Editorial

Digital Mental Health Research Priorities, Revisited for the AI and Large Language Model Era

Max Valentin Birk[1*], PhD; Shruti Kochhar[2*], MS; Keris Myrick[3*], MBA, MS; Stephen M Schueller[4*], PhD; John Torous[3*], MD, MBI

[1]Eindhoven University of Technology, Eindhoven, North Brabant, The Netherlands
[2]JMIR Publications, Toronto, ON, Canada
[3]Beth Israel Deaconess Medical Center, Boston, MA, United States
[4]University of California, Irvine, Irvine, CA, United States
*all authors contributed equally

Corresponding Author:

John Torous, MD, MBI
Beth Israel Deaconess Medical Center
20 Overland St, Suite OV-202
Boston, MA 02215
United States
Phone: 1 6176676700
Email: jtorous@bidmc.harvard.edu

Abstract

Digital mental health has become an established part of mental health care, but the rapid arrival of large language models and other artificial intelligence (AI) tools has refocused attention on the evidence needed to guide the field. This editorial updates the research priorities articulated by *JMIR Mental Health* in 2023, while reaffirming their emphasis on equity, replicability, privacy, efficacy, and engagement. While the importance of these priorities has not changed in recent years, the urgency with which they must now be applied has. As digital tools become more clinically consequential, research must move beyond demonstrating that a technology is feasible, usable, or novel. The field now needs studies that clarify how these tools work, for whom they are beneficial, under what conditions they may cause harm, and how they can be ethically integrated into care. We call for research that is transparent about the technologies being studied, grounded in meaningful clinical questions, attentive to safety, and designed to produce knowledge that remains useful as specific products and models change.

JMIR Ment Health 2026;13:e104118; doi: 10.2196/104118

Keywords: artificial intelligence; AI; large language model; LLM; mental health; apps

Introduction

Digital mental health is no longer an emerging field. Increases in teletherapy platforms and providers have increased many people's comfort and knowledge with receiving services remotely. It might already be the case that more people are using chatbots than therapists for mental health support [1]. What started as a field based on computerized programs has expanded to include apps, virtual reality, social media, digital phenotyping, rule-based chatbots, and large language models (LLMs). Many health systems and public entities are exploring the use of digital mental health to more effectively meet people's needs [2,3]. This shift raises the standard for the research we publish. Although establishing the feasibility of digital mental health interventions previously made a useful contribution, similar feasibility studies may fail to further advance the field. The field now needs more evidence that can guide clinical care, inform policy, protect patients, and identify which technologies work, for whom, under what conditions, and at what risk [4-6].

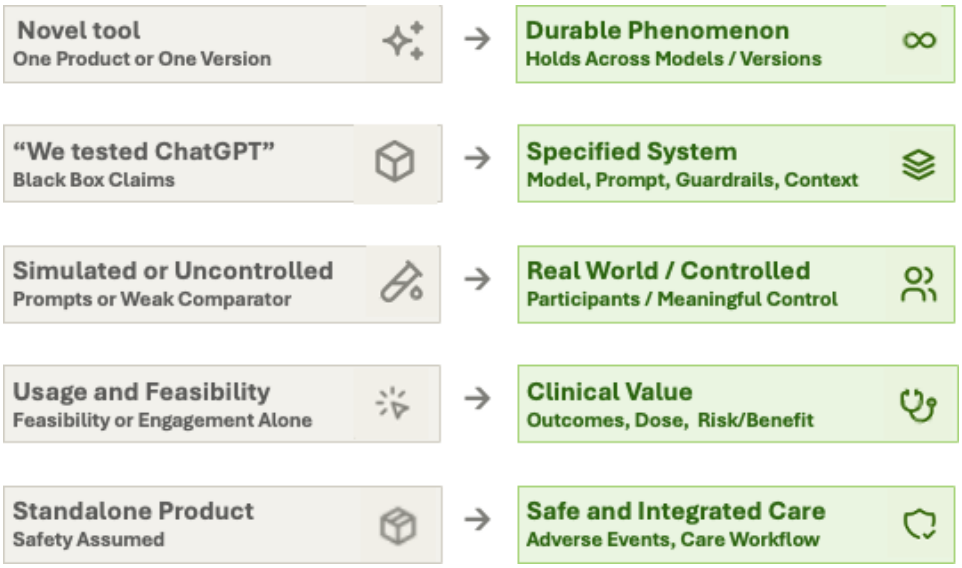
In 2023, *JMIR Mental Health* outlined priorities for digital mental health research centered on equity, replicability, privacy, efficacy, and engagement [7]. Those priorities remain. What has changed is the speed, scale, and consequence of the technologies now entering mental health care, especially with the emergence of LLMs. Although LLMs are the most visible example of the current methodological challenge, the standards described here apply across digital mental health. Given the recent volume of submissions on LLMs, better defining the scope of these papers will aid authors wishing to disseminate their work and provide

