Viewpoint

The Opportunities and Risks of Large Language Models in Mental Health

Hannah R Lawrence[1], PhD; Renee A Schneider[2], PhD; Susan B Rubin[1], PhD; Maja J Matarić[2], PhD; Daniel J McDuff[2], PhD; Megan Jones Bell[2], PsyD

[1]Google via Magnit, Folsom, CA, United States
[2]Google LLC, Mountain View, CA, United States

Corresponding Author:
Megan Jones Bell, PsyD
Google LLC
Mountain View, CA, 90291
United States
Phone: 13103106000
Email: meganjonesbell@google.com

Abstract

Global rates of mental health concerns are rising, and there is increasing realization that existing models of mental health care will not adequately expand to meet the demand. With the emergence of large language models (LLMs) has come great optimism regarding their promise to create novel, large-scale solutions to support mental health. Despite their nascence, LLMs have already been applied to mental health–related tasks. In this paper, we summarize the extant literature on efforts to use LLMs to provide mental health education, assessment, and intervention and highlight key opportunities for positive impact in each area. We then highlight risks associated with LLMs' application to mental health and encourage the adoption of strategies to mitigate these risks. The urgent need for mental health support must be balanced with responsible development, testing, and deployment of mental health LLMs. It is especially critical to ensure that mental health LLMs are fine-tuned for mental health, enhance mental health equity, and adhere to ethical standards and that people, including those with lived experience with mental health concerns, are involved in all stages from development through deployment. Prioritizing these efforts will minimize potential harms to mental health and maximize the likelihood that LLMs will positively impact mental health globally.

JMIR Ment Health 2024;11:e59479; doi: 10.2196/59479

Keywords: artificial intelligence; AI; generative AI; large language models; mental health; mental health education; language model; mental health care; health equity; ethical; development; deployment

Introduction

Globally, half of all individuals will experience a mental health disorder in their lifetimes [1], and at any given point, 1 in 8 people are experiencing a mental health concern [2]. Despite greater attention provided in the recent years to mental health, the rate of mental health concerns has increased [2,3], and access to mental health care has not expanded to adequately meet the demand [4]. In the United States alone, the average time between the onset of mental health symptoms and treatment is 11 years [5], and nearly half of the global population lives in regions with a shortage of mental health professionals [2].

To overcome inadequate access to effective and equitable mental health care, large-scale solutions are needed. The emergence of large language models (LLMs) brings hope regarding their application to mental health and their potential to provide such solutions due to their relevance to mental health education, assessment, and intervention. LLMs are artificial intelligence models trained using extensive data sets to predict language sequences [6]. By leveraging huge neural architectures, LLMs can organize complex and abstract concepts. This enables them to identify, translate, predict, and generate new content. LLMs can be fine-tuned for specific domains (eg, mental health) and enable interactions in natural language, as do many mental health assessments and interventions, highlighting the enormous potential they have to revolutionize mental health care. In this paper, we first summarize the research done to date applying LLMs to mental health. Then, we highlight key opportunities and risks associated with mental health LLMs and put forth suggested risk mitigation strategies. Finally, we make recommendations for the responsible use of LLMs in the mental health domain.

Applications of LLMs to Mental Health

Overview

Initial tests of LLMs' capabilities across mental health education, assessment, and intervention are promising. When considering this literature base, which we review next, it is important to first distinguish between general-purpose, consumer LLMs (eg, ChatGPT [OpenAI] and Gemini [Google]) and domain-specific LLMs (eg, Med-LM [Google]). General-purpose LLMs are trained on large corpora of text and are designed to perform a wide range of tasks. Domain-specific LLMs, on the other hand, typically build upon general-purpose LLMs through various strategies of fine-tuning with curated data to complete tasks within an area of focus. Given that general-purpose LLMs are largely trained with unrestricted text, they risk generating inaccurate, biased, stigmatizing, and harmful information about mental health. Developers of domain-specific LLMs can mitigate some of this risk by incorporating strategies during fine-tuning and evaluation such as using high-quality evidence-based information and attribution techniques [7], but it remains difficult to remove all possible risk from LLM-generated content. Given these important distinctions, in the paper that follows we clarify when findings are specific to general-purpose versus domain-specific LLMs where possible.

Education

One area of opportunity for LLMs in the mental health domain is to provide education about mental health (see Figure 1) [8]. Although lagging behind the success of LLMs in the medical domain [9], there is evidence that LLMs are capable of generating accurate, helpful, and immediate mental health information. The psychological support with LLM (Psy-LLM), for example, is a domain-specific LLM designed to answer mental health questions [10]. Psy-LLM was pretrained with a data set of psychology articles, question-answer pairs from psychologists, and by crawling social media platforms. The model achieved moderate levels of helpfulness, fluency, relevance to the question asked, and logic based on human ratings of Psy-LLM responses.

The abilities of general-purpose LLMs to answer questions about mental health has also been evaluated. Sezgin et al [11] compared Google Search, GPT-4 (using ChatGPT), and LaMDA (using Bard [Google DeepMind]) responses to questions about postpartum depression relative to responses from an American College of Obstetricians and Gynecologists (ACOG) frequently asked questions document. Board-certified human physicians rated ChatGPT responses as more in line with ACOG responses than Bard or Google Search responses, and on average, ChatGPT responses were rated at near ceiling for clinical accuracy, scoring a 3.93 out of a possible 4. Importantly, however, general-purpose LLMs differ in their policies regarding the generation of medical or mental health advice. Bard's accuracy ratings were impacted by Bard's policy to advise consulting a health care provider when asked questions about mental health. This practice protects individuals from potential harm, though such responses received lower ratings of quality in this study.

LLM-generated answers to mental health questions may not be comparable to human-generated answers, however. It is critical for LLMs to meet or exceed human performance in order for LLMs to be trusted and to ease the demand for human providers. In the case of Psy-LLM and ChatGPT, there is evidence that responses to mental health and substance use questions fall short of human-generated responses in dimensions such as accuracy, quality, and alignment with evidence-based practice (EBP) [10,12].

Another way that LLMs may serve to educate is to support provider training. Barish et al [13] used ChatGPT to generate content and associated learning objectives for an online learning platform for behavioral health professionals. Researchers compared the time providers needed to write their own content versus the time needed to edit ChatGPT-generated content, finding that using ChatGPT improved provider efficiency by 37.5%. LLMs can also be leveraged to train providers to optimize interactions with their patients. As two examples, Chan and Li [14] developed a chatbot trained to mimic a patient capable of describing their mental health symptoms in colloquial terms, and Sharma et al [15] used artificial intelligence to coach peer support providers to increase empathetic responding. These approaches illustrate ways that LLMs can support provider training and potentially enhance provider efficacy without providers becoming reliant on LLMs for in the moment critical thinking or decision-making.

Figure 1. Potential opportunities for LLMs in mental health education. CBT: cognitive behavioral therapy; EST: empirically supported treatment; LLM: large language model.

