A Comprehensive Literature Review on [LLMs For Mental Health Therapeutics]

Your Name
Your Affiliation

April 22, 2025

Abstract

This literature review examines the evolving landscape of [LLMs For Mental Health Therapeutics] research. It synthesizes key findings from [Number] studies published between [Year Range], focusing on [Key Themes]. The review identifies [Key Gaps/Contradictions] and proposes [Potential Future Directions].

Contents

1 Introduction

article lipsum

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

2 History of LLMs

article lipsum

2.1 Subtopic 1.1

Einstein introduced the theory of relativity in 1905 Arriba-Pérez and García-Méndez (2024).

2.2 Subtopic 1.2

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

3 Efficacy of Mental Health Chat Bots

Recent advancements in large language models (LLMs) have opened new possibilities in the field of digital mental health, offering scalable solutions for early detection, diagnosis, and support. However, these systems are not without challenges. This section explores the data sources, fine-tuning methods, performance trade-offs, and ethical implications of using AI-driven chatbots in mental health contexts. We highlight the impact of training data biases, discuss the balance between speed and accuracy, and examine real-world risks including misinformation, user trust, and demographic disparities.

3.1 Training Data and Bias

Chatbots and AI mental health models are typically trained on social media data, psychotherapy transcripts, and electronic health records (EHRs) to detect emotional distress and mental health conditions Arriba-Pérez and García-Méndez (2024). Common datasets include Dreaddit, DepSeverity, SDCNL, and CSSRS-Suicide, often sourced from Reddit with expert annotations for tasks like stress prediction and suicide risk detection Arriba-Pérez and García-Méndez (2024).

However, reliance on Reddit-based datasets introduces user and demographic biases, limiting generalizability **?**. These models often reinforce user distress rather than challenge it, increasing the likelihood of false positives Gbollie et al. (2023). Biases also emerge from lack of racial and gender diversity in training datasets, which can skew chatbot responses Arriba-Pérez and García-Méndez (2024). Even expert-annotated content can carry stereotypes and normative assumptions.

3.2 Speed vs. Accuracy Trade-offs

Larger models like GPT-4 and FLAN-T5 are optimized for speed but may lack the specificity of smaller, fine-tuned models like Mental-Alpaca, which excel in domain-specific tasks Xu et al. (2024). Streaming models allow real-time interaction but require continuous updates **?**, whereas batch models are slower but more accurate.

3.3 Fine-tuning and Performance

Fine-tuning strategies such as instruction tuning and LoRA (Low-Rank Adaptation) have improved LLM performance on Cognitive Behavioral Therapy (CBT) tasks. These models minimize prediction error using cross-entropy loss during training **?**. Mental-FLAN-T5, for example, outperforms GPT-4 by 4.8% in balanced accuracy Xu et al. (2024).

3.4 Coherence and Logical Flow

Fine-tuned models outperform general-purpose LLMs like GPT-4 in logical consistency and contextual understanding **?**. Woebot and Wysa deliver more structured responses than generative models Greco et al. (2023). GPT-based models may vary responses based on prompt phrasing, occasionally generating overly generic or logically inconsistent replies Sejnowski (2023).

3.5 Handling Ambiguity

Fine-tuned models demonstrate stronger performance on vague or ambiguous mental health queries, while zero-shot models default to binary responses. Few-shot prompting improves accuracy by approximately 4.1% in GPT-4 Xu et al. (2024).

3.6 User Engagement and Trust

While 81.1% of university students are willing to use AI chatbots for mental health support, only 4% have actually done so Gbollie et al. (2023). Privacy concerns, stigma, and fear of misdiagnosis limit adoption. However, users who avoid traditional therapy are more open to digital tools.

3.7 Limitations and Misinformation Risks

Despite high accuracy in detecting conditions like depression or cognitive impairment (96%+ and 80%, respectively) Greco et al. (2023), risks of misinformation, overgeneralization, and ethical concerns persist. LLMs often lack systematic evaluation mechanisms, making their predictions potentially unreliable for high-stakes applications.

3.8 Summary of Efficacy and Limitations

Studies highlight the promise of AI mental health models like CBT-LLM, which show superior response quality through fine-tuning **?**. However, their ability to truly understand and reason about mental health remains in question Sejnowski (2023). More rigorous evaluation, ethical safeguards, and transparency are necessary before widespread deployment **??**.

4 Privacy Concerns

article natbib
... Some text ... Gbollie et al. (2023). More text... (Sejnowski, 2023) ...

References

Arriba-Pérez, F. and García-Méndez, S. (2024). Leveraging llms for real-time mental health predictions. *Arabian Journal for Science and Engineering*.

Gbollie, E. F., Bantjes, J., and Jarvis, L. (2023). Intention to use digital mental health solutions. *Digital Health*, 9:1–19.

Greco, C. M., Simeri, A., and Tagarelli, A. (2023). Transformer-based language models for mental health issues: A survey. *Pattern Recognition Letters*, 167:204–211.

Sejnowski, T. J. (2023). Large language models and the reverse turing test. *Neural Computation*, 35:309–342.

Xu, X., Yao, B., and Dong, Y. (2024). Mental-llm: Leveraging large language models for mental health prediction via online text data. In *Proceedings of ACM Interact*.

5 Conclusion

article lipsum

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

References

Arriba-Pérez, F. and García-Méndez, S. (2024). Leveraging llms for real-time mental health predictions. *Arabian Journal for Science and Engineering*.

Gbollie, E. F., Bantjes, J., and Jarvis, L. (2023). Intention to use digital mental health solutions. *Digital Health*, 9:1–19.

Greco, C. M., Simeri, A., and Tagarelli, A. (2023). Transformer-based language models for mental health issues: A survey. *Pattern Recognition Letters*, 167:204–211.

Sejnowski, T. J. (2023). Large language models and the reverse turing test. *Neural Computation*, 35:309–342.

Xu, X., Yao, B., and Dong, Y. (2024). Mental-llm: Leveraging large language models for mental health prediction via online text data. In *Proceedings of ACM Interact*.