guidance to the field on which areas of inquiry we see as most generative.

The challenges of LLMs moving faster than the research base have been well documented across the mental health field, including cases of patient harm, ongoing lawsuits, and expanding efforts to restrict or ban some tools for care. This is the bind David Collingridge [8] described decades ago: by the time evidence accumulates about a new technology, it is often too embedded in practice, regulation, or use to redirect easily. The question facing digital mental health research is no longer whether the field can produce evidence, but whether the evidence it produces is rigorous, generalizable, timely, and useful enough to inform how these technologies are evaluated and used to improve care.

Artificial intelligence (AI) is itself part of how that evidence is now produced. Authors today may use LLMs to draft, summarize, code, and even analyze. The volume of research on AI in mental health has risen sharply, and so has concern that the quality of that research has not kept pace. This is the paradox of the current moment: the same tools that accelerate research are also raising questions about its soundness. This is why research standards matter more than ever, and why *JMIR Mental Health* seeks to advance the most rigorous and useful research with new priorities, as summarized in Figure 1. We are seeking research that helps the field learn something durable, especially given how rapidly the field changes.

Figure 1. Schematic of updated research priorities, with a focus on artificial intelligence. Across all priorities, equity, safety, transparency, implementation, and clinical relevance still apply, as noted in our prior guidance.



What We Want From Research

For digital mental health work in general, and for LLM-related work in particular, the strongest studies are designed around the phenomenon being investigated rather than only the specific tool being tested, which is aligned with the National Institute of Mental Health's experimental therapeutics approach [9]. Submissions in this space are most useful when they meet several standards.

Durable Knowledge Beyond a Single Product

Studies should be explicit about whether they are evaluating a general phenomenon, a product, a model family, a model version, a user interface, or a deployment context. Work that examines more than one model, version, setting, or population is often better positioned to help the field draw conclusions that remain useful as technologies change. The most useful studies will identify those mechanisms or active ingredients that hold across technologies, much as the Wellcome Trust's active ingredients commission sought to isolate the components driving benefit in mental health interventions [10] This extends the National Institute of Mental Health experimental therapeutics framing noted above [9]: even if the mechanisms underlying mental health conditions, and those of the digital tools acting on them, are never fully specified, research oriented toward mechanisms will prove more durable than research tied to a single product.

Transparent Characterization of the Technology

Terminology and technical specificity matter more than ever as technology rapidly evolves. The base model behind a product, the wrapper or interface layer built on top of it, the system prompt, the safety guardrails, and the deployment context all shape what a study is actually measuring and can be the difference between an AI missing a safety warning or not [11]. Submissions that collapse these components into phrases such as "we tested ChatGPT" become difficult to evaluate, replicate, or interpret. Worse, they may misinform about safety, as critical protections often differ across the model, product, and API layers of technology. Reporting frameworks such as CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence), SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence), DECIDE-AI

(Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence), MI-CLAIM (Minimum Information About Clinical Artificial Intelligence Modeling), MINIMAR (Minimum Information for Medical AI Reporting), TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence), and AI model cards published by developers can help authors describe AI systems with better precision.