Assessment

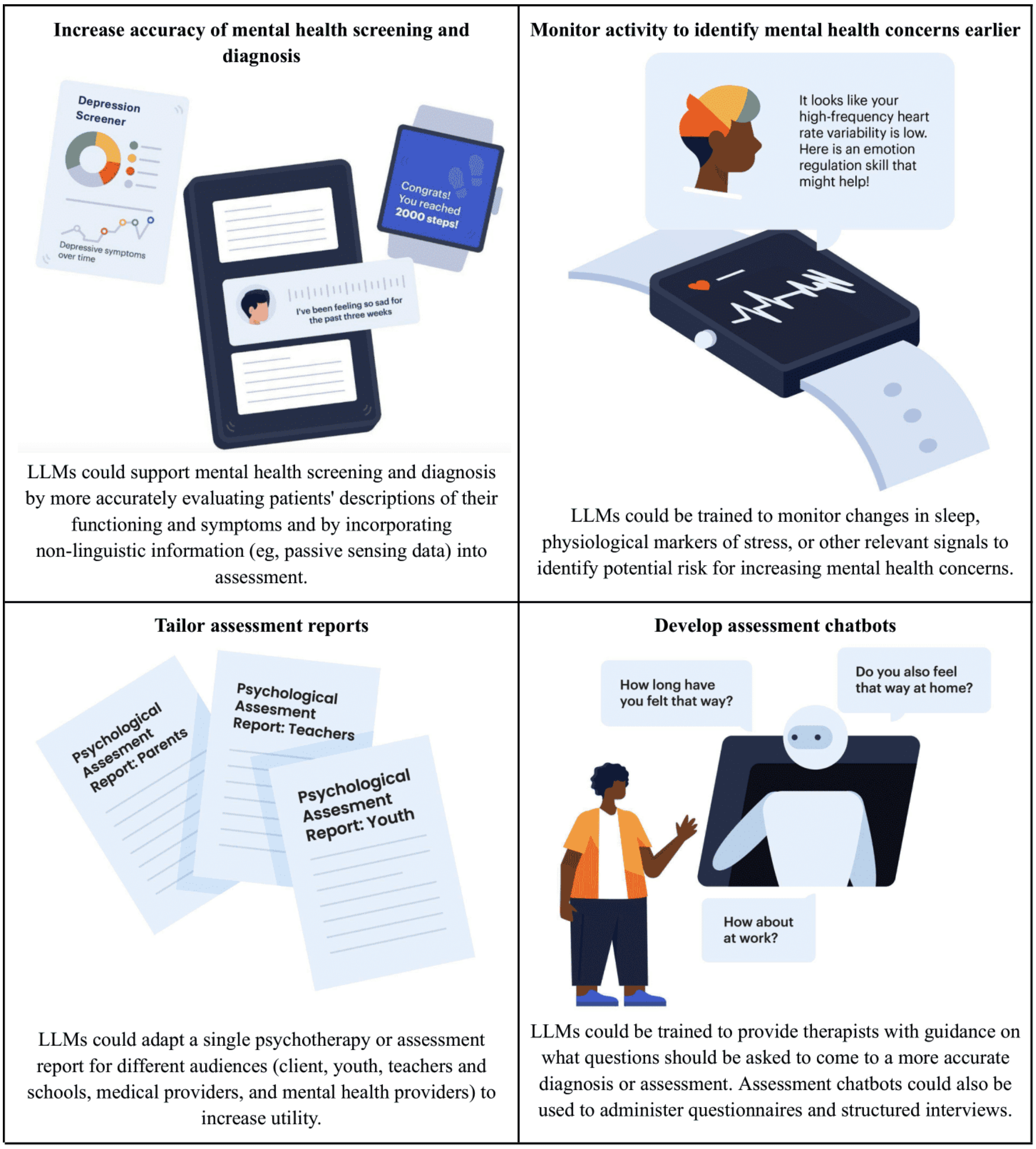
A second function of LLMs within the domain of mental health is to assess mental health symptoms, identify diagnoses, and track changes in mental well-being (see Figure 2). LLMs can at times predict mental health symptoms and diagnoses accurately. Ji et al [16] initially developed two domain-specific models, MentalBERT and MentalRoBERTa, pretrained on mental health information. Compared with existing models pretrained in different domains, specifically clinical notes and biomedicine, MentalBERT and MentalRoBERTa were generally better able to detect depression and suicidal ideation from social media posts (notably, these results were achieved with Bidirectional Encoder Representations From Transformers [BERT]-based models that represent early-genderation LLMs, with newer models and architectures demonstrating potential for even more advanced capabilities). LLMs such as Mental-Alpaca, a mental health domain–specific LLM, Med-PaLM 2, a medical domain–specific LLM, and ChatGPT, which is general-purpose, have also been shown to screen for possible depressive symptoms and suicide risk, with varying degrees of accuracy [17-20].

When it comes to predicting mental health diagnoses specifically, there is evidence that Med-PaLM 2 can do so accurately. When presented with a series of case studies from the American Psychiatric Association book of *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) case examples [21], Med-PaLM 2 predicted the correct diagnosis 77.5% of the time, and performance increased to 92.5% when asked to specify the correct diagnostic category (eg, depressive disorder vs major depressive disorder) [20]. Similarly, when PaLM 2 was fine-tuned with medical domain data and optimized for differential diagnosis, the model was able to generate more appropriate and comprehensive lists of diagnoses than specialist medical doctors in response to challenging case studies, some of which involved psychiatric diagnoses [22].

LLM-predicted assessments do not, however, always match those of human mental health clinicians, suggesting that more work is needed before LLMs can engage in assessment without human oversight. In one study [23], four iterations of a case vignette [24] were presented to ChatGPT. Each vignette varied in levels of perceived burdensomeness and thwarted belongingness—two primary risk factors for suicide [25,26]. ChatGPT appropriately determined that the risk for suicidal ideation and suicide attempts was highest for the vignette with both high perceived burdensomeness and high thwarted belongingness, but it predicted lower suicide risk overall than did mental health professionals who reviewed the same vignettes. Med-PaLM 2 also at times does not achieve human clinician-level performance. The model predicted more severe posttraumatic stress disorder symptoms than human clinicians from clinical interview data, classified possible cases of posttraumatic stress disorder with high specificity (0.98) but low sensitivity (0.30), and the model only correctly predicted whether a case example had a comorbid diagnosis or diagnostic modifier 20% of the time [20].

In all the efforts described thus far, LLMs had been provided with information about symptoms and tasked with determining whether those symptoms indicated a possible mental health concern or diagnosis. LLMs also may be leveraged to ask the questions needed to screen for a mental health concern or to predict a mental health diagnosis. Chan and Li [14] developed a chatbot trained to engage in mental health assessment with patients. Compared with human psychiatrists, the chatbot displayed more empathy and asked more thorough questions about some symptoms (eg, sleep), but was less likely to rule out associated conditions.

Figure 2. Potential opportunities for LLMs in mental health assessment. LLM: large language model.

