VIEWPOINT

Global Mental Health Services and the Impact of Artificial Intelligence–Powered Large Language Models

Alastair C. van Heerden, PhD
Center for Community Based Research, Human Sciences Research Council, Pietermaritzburg, South Africa; and SAMRC/Wits Developmental Pathways for Health Research Unit, University of the Witwatersrand, Johannesburg, South Africa.

Julia R. Pozuelo, PhD
Department of Global Health and Social Medicine, Harvard Medical School, Harvard University, Boston, Massachusetts; and Department of Psychiatry, University of Oxford, Oxford, United Kingdom.

Brandon A. Kohrt, MD, PhD
Center for Global Mental Health Equity, Department of Psychiatry and Behavioral Sciences, George Washington School of Medicine and Health Sciences, Washington, DC.

Corresponding Author: Alastair C. van Heerden, PhD, Human Sciences Research Council, Old Bus Depot, Pietermaritzburg 3201, South Africa (avanheerden@hsrc.ac.za).

There is a large and growing need for mental health services worldwide, but there is a massive shortage of mental health specialists to meet these needs—particularly in humanitarian emergencies, low-income countries, and other areas with limited resources. One strategy that has emerged to address treatment gaps is to rely on nonspecialists (eg, lay health workers, teachers, social workers, and peer mentors) to provide mental health services. Although this approach can be effective, current strategies demand substantial training and supervision.[1] They also require highly standardized interventions, which may paradoxically limit more person-centered treatments.[2] Concurrently, the field of artificial intelligence (AI) is evolving rapidly and changing how we detect and treat mental health disorders. Artificial intelligence applications in psychiatry are varied and include developing prediction models for disease detection and prognosis, creating algorithms that can help clinicians choose the right treatment plan, monitoring patient progress based on data from wearable devices, building chatbots that deliver more personalized and timely interventions, and using AI techniques to analyze therapy session transcripts to improve treatment fidelity and quality.[3-5]

One breakthrough that may be particularly relevant for global mental health is the development of autoregressive large language models (LLMs), such as GPT (Generative Pretrained Transformer)[6] and BERT (Bidirectional Encoder Representations from Transformers).[7] These models use deep learning algorithms trained on big data sets scraped from the internet to predict the next word or sequence of words in a given text based on a prompt or question. Large language models are proving surprisingly capable on various tasks, including the ability to generate long, coherent, and convincing text that seems close to human quality.[8] As data for training and computation power continue to increase, these models will become even more powerful. Given current trends, it seems possible that LLM-based agents could be fine-tuned on digitized texts in psychology and psychiatry, including textbooks and manuals, alongside many years' worth of therapeutic transcripts, to offer an inexpensive tool capable of delivering complex and tailored therapeutic models with high fidelity, compassion, and perfect recall that can engage with thousands of clients simultaneously.

The potential of LLMs for improving mental health services raises many questions. Might LLM-enhanced therapy be more useful or acceptable in certain settings? Might certain psychological therapeutic techniques, such as cognitive behavioral therapy and guided self-help, align with LLMs more readily than other approaches? Who is ethically responsible for making clinical decisions in situations where AI replaces human decision-making? One of the most important questions is how LLMs could contribute to bridging the gap between mental health needs and available services in the settings and populations with the greatest dearth of specialists: low-income countries.

The unavailability of supervision has been a major bottleneck in expanding mental health services in low-income countries, whether that be supervision of primary care physicians taking on diagnosis and pharmacological treatment of mental health conditions or community health workers delivering psychological services. Although supervision is a critical component of service delivery, it is time and resource intensive as it involves regular meetings and continuous review of audio-recorded therapy sessions against checklists to ensure competency and fidelity. In many global programs, nonspecialists receive brief training but no supervision or very limited ongoing support. Given the shortage of trained supervisors, these train-and-hope models (train nonspecialists, then hope everything works out) are often the only alternative available in low-resource settings.

Large language models could help reduce this bottleneck by supporting the training and supervision of the human workforce. For example, LLMs could act as clients with whom nonspecialists could practice their skills, provide nonspecialists with customized learning materials, review session transcripts, and provide feedback based on competency rating tools.[9] In other words, LLMs could assist nonspecialists in acquiring foundational knowledge in an engaging, tailored, and interactive manner. This strategy would free up specialists and allow them to focus on more complicated clinical supervision challenges and expand the human workforce for delivering quality mental health care. In addition, LLMs could also help with language translation, especially with transcription and automatic qualitative coding emerging rapidly. Although LLMs, as they currently stand, cannot be comprehensive replacements for specialists, they may help to ensure the competence of nonspecialists in settings in which supervisors are simply not available.