Meaningful Comparators and Controls

The choice of comparator greatly determines what a paper can offer the field. Meaningful comparators in this space are those that control for the nonspecific effects of technology itself, including structure, attention, credibility, expectancy, and the sense of receiving a plausible intervention. Studies should be designed to isolate the active ingredient they claim to test. Recent work in this journal showed that ecologically valid online controls produced credibility, expectancy, and outcome trajectories nearly indistinguishable from purpose-built digital interventions. These controls included curated libraries of the kind of popular content that users actually encounter when seeking help online and showed a small between-condition effect size (Cohen d<0.15) [12]. Other recent pilot work found that an unstructured general-purpose LLM produced reductions in depression symptoms comparable to a purpose-built mental health chatbot, with both outperforming an assessment-only control [13]. Whether we call this a digital placebo effect or not, these examples demonstrate empirical data on the nonspecific effects of digital engagement.

Clinical Outcomes and Connection to Care

The field has produced many studies showing that AI systems can respond plausibly to clinical prompts, but fewer that show these systems improve patients' lives [14-16]. LLM mental health work has generated impressive results on simulated cases, standardized vignettes, and short interaction tasks. A recent scoping review of ChatGPT's clinical applications in mental health found that 83% of the 60 included studies were prompt-based experiments without human participants, with only 3 controlled trials and only 6 studies enlisting clinical populations [16]. Models that perform well on benchmarks may behave less reliably in real interactions, and even models that demonstrate safe responses in single-turn exchanges may pose harms across long conversations [17]. Submissions should use measures that are practical and actionable in real-world settings and hold value for patients and clinicians, as these are most likely to shape practice. These measures move beyond symptom change to also include functional improvement, quality of life, safety, and integration into existing care pathways.

Human Support, Safety, and Accountability

Human support should be described with the same specificity as the technology itself. Digital navigators [18], peer specialists [19,20], clinicians, and other support roles can shape engagement, safety, and outcomes, but vague claims about human oversight are not enough. At the same time, recent work has cautioned that "human-in-the-loop" oversight, when invoked without specification, can function more as symbolic reassurance than substantive protection [21]. Human-in-the-loop should not be treated as a safety claim unless the loop is described with enough detail to evaluate who is responsible, what they can see, when they act, and what authority they have to change care. Studies that include human roles should specify details such as who performs those roles, what training they have received, what data they monitor, and when they intervene.

Reporting on Adverse Events

Clinical trials in digital mental health still lack a consensus framework for identifying and reporting adverse events, in contrast to pharmacological trials, which operate under strict regulatory oversight. Potential harms, such as worsening symptoms, distressing disengagement, and inappropriate responses to crises, are inconsistently tracked and reported across studies [22,23]. Prior efforts have proposed taxonomies of harm [24], yet these are rarely taken up in practice. AI-based interventions introduce a further complication: unlike static digital tools, the intervention itself may exhibit considerable variance, meaning that inappropriate crisis responses and problematic use require additional vigilance. Rigorous research should clearly define adverse event criteria, specify monitoring protocols, identify escalation thresholds, and report safety findings transparently. AI harms can be clear, but they can also accumulate slowly over long conversations and the formation of a dangerous parasocial relationship that may be harder to detect, yet more deadly [25].

Theory and Data on Emerging Phenomena

New phenomena tend to outpace the literature, and that appears to be the case with AI. "AI psychosis," while not a formal diagnostic category, has become a shorthand for concerns about delusional ideation, overreliance, reality testing, and harmful narrative reinforcement in interactions with conversational agents [26]. Theoretical contributions should do real conceptual work by locating such phenomena in relation to existing concepts, proposing mechanisms, or specifying who may be at risk and why. Empirical contributions should rest on more than illustrative examples. Ideally, new datasets can be created and shared. For phenomena where the published evidence base remains nascent, we particularly welcome detailed clinical case reports and work that includes patients as authors, whose own accounts shape the analysis rather than appearing only as data.

