. **Xin Jing:** Writing – review & editing. **Bin Hu:** Supervision, Resources, Funding acquisition. **Yoshiharu Yamamoto:** Writing – review & editing, Supervision. **Björn W. Schuller:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key R&D Program of China (2023YFC2506804), the National Natural Science Foundation of China (62272044 and 62227807), the Beijing Natural Science Foundation (L243034), the Ministry of Science and Technology of the People's Republic of China with the STI2030-Major Projects (2021ZD0201900), the Japan Society for the Promotion of Science (S24116), and the Teli Young Fellow Program from the Beijing Institute of Technology, China.

References

1. Hu B, Fan J, Zhen W, et al. Advances in computational psychophysiology. *Science.* 2015;350(6256) 114–114.
2. Malgaroli M, Schulterbrauck K, Myrick KJ, et al. Large language models for the mental health community: Framework for translating code to care. *Lancet Digit Health.* 2025;7(4):e282–e285.
3. Liu J M., Li D., Cao H., et al. Chatcounselor: a large language models for mental health support. arXiv:2309.15461. 2023.
4. Guo Z, Lai A, Thygesen JH, et al. *Large language models for mental health applications: systematic review. JMIR Mental Health.* 2024;11(1):e57400.
5. Hua Y., Liu F., Yang K., et al. Large language models in mental health care: a scoping review. arXiv:2401.02984. 2024.
6. Triantafyllopoulos A, Terhorst Y, Tsangko I, et al. Large language models for mental health. arXiv:2411.11880. 2024.
7. Shen L, Zhang H, Zhu C, et al. Enhancing emotion regulation in mental disorder treatment: An AIGC-based closed-loop music intervention system. *IEEE Trans Affect Comput.* 2025:1–16. https://doi.org/10.1109/TAFFC.2025.3557873
8. Tian F, Zhang H, Tan Y, et al. An on-board executable multi-feature transfer-enhanced fusion model for three-lead EEG sensor-assisted depression diagnosis. *IEEE J Biomed Health Inf.* 2025;29(1):152–165.
9. Hu J, Dong T, Gang L, et al. Enhancing llm for psychological understanding and evaluation. *IEEE Trans on Comput Soc Syst.* 2024;12(2):539–551.
10. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digit Med.* 2023;6(1):236.
11. Zhang Y, Wu W, Toll RT, et al. Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nat Biomed Eng.* 2021;5(4):309–323.
12. Hauser TU, Skvortsova V, De Choudhury M, Koutsouleris N. The promise of a model-based psychiatry: Building computational models of mental ill health. *Lancet Digit Health.* 2022;4(11):e816–e828.
13. Mao Y, Wei D, Yang W, et al. A neural predictive model of negative emotions for COVID-19. *IEEE Trans on Affect Comput.* 2022;14(4):2646–2656.
14. Lawrence HR, Schneider RA, Rubin SB, et al. The opportunities and risks of large language models in mental health. *JMIR Mental Health.* 2024;11(1):e59479.
15. Stade EC, Stirman SW, Ungar LH, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. NPJ Mental Health. *Research.* 2024;3(1):12.
16. Coppock H, Jones L, Kiskin I, Schuller B. COVID-19 detection from audio: Seven grains of salt. *Lancet Digit Health.* 2021;3(9):e537–e538.