One known limitation of current LLMs is their propensity to output incorrect or fabricated information. This can be particularly problematic when the language used by the LLM is confident and assertive, which can easily fool an unsuspecting user. Therefore, it remains important for health care specialists to vet and take responsibility for the use of such information in the health care setting to ensure that patients receive accurate and reliable information.

Careful consideration, however, is needed when deploying these tools in low-resource settings, as their use is subject to a range of technological, logistical, and cultural challenges, such as limited internet connectivity, data privacy and security, cultural mistrust, and cultural relevance of therapy materials. Addressing these challenges is essential to ensure equity, ethical standards, and effectiveness in the implementation of LLMs in low- and middle-income countries. Furthermore, a collaboration between relevant stakeholders, such as health systems, technology providers, and communities, is imperative.

There is an urgent need to continue discussing and developing guidelines for the use of LLMs that take into consideration the needs and conditions of the population being served, the availability and quality of existing mental health resources, and the ethical and technical implications of using LLMs in therapy. Ethical frameworks and human oversight are essential to ensure AI-powered systems are used appropriately, and technical approaches, such as explainable AI,[10] may be useful to these endeavors. Another challenge that needs addressing, as acknowledged by both critics and developers of LLMs, is the perpetuation of existing cultural, gender, and language biases. The models are no better than the global pool of text on which they are based, and this pool is rooted in male, Western, colonial biases, thus generally perpetuating a homogeneous view of what a healthy psyche looks like and how to achieve it. Thus, it is crucial that those in the field of global mental health expand data on populations around the world in partnership with people with lived experience of mental health conditions in these settings. Tackling the mental health inequities for LGBTQ+ persons, historically minoritized groups, and those whose voices have traditionally been silenced will require that LLMs are built and refined with more content and guidance from these groups. In addition, the diversity of the world's languages needs to be considered when tailoring how LLMs are used for locally led mental health care; otherwise, the biases of English-speaking WEIRD (Western, educated, industrialized, rich, and democratic) populations will continue to dominate care models and services. We postulate that the low-hanging fruit in this discussion is to use LLMs to address the lack of specialist trainers and supervisors. We imagine a world in which the human nonspecialist workforce expands while LLMs take over some roles of specialists as supervisors and trainers of this human workforce.

Large language models and other forms of AI will fundamentally change how we treat mental disorders, allowing us to move away from the current model in which most of the world's population does not have access to quality mental health services. Instead, the advances can diversify how care is provided, where, and by whom, and assure greater quality to improve lives around the world. To realize this potential, while reducing the potential for harm, substantial effort should go toward the careful and thoughtful introduction of these AI technologies into the field of global mental health.

ARTICLE INFORMATION

Published Online: May 17, 2023. doi:10.1001/jamapsychiatry.2023.1253

Conflict of Interest Disclosures: Dr Pozuelo reported receiving grants from MRC Newton Fund, Wellspring Philanthropic Foundation, and National Institute of Mental Health. No other disclosures were reported.

REFERENCES

1. van Ginneken N, Chin WY, Lim YC, et al. Primary-level worker interventions for the care of people living with mental disorders and distress in low- and middle-income countries. *Cochrane Database Syst Rev*. 2021;8(8):CD009149.

2. Patel V. Scaling up person-centered psychosocial interventions: global mental health's next challenge. *SSM Ment Health*. 2022;2(100072). doi:10.1016/j.ssmmh.2022.100072

3. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med*. 2019; 49(9):1426-1448. doi:10.1017/S0033291719000151

4. Kessler RC, Luedtke A. Pragmatic precision psychiatry—a new direction for optimizing treatment selection. *JAMA Psychiatry*. 2021;78(12): 1384-1390. doi:10.1001/jamapsychiatry.2021.2500

5. Ewbank MP, Cummins R, Tablan V, et al. Quantifying the association between psychotherapy content and clinical outcomes using deep learning. *JAMA Psychiatry*. 2020;77(1):35-43. doi:10.1001/jamapsychiatry.2019.2664

6. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. *arXiv*. Posted online May 28, 2020. Updated July 22, 2020. https://arxiv.org/abs/2005.14165

7. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv*. Posted online October 11, 2018. Updated May 24, 2019. doi:10.48550/arXiv.1810.04805

8. Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models. *arXiv*. Posted online June 15, 2022. Updated October 26, 2022. doi:10.48550/arXiv.2206.07682

9. Kohrt BA, Jordans MJD, Rai S, et al. Therapist competence in global mental health: development of the ENhancing Assessment of Common Therapeutic factors (ENACT) rating scale. *Behav Res Ther*. 2015;69:11-21. doi:10.1016/j.brat.2015.03.009

10. Doran D, Schulz S, Besold TR. What does explainable AI really mean? A new conceptualization of perspectives. *arXiv*. Posted online October 2, 2017. doi:10.48550/arXiv.1710.00794