Equity, Voice, and Context

The 2023 editorial called for equity in access, in research populations, and in authorship, and the years since have made all three calls more pressing. Models trained predominantly on English-language sources from high-resource areas perform less reliably in other languages and contexts, and research populations remain skewed away from the needs

of low- and middle-income countries. The mismatch is highlighted in a recent conversation-level analysis [27] of ChatGPT use in India, Nigeria, Brazil, and Pakistan, which found that health-related queries occurred at substantially higher rates than in Western comparators. Equity requires asking whether the language, culture, infrastructure, costs, literacy, and care pathways assumed by a technology align with the intended user group and population. We value submissions conducted in low- and middle-income settings, work in languages other than English, studies that surface cultural variation in how digital tools are used and understood, and work that brings lived-experience perspectives into authorship.

A Note on Scope

Much of this editorial focuses on LLMs because they raise questions for the field that are especially acute. The journal's interests, however, remain broader, and an LLM is often not the right, or only, tool for mental health. We continue to welcome rigorous work on mental health apps, virtual and augmented reality, rule-based chatbots, digital phenotyping, online interventions, remote monitoring, digital therapeutics, social media, clinical prediction, hybrid care models, and other topics. The standards described here apply across these areas: submissions should be replicable, clinically meaningful, attentive to equity and engagement, and grounded in appropriate comparisons.

Engagement remains one of the field's foremost challenges [28,29]. We are thus interested in work that moves beyond reporting low engagement (or perhaps in some cases, too much engagement with AI or social media) toward examining how engagement matters on a personal level. Does greater use predict better outcomes, or does it simply reflect higher baseline need? Is there a dose-response relationship, and how do we even measure dose? Should we measure minutes, sessions, completed modules, conversational turns, therapeutic content, or human contact?

With LLMs, the engagement question is now whether it is too little, as was often the case with smartphone apps. Recent passive-sensing data from over 6000 US youth found that median daily use of generative AI apps was under a minute, while a small subset engaged for more than 40 minutes a day [30]. This pattern suggests that average exposure can conceal very different individual experiences. Cases of compulsive use and parasocial attachment raise empirical questions about what too much engagement looks like, who is vulnerable to it, and what consequences may follow.

Implementation

Implementation deserves more attention. The most clinically useful tools are not always the most technically advanced ones. Research showing how a technology fits a setting, workflow, payment model, workforce, or population is crucial. LLMs raise new challenges in implementation, design (eg, accessibility and access to data), and output (eg, bias and correctness) [31]. Real-world implementation is shaped by infrastructure, cost, training, reliability, privacy, and legal accountability as much as by underlying capability [32]. For example, recent research on the current generation of AI scribes has suggested that they may actually save negligible time, [33] and research examining the costs of running LLMs for medical coding suggests the cost may be far higher than expected [34]. Research that takes those constraints seriously rather than designing around them is especially productive.

Theories, frameworks, and models from implementation science might support structuring this research in ways to support learnings across projects. It is also worth considering the extent to which these theories, frameworks, and models apply to digital mental health and what adaptations might be necessary [35]. Given the established evidence base for many digital mental health interventions, better understanding how to successfully integrate them into routine care settings is an opportunity for future contributions.

Real-World Evidence

As implementation increases, data generated from real-world settings has become more common. Although these investigations overcome the challenges posed by conducting clinical trials in artificial environments that do not mimic real-world deployments, they face limitations that may limit the insights they can offer the field. When used, real-world data need to be appropriate for the proposed research question, including aspects such as the size and representativeness of the sample. Information about engagement, outcome, covariates, and potential confounders is always critical. In many examples of real-world data, it may not be possible to use as rigorous a comparison of conditions as might be expected in clinical trial settings; however, this does not mean that other considerations of rigor and reproducibility are not relevant. Studies that merely examine convenience samples from deployed interventions may contain a degree of bias, and any conclusions may raise more questions than answers.

