AI-based psychoeducation and emotion regulation applications - progress and challenges

Adam Wściubiak,
Volha Tsekalo

Mental disorders account for nearly 30% of non-fatal global health problems, with limited access to care leaving around 70% of those in need without professional support. High costs and underfunded public services make it increasingly challenging to meet the rising demand for individualized psychotherapy, underscoring the urgent need for scalable, innovative mental health solutions alongside traditional therapeutic approaches.

The rapid expansion of mental health applications reflects this demand, yet both deterministic and AI-driven solutions face significant challenges. Pre-programmed interventions offer limited personalization and context-specific adaptability. While these tools can effectively teach specific techniques, most users require broader, more dynamic support. Large Language Models (LLMs) offer a promising alternative by analyzing users' mental states in real-time and tailoring interventions accordingly. Research suggests AI-powered systems can detect affective disorders, recognize stressors, and provide emotional support. However, concerns remain regarding AI's ability to replicate therapeutic relationships and navigate the complex mechanisms of psychotherapy.

Compounding these challenges is widespread "AI hype," amplified by the anthropomorphic design of LLMs that use human-like language and first-person pronouns. While such features can boost user engagement, they may also foster misconceptions about AI's capabilities, potentially leading to unwarranted emotional attachments. Some experts argue that, if paired with user education and safety measures, anthropomorphization can be beneficial in specialized contexts. Since AI is here to stay and humans naturally anthropomorphize, the key is to manage these tendencies responsibly.

Ensuring safety and ethical accountability in AI demands thoughtful model design and oversight. Techniques like pre-training and fine-tuning on specialized data or Reinforcement Learning with Human Feedback (RLHF) help align AI with ethical standards and user well-being -particularly when mental health professionals inform the feedback process. Instruction tuning further structures AI behaviour, while few-shot and zero-shot learning enhances the system's adaptability. Robust safeguards and user education can mitigate risks and foster healthy therapeutic relationships with ethically developed mental health chatbots.

A notable example is the Wysa chatbot—launched in 2018 and recently recognized by the U.S. Food and Drug Administration (FDA) for delivering clinically effective cognitive-behavioural therapy for individuals struggling with pain. Initially reliant on pre-programmed scenarios, Wysa integrated ChatGPT in 2023, introducing both benefits and

risks for AI-driven therapy. Whether these have been properly resolved remains open to debate.

AI-driven mental health solutions hold immense potential. Given the limited access to human-based psychotherapy, responsibly developed AI systems could offer a valuable supplement. The challenge now is to ensure these systems are not only effective but also safe, ethical, and aligned with best practices in psychological care.