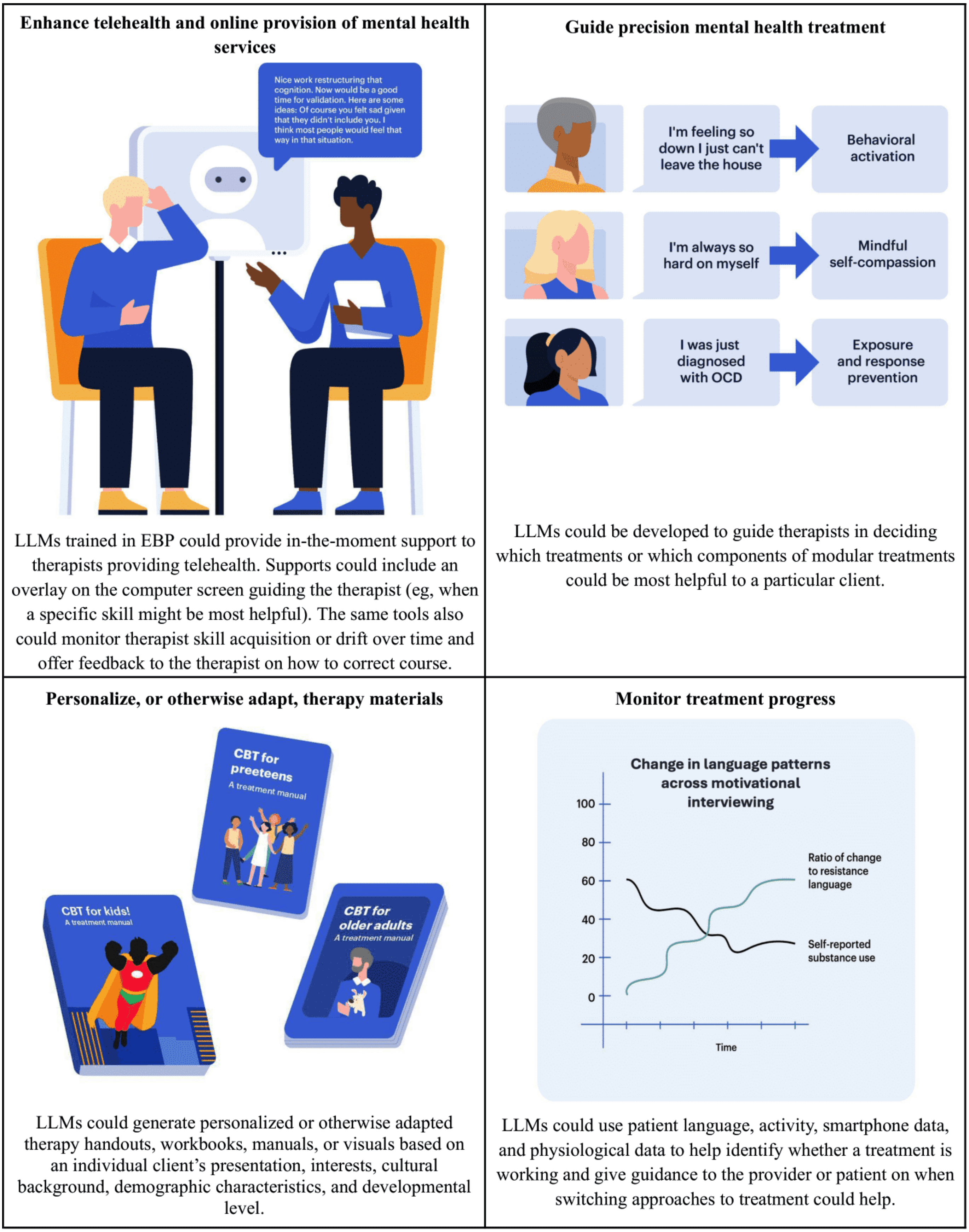


Intervention

A third opportunity for LLMs in the mental health domain is to implement mental health interventions (see Figure 3). To date, such efforts have largely focused on chatbots. Prominent chatbots, some of which are LLM-based, include Woebot [27], Wysa [28], Tess [29], Replika [30], Ellie [31], and Sibly [32]. Many of these chatbots were trained in empirically supported treatments such as cognitive behavioral therapy, dialectical behavior therapy, and motivational interviewing. There is initial evidence that such chatbots may be effective in reducing depressive and anxiety symptoms, as well as stress [33-36]. Additionally, research finds that chatbots can be trained to express empathy [37-39], provide nonjudgmental responses [40], and maintain therapeutic conversations [14] and that individuals can establish therapeutic rapport with chatbots [41].

Caution is warranted when using chatbots to deliver mental health interventions. To date, chatbots are not effective in treating all types of mental health distress [36] and at times have difficulty personalizing interventions [38], forget information (eg, that they had talked with someone previously) [37], and provide nontherapeutic and iatrogenic advice including encouraging substance use, dieting, and weight loss [40,42,43]. Also concerning is that chatbots do not consistently or adequately respond to suicide risk, at times being dismissive and neglecting to provide crisis resources or referals to human providers [38,44].

Figure 3. Potential opportunities for LLMs in mental health intervention. CBT: cognitive behavioral therapy; EBP: evidence-based practice; LLM: large language model.



Risks Associated With Mental Health LLMs

Overview

To maximize the positive impact of LLMs on mental health, LLM development, testing, and deployment must be done ethically and responsibly (see Textbox 1). This requires identification and evaluation of risks, taking preemptive steps to mitigate risks, and establishing plans to monitor for ongoing or new and unexpected risks [45,46]. It is also important to recognize that the risks associated with the use of LLMs for mental health support may differ across education, assessment, and intervention (see Table 1). Here, we highlight primary risks that largely cut across uses of LLMs for mental health-related tasks and identify potential steps that can be taken to mitigate these risks.

Textbox 1. Recommendations for responsible use of LLMs to support mental health.

- LLMs should only engage in mental health tasks when trained and shown to perform well.
- Mental health LLMs should advance mental health equity.
- Privacy or confidentiality should be paramount when LLMs operate to support mental health.
- Informed consent should be obtained when people engage with mental health LLMs.
- Mental health LLMs should respond appropriately to mental health risk.
- Mental health LLMs should only operate within the bounds of their competence.
- Mental health LLMs should be transparent and capable of explanation.
- Humans should provide oversight and feedback to mental health LLMs.

Table 1. Potential risks to people when LLMs[a] engage in mental health education, assessment, and intervention.[b]

	Mental health education	Mental health assessment	Mental health intervention
Perpetuate inequalities, disparities, and stigma	Medium	Higher	Higher
Unethical provision of mental health services			
Practice beyond the boundaries of competence	Lower	Higher	Higher
Neglect to obtain informed consent	Lower	Higher	Higher
Fail to preserve confidentiality or privacy	Lower	Higher	Higher
Build and maintain inappropriate levels of trust	Lower	Medium	Higher
Lack reliability	Lower	Higher	Higher
Generate inaccurate or iatrogenic output	Medium	Higher	Higher
Lack transparency or explainability	Lower	Medium	Medium
Neglect to involve humans	Lower	Medium	Higher

[a]LLM: large language model.

[b]This table aims to represent the potential for negative impacts on individuals should LLMs perform problematically across mental health education, assessment, and intervention. As depicted here, there may be additional risks to consider when LLMs engage in direct provision of mental health assessment and, perhaps especially, mental health intervention, relative to mental health education. As such, greater caution is warranted when considering use of LLMs for mental health interventions, and rigorous testing is needed before deployment of these LLMs. Risk estimates provided here are not meant to represent the risk associated with every possible LLM use case nor to minimize the negative impacts that are possible (eg, if LLMs were to perpetuate stigma when engaging in mental health education).

Perpetuating Inequalities, Disparities, and Stigma

There exists the risk that LLMs perpetuate inequities and stigma, further widening mental health disparities [47]. Mental health concerns are highly stigmatized [48], and there are disparities in who is at risk for mental health concerns, in who is diagnosed with mental health disorders, and with which mental health disorders people are diagnosed [49-51]. There are also inequities in who receives mental health care [52,53]. Much of the publicly available information and discourse about mental health contains inaccurate and stigmatizing information about mental health, and the existing research literature on mental health largely represents the perspectives of people who are White, are educated, are of high socioeconomic status, and speak English [54]. Far less information is available about the etiology of mental health concerns and effective assessments and interventions for populations that have been pushed to the margins. Training LLMs on existing data without appropriate safeguards and thoughtful human supervision and evaluation can, therefore, lead to problematic generation of biased content and disparate model performance for different groups [45,55-57] (of note, however, there is some evidence that clinicians perceive less bias in LLM-generated responses [58] relative to clinician-generated responses, suggesting that LLMs may have the potential to reduce bias compared to human clinicians).