Expanding Beyond Current Knowledge

A few categories of submission are unlikely to fit, less because they lack effort than because they no longer move the field forward without a stronger justification. Studies whose main finding is that people can use a piece of technology or find it acceptable are rarely sufficient in 2026. Cross-sectional surveys of technology use, particularly surveys measuring generic interest in AI for mental health, have been well covered. Research focused on older AI models that are no longer in use is generally of historical rather than practical interest.

Effect sizes for digital mental health interventions have also been well characterized. A meta-analysis of 176 randomized trials placed app-based effects for depression and anxiety in the small-to-modest range (Hedges g was

approximately 0.26 to 0.28 for both depression and anxiety) [36], and a recent meta-analysis of 39 AI chatbot trials for depression and anxiety reported nearly identical estimates (Hedges *g* ranged from 0.28 to 0.31) [37]. While these effect sizes will vary based on human support, follow-up duration, clinical severity, and other factors, they at least offer a baseline to orient new findings. Papers reporting substantially larger effects are welcome, but the rigor of evidence behind them should match the size of the claim. Finally, work raising ethical concerns, including but not limited to the use of patient data without appropriate permissions, will not be considered regardless of the strength of the underlying findings.

Transparency in AI-Assisted Research

The profound impact of generative AI tools in scholarly writing is now well known—the same tools that are shaping how digital mental health research is conducted are also driving how related scholarly work is produced [38]. These tools introduce risks spanning factual accuracy, fabricated content, hallucinated references, copyright infringement, and ethical breaches. As a result, manuscripts that do not address AI assistance at all are increasingly difficult to interpret both scientifically and ethically, challenging editors, reviewers, and readers. Authors should therefore report how and when generative AI was used in the research or writing process, specifying the tools, the nature of the AI-generated content, and, where relevant, the prompts that shaped it [39], per guidelines developed by JMIR leadership. We also appreciate and expect that many new research papers will use AI in a larger role and welcome that when fully disclosed.

Conclusion

The potential of digital mental health has always been that technology can expand access to care while improving outcomes that matter to patients, clinicians, and communities. Today that promise is more visible, more commercially active, and more clinically consequential than ever. But promise is not evidence, and excitement is not a substitute for rigorous research, as earlier instances of enthusiasm around apps, virtual reality, and other digital tools have shown. The window Collingridge [8] described has not closed, but it is narrower than it was when *JMIR Mental Health* published its 2023 editorial guidance.

The Society of Digital Psychiatry, in partnership with the journal, has set out a roadmap for translating research into practice through education, AI standards, and digital navigators [40]. The publication standards described here are the upstream complement: research must be designed to be worth translating. Whether the next 3 years bear out the promise of digital mental health will depend less on the tools than on the evidence that shapes their use.

Acknowledgments

ChatGPT (GPT-5.5 Thinking; OpenAI) was used to copyedit the manuscript after drafting, on June 8 and 20, 2026, and to format the figure on June 19, 2026. All AI-assisted output was reviewed and approved by the authors, who take full responsibility for the content.

Funding

The authors declare no financial support was received for this work.

Data Availability

Not applicable.

Conflicts of Interest

JT holds editor-in-chief role, SK holds managing editor role, and MB, SS, and KM hold editorial board member roles for *JMIR Mental Health* at the time of this publication.