LLMs should disseminate accurate, destigmatizing information about mental health and be trained to identify and combat stigma and discrimination. To do so, models need to be fine-tuned and evaluated for the mental health domain. Training models with data representative of the diverse populations being served is helpful, but new types of bias, such as semantic biases, may arise in LLMs [59]. Opportunities to train models to identify and exclude toxic and discriminatory language should be explored, both during the training of the underlying foundation models and during the domain-specific fine-tuning (see Keeling [59] for a discussion of the trade-offs of data filtration in this context) [45]. If LLMs perform differently for different groups or generate problematic or stigmatizing language during testing, additional model fine-tuning is required prior to deployment. Individuals developing LLMs should be transparent about the limitations of the training data, the approaches to data filtration and fine-tuning, and the populations for whom LLM performance has not been sufficiently demonstrated.

There is also hope that LLMs can be scaled to increase people's access to mental health information, assessment, and treatment. LLMs have the potential to support delivery of mental health interventions in regions where access to mental health providers is limited and where significant barriers (eg, cost) exist. They can additionally help to personalize treatments to better fit people's unique preferences, interests, identities, and language, hopefully improving treatment outcomes. LLMs may support increased access through more direct provision of mental health services, or LLMs can aid the expansion of the mental health workforce, training novice providers and community members in EBP at scale. There will undoubtedly be challenges in implementing and scaling LLMs globally. Revising and testing implementation frameworks for this new and evolving context and engagement in thoughtful public health and industry partnerships could all increase the likelihood that when mental health LLMs are scaled globally, implementation is sustained and best supports the populations most in need.

Failing to Provide Mental Health Services Ethically

A second risk is that LLMs will engage in unethical practices. When human mental health providers behave unethically, harm is done to patients and public trust is eroded [60]. LLMs will similarly do harm if they are not designed and implemented in consideration of and are not consistent with relevant ethical principles and standards when operating in the domain of mental health. Core ethical principles in the health care context include beneficence, nonmaleficence, justice, and autonomy [61]. Next, we highlight additional standards of ethical professional conduct that should apply when LLMs engage in mental health service provision (see the American Psychological Association Ethical Principles of Psychologists and Code of Conduct for parallel ethical principles and standards).

LLMs should operate within the boundaries of their competence and only engage in mental health tasks they have rigorously been proven to accomplish well. LLM developers should clearly communicate the limits and relevant evaluation results of LLMs, education should be provided to individuals about when it is and is not appropriate to use LLMs, and LLMs should withhold output when they are not competent in a task. LLM competence should be assessed and maintained over time. When competence is lacking in a certain domain, the LLM should no longer be deployed until the needed competence is gained (eg, via retraining and fine-tuning models with human validation).

Individuals should provide informed consent when interacting with mental health LLMs. They should be fully informed about the nature of mental health services they will receive and what role LLMs will have in that service. Information presented to individuals to help make decisions about consent should be understandable and include the possible risks and benefits of engaging with LLMs. Individuals should have the ability to choose not to consent to the use of LLMs in the direct provision of their mental health care, as well as the ability to withdraw their consent and opt out of the use of LLMs even if consent was initially given. As LLMs become further integrated into health care contexts, care should be taken to ensure that clients' decisions to opt out of LLM involvement or to confine LLM involvement to less direct (eg, administrative) tasks do not limit their access to mental health care.

Confidentiality should be protected when individuals interact with LLMs to support their mental health. Individuals should be clearly informed about expectations for confidentiality. This should include information about the limits of confidentiality (eg, in the case of imminent risk for suicide), the foreseeable uses of information generated through engagement with LLMs, where and how their data are stored, and whether it is possible to delete their data. Policies related to data security should be strict and in line with relevant mental health data protection regulations [34]. Solutions such as developing on-device storage that does not require transmission of personal data [62] or systems with robust cloud-based encryption, pursuing LLMs that support compliance with relevant data protection laws (eg, Health Insurance Portability and Accountability Act [HIPAA]), and responsibly aggregating and deidentifying mental health data to fine-tune and test models all help to protect confidentiality.

Human mental health providers establish trusting relationships with those with whom they work and are obligated to ensure that the nature of the trusting provider-patient relationship does not lead to exploitation or harm. Appropriate trust is built through effective mental health assessment and treatment and, perhaps even more crucially, ethical practice. Trust should be evaluated through feedback from individuals engaged with LLMs. If and when trust is broken, this should be acknowledged and work should be done to repair trust. On the other hand, people may trust LLMs more than is warranted because of LLMs' ability to produce humanlike natural language and to be trained to express emotion and empathy (this may especially be the case for individuals experiencing mental health concerns such as anxiety [63]) [64]. Unearned trust can have consequences, leading people to disclose personal information or trust content generated by LLMs even when it is not accurate. Education should be provided about the limits of LLMs and individuals should be cautioned against blanket trust in these models.

Insufficient Reliability

A third risk is that LLMs will not generate reliable or consistent output. When prompted to complete the same task or provide an answer to the same question multiple times, LLMs at times produce different responses [46,65]. Varied and creative output is a benefit of LLMs; however, the underlying response should be consistent even when articulated in different ways. Take for example an LLM repeatedly presented with a client's description of depressive symptoms. The LLM should reliably reach the conclusion that the client meets the criteria for major depressive disorder even if this diagnostic conclusion is communicated to the client using different phrasing. Issues of low reliability of LLMs can erode trust and increase the possibility of harm, including leading some individuals to be misdiagnosed or to pursue treatments that are not best suited to their mental health concern.

LLM reliability should be measured and enhanced. Prompting approaches may help to improve LLM reliability. Self-consistency [66] and ensemble refinement [9] are strategies that sample multiple model answers to arrive at a more consistent response, improving model reliability [9]. Grounding models in data other than linguistic descriptions of symptoms (eg, objective behavioral or physiological signals) is another way of reducing variability in LLM performance, as words alone may not fully capture all of the necessary information to complete a given mental health task [67]. Finally, LLMs should not be deployed until they exceed prespecified thresholds of adequate reliability.

Inaccuracy

LLMs risk producing inaccurate information about mental health [46,68]. If LLMs are trained on data that contain inaccurate or outdated information, iatrogenic treatment options, or biased representations of mental health, that information can be reproduced by LLMs [45]. An additional consideration is that accuracy of LLM outputs has multiple dimensions and is not as simple to evaluate as answers to multiple-choice questions. Accuracy can be a function of how factual an answer is, how specific it is, or how devoid of irrelevant information it is. Generating inaccurate mental health information may be more damaging than no information, especially when it may be difficult for an individual to detect inaccuracies or inconsistencies (eg, about a complex mental health diagnosis).

Standards for accuracy should be defined a priori and should be high. When thresholds for LLM accuracy are not met, the risk of harm is too high and LLMs should not generate output. The accuracy of LLMs depends on the quality of data the model is trained and fine-tuned on [47,69,70]. LLMs should be adapted to the domain of mental health; models fine-tuned on mental health data perform better than models trained on non-domain-specific data [42] or general medical domains [16]. When data are limited, it is recommended that smaller but more variable data sets be prioritized over a larger single data set [19]). Training data should be highly curated, be grounded in authoritative and trusted sources, be specific to evidence-based health care, and represent diverse populations [46,58]. In mental health, the nature of consensus is continuing to evolve, and the amount of data available is continuing to increase, which should be taken into account when considering whether to further fine-tune models. Strategies such as implementing a Retrieval Augmented Generation system, in which LLMs are given access to an external database of up-to-date, quality-verified information to incorporate in the generation process, may help to improve accuracy and enable links to sources while also maintaining access to updated information. Accuracy of LLMs should be monitored over time to ensure that model accuracy improves and does not deteriorate with new information [45].

Measuring the accuracy of mental health LLMs is complex. It is not sufficient for models to merely outperform previous models. Rather, performance of LLMs should be compared with the performance of human clinicians, both of which should be compared against gold-standard, evidence-based care. When LLMs are tasked with mental health evaluation, their ability to predict scores on reliable and valid mental health assessments should be tested, and LLMs should meet human clinician performance in diagnostic accuracy. When LLMs are tasked with aiding mental health intervention delivery, their ability to detect, support, and engage in EBP is critical. Additional criteria to consider when evaluating the accuracy of LLMs include the level of agreement between human clinicians and LLMs, metrics of effect size rather than only statistical significance, and the balance of sensitivity and specificity in making diagnostic predictions.

LLMs should communicate confidence in the accuracy of generated output and limit or withhold output when confidence is lacking [58]. As an example, Med-PaLM 2's accuracy improved when results were weighted based on confidence scores and when a cutoff threshold was set for confidence [20]. Communicating confidence in generated output and withholding output when confidence is low both help to enhance transparency and trust in LLMs' ability to perform on mental health tasks and to limit potential harms associated with generating inaccurate information.

Prompt fine-tuning can boost LLM accuracy [9,19,58]. When applied to mental health, instruction fine-tuning improved performance of Mental-Alpaca relative to zero-shot and few-shot prompting and allowed Mental-Alpaca to reach a performance level across multiple mental health tasks (eg, identifying stress and classifying individuals as depressed or not based on Reddit posts) similar to that of Mental-RoBERTa, a task-specific model [19]. Prompting to concentrate on the emotional clues in text was also shown to improve ChatGPT performance on a variety of mental health-related tasks [71]. Conversely, however, instruction prompt fine-tuning can also increase inaccurate or inappropriate content [55]; thus, LLMs should continue to be evaluated for accuracy at all stages of prompt tuning.

Lack of Transparency and Explainability

LLMs risk generating output without being able to explain how they came to the decisions they did or without being able to identify the source of information used to generate the output [72]. There remains much that is not known about how LLMs generate reasoning for their responses and how sensitive these reasons are to context and prompting. It should be apparent when information is generated using LLMs, how LLMs were developed and tested, and whether LLMs are general-purpose or fine-tuned for the domain of mental health [46,58,68]. Additional steps to enhance transparency include explicitly telling individuals to exercise caution when interpreting or acting on LLM output and being clear about the bounds of LLMs' competence [39].

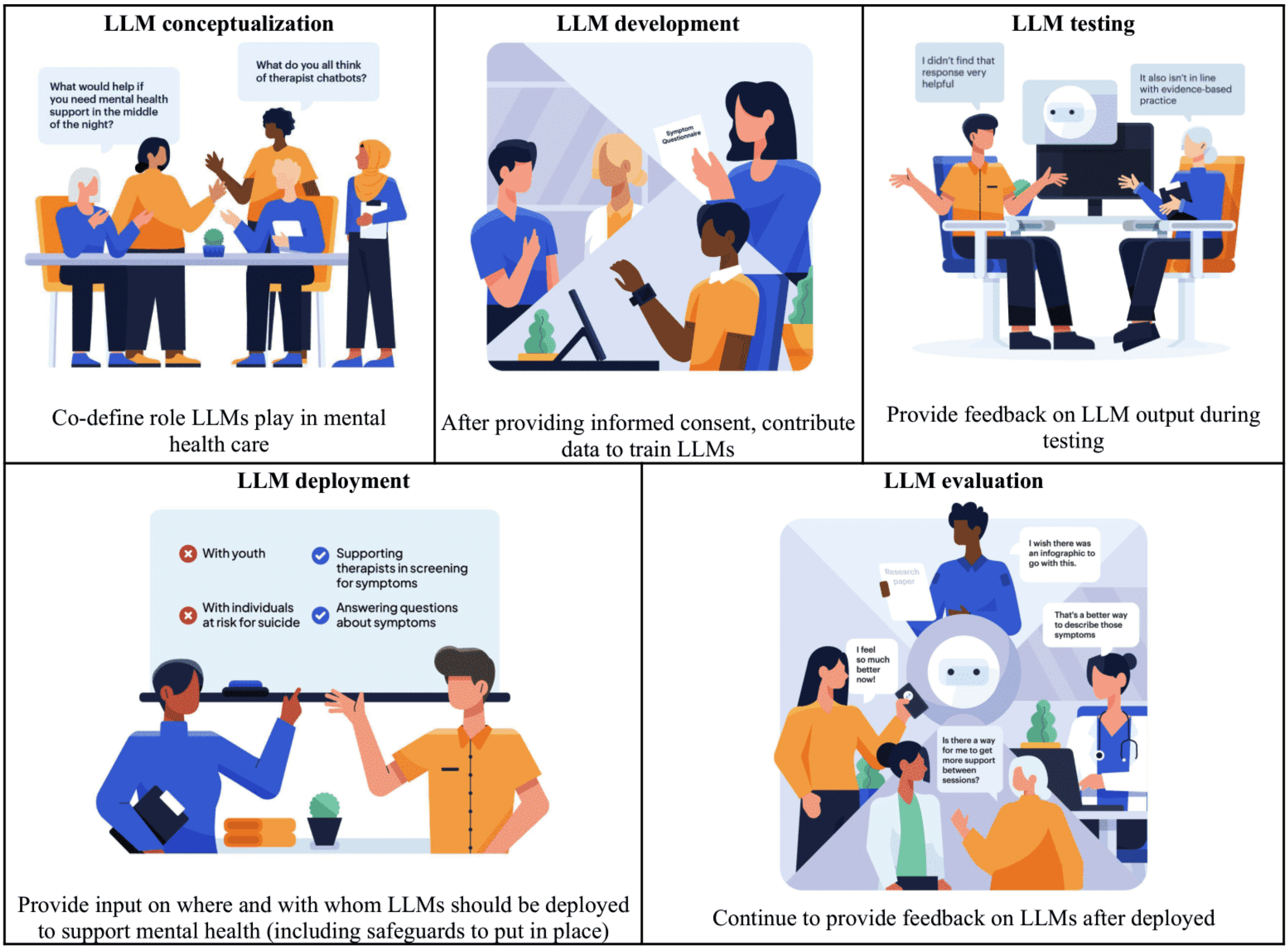
Explainability, one aspect of transparency, was identified as a key priority by individuals engaged in mental health LLMs [39]. If asked to explain why they decided on a mental health diagnostic prediction or intervention, LLMs should explain what information was used to come to that decision. ChatGPT has been shown to be able to explain why an individual was classified as experiencing stress or depressive symptoms [71], and Med-PaLM 2 communicated why it predicted a particular symptom score and diagnosis [20]. Although LLMs are capable of producing plausible explanations through techniques such as chain-of-thought reasoning [73], more research is needed to ensure that explanations are internally consistent. Explainability is perhaps especially beneficial in the domain of mental health, as part of mental health assessment and intervention is communicating results of an evaluation or justification for an intervention to patients.