References

1. Rousmaniere T, Goldberg SB, Torous J. Large language models as mental health providers. Lancet Psychiatry. Jan 2026;13(1):7-9. [doi: 10.1016/S2215-0366(25)00269-X] [Medline: 40939602]
2. Zhao X, Stadnick NA, Ceballos-Corro E, et al. Facilitators of and barriers to integrating digital mental health into county mental health services: qualitative interview analyses. JMIR Form Res. May 16, 2023;7(1):e45718. [doi: 10.2196/45718] [Medline: 37191975]
3. Hitcham L, Gómez Bergin A, Ito-Jaeger S, Reeves S, Perez Vallejos E. Adoption of digital mental health interventions in National Health Service England, Scotland, and Wales: freedom of information questionnaire study. JMIR Ment Health. May 14, 2026;13:e92187. [doi: 10.2196/92187] [Medline: 42133938]
4. Benjet C, Zainal NH, Albor Y, et al. A precision treatment model for internet-delivered cognitive behavioral therapy for anxiety and depression among university students: a secondary analysis of a randomized clinical trial. JAMA Psychiatry. Aug 1, 2023;80(8):768-777. [doi: 10.1001/jamapsychiatry.2023.1675] [Medline: 37285133]

5. Pigeon WR, Bishop TM, Bossarte RM, Schueller SM, Kessler RC. A two-phase, prescriptive comparative effectiveness study to optimize the treatment of co-occurring insomnia and depression with digital interventions. Contemp Clin Trials. Sep 2023;132:107306. [doi: 10.1016/j.cct.2023.107306] [Medline: 37516163]
6. Wanniarachchi VU, Greenhalgh C, Choi A, Warren JR. Personalization variables in digital mental health interventions for depression and anxiety in adolescents and youth: a scoping review. Front Digit Health. 2025;7:1500220. [doi: 10.3389/fdgth.2025.1500220] [Medline: 40444184]
7. Torous J, Benson NM, Myrick K, Eysenbach G. Focusing on digital research priorities for advancing the access and quality of mental health. JMIR Ment Health. Apr 24, 2023;10:e47898. [doi: 10.2196/47898] [Medline: 37093624]
8. Collingridge D. The Social Control of Technology. St Martin's Press; 1980.
9. Graham AK, Lattie EG, Mohr DC. Experimental therapeutics for digital mental health. JAMA Psychiatry. Dec 1, 2019;76(12):1223-1224. [doi: 10.1001/jamapsychiatry.2019.2075] [Medline: 31433448]
10. Wolpert M, Pote I, Sebastian CL. Identifying and integrating active ingredients for mental health. Lancet Psychiatry. Sep 2021;8(9):741-743. [doi: 10.1016/S2215-0366(21)00283-2] [Medline: 34419174]
11. Ramaswamy A, Tyagi A, Hugo H, et al. ChatGPT Health performance in a structured test of triage recommendations. Nat Med. May 2026;32(5):1671-1675. [doi: 10.1038/s41591-026-04297-7] [Medline: 41731097]
12. Kaveladze BT, Schueller SM, Mohr DC. Popular online content as a treatment-as-usual control in digital mental health intervention trials: secondary analysis of two online randomized controlled trials with repeated measures. JMIR Ment Health. Apr 20, 2026;13:e83707. [doi: 10.2196/83707] [Medline: 42008585]
13. Kuta B, Novak L, Zidkova R, et al. Effectiveness of a fully automated mobile therapeutic versus a general chatbot in reducing depression and anxiety and improving well-being: feasibility randomized controlled trial. JMIR Ment Health. Apr 22, 2026;13:e82642. [doi: 10.2196/82642] [Medline: 42018980]
14. Hua Y, Siddals S, Ma Z, et al. Charting the evolution of artificial intelligence mental health chatbots from rule-based systems to large language models: a systematic review. World Psychiatry. Oct 2025;24(3):383-394. [doi: 10.1002/wps.21352] [Medline: 40948070]
15. Hua Y, Xia W, Bates D, et al. Standardizing and scaffolding health care AI-chatbot evaluation: systematic review. JMIR AI. Nov 7, 2025;4(1):e69006. [doi: 10.2196/69006] [Medline: 41202290]