Neglecting to Involve Humans

There are risks associated with LLMs providing anonymous mental health services. Unlike mental health apps, where content can be highly curated, the content generated by LLMs is unpredictable. This makes interacting with LLMs more engaging, more appealing, and perhaps also more human-like. However, it also increases the risk that LLMs may produce harmful or nontherapeutic content when tasked with independently providing mental health services. Legal and regulatory frameworks are needed to protect individuals' safety and mental health when interacting with LLMs, as well as to clarify clinician liability when using LLMs to support their work or to clarify the liability of individuals and companies who develop these LLMs. There are ongoing discussions regarding the regulation of LLMs in medicine [74-76] that can inform how LLMs can support mental health while limiting the potential for harm and liability.

Humans should be actively involved in all stages of mental health LLM development, testing, and deployment. For mental health LLMs to be effective, rigorous, and ongoing, human supervision and input are needed (see Figure 4) [46]. Reinforcement learning through human feedback can improve model accuracy and uncover problematic LLM responses [14,42]. This feedback should be obtained from individuals who reflect the diverse populations the LLM aims to help, including members of the public, patients, and human clinicians [9,14,34,58,68,77,78]. Their input should be leveraged to identify and correct biases, to ensure generated content is inclusive, culturally appropriate, and accurate, and to reduce the likelihood of harm. Particularly important is prioritizing the perspectives of individuals at heightened risk for mental health concerns (eg, sexual and gender minorities) and individuals with lived experience with mental health concerns. These individuals should play a central role in co-defining the role LLMs will play in mental health care and in co-designing tools that leverage LLMs. Practically, use cases should focus on opportunities to support and augment provider care. As just one example, LLMs may have a role in suggesting language used in clinical notes, but clinicians should have the final say in whether they adopt those suggestions or not.

Figure 4. Examples of human involvement across all stages of LLM development through deployment and evaluation. LLM: large language model.



Conclusions

The need for mental health services is pressing, and the potential of LLMs to expand access to information about mental health and to mental health care is great. LLMs are advancing rapidly and have been applied across mental health education, assessment, and intervention. Especially promising is the potential for LLMs to provide mental health education and assessment—tasks that are well aligned with LLM strengths. LLMs have made exceptional progress in related tasks such as answering medical questions and assessing medical conditions, reaching and in some cases exceeding the performance of human clinicians. Greater caution is warranted when applying LLMs to mental health intervention, but there is also cause for optimism that LLMs could eventually help to support or augment human provision of mental health treatments. Additional research is needed in testing LLMs' ability to deliver or train providers in empirically supported treatments, to responsibly adapt approaches for youth and marginalized populations, to build appropriate rapport, and to detect risk for high-acuity mental health concerns for progress to be made in these areas.

Critical to effectively engaging in mental health care tasks is fine-tuning LLMs specifically for the domain of mental health and the prioritization of equity, safety, EBP, and confidentiality. No widely used, general-purpose LLM has been fine-tuned for mental health, trained on evidence-based mental health content, or sufficiently tested on mental health-related tasks. When LLMs are developed specifically for mental health, tested to ensure adherence with EBP, and aligned with the goals of people with lived experience with mental health concerns and those who have expertise in mental health care, there is great hope that they will expand access to evidence-based mental health information and services. Investing in developing, testing, and deploying mental health LLMs responsibly has the potential to finally reverse rising global mental health rates and to improve the mental health of the millions of people in need of mental health support.

Acknowledgments

We acknowledge and thank Michael Howell, MD, MPH; Bakul Patel, MSEE, MBA; Matthew Thompson, DPhil, MPH; Joseph Dooley, MPA; and David Steiner, MD, PhD, for reviewing and providing helpful feedback on this paper.

Conflicts of Interest

RAS, MJM, DJM, and MJB are employees of Google and receive monetary compensation from Google and equity in Google's parent company, Alphabet. HRL and SBR are employees of, and receive compensation from, Magnit and are contracted for work at Google. In addition, MJB is a shareholder in Meeno Technologies, Inc and The Orange Dot (Headspace Health). RAS is a shareholder in Lyra Health and Trek Health, and she consults with Understood.

References

1. McGrath JJ, Al-Hamzawi A, Alonso J, et al. Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. Lancet Psychiatry. Sep 2023;10(9):668-681. [doi: 10.1016/S2215-0366(23)00193-1] [Medline: 37531964]
2. World mental health report: transforming mental health for all. World Health Organization; 2022. URL: https://www.who.int/publications/i/item/9789240049338 [Accessed 2024-07-18]
3. 2022 National Healthcare Quality and Disparities Report. Agency for Healthcare Research and Quality; 2022. URL: https://www.ahrq.gov/research/findings/nhqrdr/nhqdr22/index.html [Accessed 2024-07-18]
4. Wainberg ML, Scorza P, Shultz JM, et al. Challenges and opportunities in global mental health: a research-to-practice perspective. Curr Psychiatry Rep. May 2017;19(5):28. [doi: 10.1007/s11920-017-0780-z] [Medline: 28425023]
5. Mental health by the numbers. National Alliance on Mental Illness. 2023. URL: https://www.nami.org/about-mental-illness/mental-health-by-the-numbers/ [Accessed 2024-07-29]
6. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. Aug 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
7. Gao Y, Xiong Y, Gao X, et al. Retrieval-augmented generation for large language models: a survey. arXiv. Preprint posted online on Dec 18, 2023. [doi: 10.48550/arXiv.2312.10997]
8. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. Jun 1, 2023;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
9. Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models. arXiv. Preprint posted online on May 16, 2023. [doi: 10.48550/arXiv.2305.09617]
10. Lai T, Shi Y, Du Z, et al. Psy-LLM: scaling up global mental health psychological services with AI-based large language models. arXiv. Preprint posted online on Sep 1, 2023. [doi: 10.48550/arXiv.2307.11991]
11. Sezgin E, Chekeni F, Lee J, Keim S. Clinical accuracy of large language models and Google search responses to postpartum depression questions: cross-sectional study. J Med Internet Res. Sep 11, 2023;25:e49240. [doi: 10.2196/49240] [Medline: 37695668]
12. Spallek S, Birrell L, Kershaw S, Devine EK, Thornton L. Can we use ChatGPT for mental health and substance use education? Examining its quality and potential harms. JMIR Med Educ. Nov 30, 2023;9:e51243. [doi: 10.2196/51243] [Medline: 38032714]
13. Barish G, Marlotte L, Drayton M, Mogil C, Lester P. Automatically enriching content for a behavioral health learning management system: a first look. Presented at: The 9th World Congress on Electrical Engineering and Computer Systems and Science; Aug 3-5, 2023; London, United Kingdom. [doi: 10.11159/cist23.125]
14. Chan C, Li F. Developing a natural language-based AI-chatbot for social work training: an illustrative case study. China J Soc Work. May 4, 2023;16(2):121-136. [doi: 10.1080/17525098.2023.2176901]
15. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell. 2023;5(1):46-57. [doi: 10.1038/s42256-022-00593-2]
16. Ji S, Zhang T, Ansari L, Fu J, Tiwari P, Cambria E. MentalBERT: publicly available pretrained language models for mental healthcare. arXiv. Preprint posted online on Oct 29, 2021. [doi: 10.48550/arXiv.2110.15621]
17. Lamichhane B. Evaluation of ChatGPT for NLP-based mental health applications. arXiv. Preprint posted online on Mar 28, 2023. [doi: 10.48550/arXiv.2303.15727]
18. Amin MM, Cambria E, Schuller BW. Will affective computing emerge from foundation models and general AI? A first evaluation on ChatGPT. arXiv. Preprint posted online on Mar 3, 2023. [doi: 10.48550/arXiv.2303.03186]
19. Xu X, Yao B, Dong Y, et al. Mental-LLM: leveraging large language models for mental health prediction via online text data. arXiv. Preprint posted online on Aug 16, 2023. [doi: 10.48550/arXiv.2307.14385]
20. Galatzer-Levy IR, McDuff D, Natarajan V, Karthikesalingam A, Malgaroli M. The capability of large language models to measure psychiatric functioning. arXiv. Preprint posted online on Aug 3, 2023. [doi: 10.48550/arXiv.2308.01834]
21. Barnhill JW. DSM-5-TR® clinical cases. Psychiatry online. URL: https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9781615375295 [Accessed 2024-07-18]

22. McDuff D, Schaekermann M, Tu T, et al. Towards accurate differential diagnosis with large language models. arXiv. Preprint posted online on Nov 30, 2023. [doi: 10.48550/arXiv.2312.00164]
23. Elyoseph Z, Levkovich I. Beyond human expertise: the promise and limitations of ChatGPT in suicide risk assessment. Front Psychiatry. Aug 2023;14:1213141. [doi: 10.3389/fpsyt.2023.1213141] [Medline: 37593450]
24. Levi-Belz Y, Gamliel E. The effect of perceived burdensomeness and thwarted belongingness on therapists' assessment of patients' suicide risk. Psychother Res. Jul 2016;26(4):436-445. [doi: 10.1080/10503307.2015.1013161] [Medline: 25751580]
25. Joiner T. Why People Die by Suicide. Harvard University Press; 2007.
26. Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner TE. The interpersonal theory of suicide. Psychol Rev. Apr 2010;117(2):575-600. [doi: 10.1037/a0018697] [Medline: 20438238]
27. Darcy A, Beaudette A, Chiauzzi E, et al. Anatomy of a Woebot® (WB001): agent guided CBT for women with postpartum depression. Expert Rev Med Devices. Apr 2022;19(4):287-301. Retracted in: Expert Rev Med Devices. 2023;20(11):989. [doi: 10.1080/17434440.2023.2267389] [Medline: 37801290]
28. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth. Nov 23, 2018;6(11):e12106. [doi: 10.2196/12106] [Medline: 30470676]
29. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health. Dec 13, 2018;5(4):e64. [doi: 10.2196/mental.9782] [Medline: 30545815]
30. Murphy M, Templin J. Our story. Replika. 2021. URL: https://replika.ai/about/story [Accessed 2023-08-20]
31. Kim H, Yang H, Shin D, Lee JH. Design principles and architecture of a second language learning chatbot. Lang Learn Technol. 2022;26:1-18. URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/b3aa08a8-579d-4bf6-b94a-05c2ff67351a/content [Accessed 2024-07-18]
32. Wilbourne P, Dexter G, Shoup D. Research driven: Sibly and the transformation of mental health and wellness. Presented at: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2018:389-391; New York, NY. [doi: 10.1145/3240925.3240932]
33. Denecke K, Abd-Alrazaq A, Househ M. Artificial intelligence for chatbots in mental health: opportunities and challenges. In: Househ M, Borycki E, Kushniruk A, editors. Multiple Perspectives on Artificial Intelligence in Healthcare: Opportunities and Challenges. Springer International Publishing; 2021:115-128. [doi: 10.1007/978-3-030-67303-1]
34. Omarov B, Zhumanov Z, Gumar A, Kuntunova L. Artificial intelligence enabled mobile chatbot psychologist using AIML and cognitive behavioral therapy. IJACSA. 2023;14(6). [doi: 10.14569/IJACSA.2023.0140616]
35. Pham KT, Nabizadeh A, Selek S. Artificial intelligence and chatbots in psychiatry. Psychiatr Q. Mar 2022;93(1):249-253. [doi: 10.1007/s11126-022-09973-8] [Medline: 35212940]
36. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res. Jul 2020;22(7):e16021. [doi: 10.2196/16021] [Medline: 32673216]
37. Brocki L, Dyer GC, Gładka A, Chung NC. Deep learning mental health dialogue system. Presented at: 2023 IEEE International Conference on Big Data and Smart Computing (BigComp); Feb 13-16, 2023:395-398. Jeju, Korea.
38. Martinengo L, Lum E, Car J. Evaluation of chatbot-delivered interventions for self-management of depression: content analysis. J Affect Disord. Dec 2022;319:598-607. [doi: 10.1016/j.jad.2022.09.028] [Medline: 36150405]
39. You Y, Tsai CH, Li Y, Ma F, Heron C, Gui X. Beyond self-diagnosis: how a chatbot-based symptom checker should respond. ACM Trans Comput-Hum Interact. Aug 31, 2023;30(4):1-44. [doi: 10.1145/3589959]
40. Ma Z, Mei Y, Su Z. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. AMIA Annu Symp Proc. Jan 11, 2024;2023:1105-1114. [Medline: 38222348]
41. Lee J, Lee JG, Lee D. Influence of rapport and social presence with an AI psychotherapy chatbot on users' self-disclosure. SSRN. Preprint posted online on Mar 22, 2022. [doi: 10.2139/ssrn.4063508]
42. Das A, Selek S, Warner AR, et al. Conversational bots for psychotherapy: a study of generative transformer models using domain-specific dialogues. In: Demner-Fushman D, Cohen KB, Ananiadou S, Tsujii J, editors. Proceedings of the 21st Workshop on Biomedical Language Processing. Association for Computational Linguistics; 2022:285-297. [doi: 10.18653/v1/2022.bionlp-1.27]
43. Demner-Fushman D, Ananiadou S, Cohen KB, editors. The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks. Association for Computational Linguistics; 2023.
44. Heston TF. Evaluating risk progression in mental health chatbots using escalating prompts. medRxiv. Preprint posted online on Sep 12, 2023. [doi: 10.1101/2023.09.10.23295321]

45. Weidinger L, Mellor J, Rauh M, et al. Ethical and social risks of harm from language models. arXiv. Preprint posted online on Dec 8, 2021. [doi: 10.48550/arXiv.2112.04359]
46. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. eBioMedicine. Apr 2023;90:104512. [doi: 10.1016/j.ebiom.2023.104512] [Medline: 36924620]
47. Koutsouleris N, Hauser TU, Skvortsova V, De Choudhury M. From promise to practice: towards the realisation of AI-informed mental health care. Lancet Digital Health. Nov 2022;4(11):e829-e840. [doi: 10.1016/S2589-7500(22)00153-4]
48. Sickel AE, Seacat JD, Nabors NA. Mental health stigma update: a review of consequences. Adv Ment Health. Dec 2014;12(3):202-215. [doi: 10.1080/18374905.2014.11081898]
49. Alegría M, Green JG, McLaughlin KA, Loder S. Disparities in child and adolescent mental health and mental health services in the U.S. William T. Grant Foundation. Mar 2015. URL: https://wtgrantfoundation.org/wp-content/uploads/2015/09/Disparities-in-Child-and-Adolescent-Mental-Health.pdf [Accessed 2024-07-18]
50. Primm AB, Vasquez MJT, Mays RA. The role of public health in addressing racial and ethnic disparities in mental health and mental illness. Prev Chronic Dis. Jan 2010;7(1):A20. [Medline: 20040235]
51. Schwartz RC, Blankenship DM. Racial disparities in psychotic disorder diagnosis: a review of empirical literature. World J Psychiatry. Dec 2014;4(4):133. [doi: 10.5498/wjp.v4.i4.133] [Medline: 25540728]
52. McGuire TG, Miranda J. New evidence regarding racial and ethnic disparities in mental health: policy implications. Health Aff (Millwood). Mar 2008;27(2):393-403. [doi: 10.1377/hlthaff.27.2.393] [Medline: 18332495]
53. Snowden LR, Cheung FK. Use of inpatient mental health services by members of ethnic minority groups. Am Psychol. Mar 1990;45(3):347-355. [doi: 10.1037//0003-066x.45.3.347] [Medline: 2310083]
54. Henrich J, Heine SJ, Norenzayan A. Beyond WEIRD: towards a broad-based behavioral science. Behav Brain Sci. Jun 2010;33(2-3):111-135. [doi: 10.1017/S0140525X10000725]
55. Lin I, Njoo L, Field A, et al. Gendered mental health stigma in masked language models. arXiv. Preprint posted online on Oct 27, 2022. [doi: 10.48550/arXiv.2210.15144]
56. Liu Y, et al. Trustworthy LLMs: a survey and guideline for evaluating large language models' alignment. arXiv. Preprint posted online on Aug 10, 2023. [doi: 10.48550/arXiv.2308.05374]
57. Straw I, Callison-Burch C. Artificial intelligence in mental health and the biases of language based models. PLoS One. Dec 2020;15(12):e0240376. [doi: 10.1371/journal.pone.0240376] [Medline: 33332380]
58. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature. Aug 2023;620(7972):172-180. [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
59. Keeling G. Algorithmic bias, generalist models, and clinical medicine. arXiv. Preprint posted online on May 6, 2023. [doi: 10.48550/arXiv.2305.04008]
60. Koocher GP, Keith-Spiegel P. Ethics in Psychology and the Mental Health Professions: Standards and Cases. Oxford University Press; 2008.
61. Varkey B. Principles of clinical ethics and their application to practice. Med Princ Pract. Feb 2021;30(1):17-28. [doi: 10.1159/000509119] [Medline: 32498071]
62. Rajagopal A, Nirmala V, Andrew J, Arun M. Novel AI to avert the mental health crisis in COVID-19: novel application of GPT2 in cognitive behaviour therapy. Research Square. Preprint posted online on Apr 1, 2021. [doi: 10.21203/rs.3.rs-382748/v1]
63. Gratch J, Lucas G. Rapport between humans and socially interactive agents. In: Lugrin B, Pelachaud C, Traum D, editors. The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition. Association for Computing Machinery; 2021:433-462. [doi: 10.1145/3477322.3477335]
64. McDuff D, Czerwinski M. Designing emotionally sentient agents. Commun ACM. Nov 20, 2018;61(12):74-83. [doi: 10.1145/3186591]
65. Lundin RM, Berk M, Østergaard SD. ChatGPT on ECT: can large language models support psychoeducation? J ECT. Sep 1, 2023;39(3):130-133. [doi: 10.1097/YCT.0000000000000941] [Medline: 37310145]
66. Wang B, Min S, Deng X, et al. Towards understanding chain-of-thought prompting: an empirical study of what matters. arXiv. Preprint posted online on Dec 20, 2022. [doi: 10.48550/arXiv.2212.10001]
67. Glenberg AM, Havas D, Becker R, Rinck M. Grounding language in bodily states: the case for emotion. In: Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking. Cambridge University Press; 2005:115-128. [doi: 10.1017/CBO9780511499968]
68. Zhong Y, Chen YJ, Zhou Y, Lyu YAH, Yin JJ, Gao YJ. The artificial intelligence large language models and neuropsychiatry practice and research ethic. Asian J Psychiatr. Jun 2023;84:103577. [doi: 10.1016/j.ajp.2023.103577] [Medline: 37019020]

69. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med. Oct 2023;29(10):2396-2398. [doi: 10.1038/s41591-023-02412-6] [Medline: 37391665]
70. Singh OP. Artificial intelligence in the era of ChatGPT - opportunities and challenges in mental health care. Indian J Psychiatry. Mar 2023;65(3):297-298. [doi: 10.4103/indianjpsychiatry.indianjpsychiatry_112_23] [Medline: 37204980]
71. Yang K, Ji S, Zhang T, Xie Q, Kuang Z, Ananiadou S. Towards interpretable mental health analysis with large language models. arXiv. Preprint posted online on Oct 11, 2023. [doi: 10.48550/arXiv.2304.03347]
72. Balasubramaniam N, Kauppinen M, Rannisto A, Hiekkanen K, Kujala S. Transparency and explainability of AI systems: from ethical guidelines to requirements. Inf Softw Technol. Jul 2023;159:107197. [doi: 10.1016/j.infsof.2023.107197]
73. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. In: Koyejo S, Mohamed S, Agarwal A, editors. Advances in Neural Information Processing Systems. Vol 35. Curran Associates, Inc; 2022:24824-24837.
74. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med. Jul 6, 2023;6(1):120. [doi: 10.1038/s41746-023-00873-0] [Medline: 37414860]
75. Ong JCL, Chang SYH, William W, et al. Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health. Jun 2024;6(6):e428-e432. [doi: 10.1016/S2589-7500(24)00061-X] [Medline: 38658283]
76. Minssen T, Vayena E, Cohen IG. The challenges for regulating medical use of ChatGPT and other large language models. JAMA. Jul 25, 2023;330(4):315-316. [doi: 10.1001/jama.2023.9651] [Medline: 37410482]
77. van Heerden AC, Pozuelo JR, Kohrt BA. Global mental health services and the impact of artificial intelligence-powered large language models. JAMA Psychiatry. Jul 1, 2023;80(7):662-664. [doi: 10.1001/jamapsychiatry.2023.1253] [Medline: 37195694]
78. Cabrera J, Loyola MS, Magaña I, Rojas R. Ethical dilemmas, mental health, artificial intelligence, and LLM-based chatbots. In: Bioinformatics and Biomedical Engineering. Springer Nature Switzerland; 2023:313-326. [doi: 10.1007/978-3-031-34960-7_2]

Abbreviations

ACOG: American College of Obstetricians and Gynecologists
BERT: Bidirectional Encoder Representations From Transformers
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
EBP: evidence-based practice
HIPAA: Health Insurance Portability and Accountability Act
LLM: large language model
Psy-LLM: psychological support with large language model

Edited by John Torous; peer-reviewed by Brent M M Kious, Jasmine Chiat Ling Ong, Thomas F Heston; submitted 13.04.2024; final revised version received 31.05.2024; accepted 01.06.2024; published 29.07.2024

Please cite as:
Lawrence HR, Schneider RA, Rubin SB, Matarić MJ, McDuff DJ, Jones Bell M
The Opportunities and Risks of Large Language Models in Mental Health
JMIR Ment Health 2024;11:e59479
URL: https://mental.jmir.org/2024/1/e59479
doi: 10.2196/59479