16. Balan R, Gumpel TP. ChatGPT clinical use in mental health care: scoping review of empirical evidence. JMIR Ment Health. Dec 24, 2025;12:e81204. [doi: 10.2196/81204] [Medline: 41442647]
17. Bean AM, Payne RE, Parsons G, et al. Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. Nat Med. Feb 2026;32(2):609-615. [doi: 10.1038/s41591-025-04074-y] [Medline: 41663592]
18. Shin HD, Kemp J, Kassam I, et al. A scoping review of the characteristics, responsibilities, implementations and evaluations of digital navigators in healthcare. NPJ Digit Med. Apr 18, 2026. [doi: 10.1038/s41746-026-02647-w] [Medline: 42000918]
19. Myrick KJ. From access to impact: Why digital literacy and lived experience must guide digital health tools. Schizophr Res. Jul 2026;293:v-vi. [doi: 10.1016/j.schres.2026.03.022]
20. Davis K. Peer support principles and navigation for LLMs in mental health. Curr Treat Options Psych. Mar 31, 2026;13(1):8. [doi: 10.1007/s40501-026-00378-z]
21. Abulibdeh R, Agyemang GO, Celi LA, et al. Who's really in the loop? Rethinking oversight in AI-assisted health care. The Lancet. Jun 2026;407(10545):2340-2344. [doi: 10.1016/S0140-6736(26)00204-7]
22. Gómez Bergin AD, Valentine AZ, Rennick-Egglestone S, Slade M, Hollis C, Hall CL. Identifying and categorizing adverse events in trials of digital mental health interventions: narrative scoping review of trials in the International Standard Randomized Controlled Trial Number Registry. JMIR Ment Health. Feb 22, 2023;10:e42501. [doi: 10.2196/42501] [Medline: 36811940]
23. Linardon J, Fuller-Tyszkiewicz M, Firth J, et al. Systematic review and meta-analysis of adverse events in clinical trials of mental health apps. NPJ Digit Med. Dec 18, 2024;7(1):363. [doi: 10.1038/s41746-024-01388-y] [Medline: 39695173]
24. Rozental A, Andersson G, Boettcher J, et al. Consensus statement on defining and measuring negative effects of Internet interventions. Internet Interv. Mar 2014;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]
25. Moore J, Mehta A, Agnew W, et al. Characterizing delusional spirals through human-LLM chat logs. arXiv. Preprint posted online on Mar 17, 2026. [doi: 10.48550/arXiv.2603.16567]
26. Flathers M, Roux S, Torous J. Beyond artificial intelligence psychosis: a functional typology of large language model-associated psychotic phenomena. Lancet Digit Health. Apr 2026;8(4):100974. [doi: 10.1016/j.landig.2025.100974] [Medline: 41833467]
27. Roy Chowdhury S, Garimella K. How people use ChatGPT: conversation-level evidence from India, Nigeria,Brazil and Pakistan. GitHub. 2025. URL: https://gvrkiran.github.io/content/How_people_use_ChatGPT.pdf [Accessed 2026-07-03]

28. Linardon J, Torous J, Messer M, et al. Association between user engagement and clinical outcomes in smartphone apps for depression and anxiety: A systematic review and meta-analysis. Psychiatry Res. Jan 2026;355:116864. [doi: 10.1016/j.psychres.2025.116864] [Medline: 41319621]
29. Smith KA, Ward T, Lambe S, et al. Engagement and attrition in digital mental health: current challenges and potential solutions. NPJ Digit Med. Jul 2, 2025;8(1):398. [doi: 10.1038/s41746-025-01778-w] [Medline: 40604240]
30. Maheux AJ, Akre-Bhide S, Boeldt D, et al. Generative artificial intelligence applications use among US youth. JAMA Netw Open. Feb 2, 2026;9(2):e2556631. [doi: 10.1001/jamanetworkopen.2025.56631] [Medline: 41627817]
31. Nilsen P, Svedberg P, Nygren J, Frideros M, Johansson J, Schueller S. Accelerating the impact of artificial intelligence in mental healthcare through implementation science. Implement Res Pract. 2022;3:26334895221112033. [doi: 10.1177/26334895221112033] [Medline: 37091110]
32. Schueller SM, Torous J. Scaling evidence-based treatments through digital mental health. Am Psychol. Nov 2020;75(8):1093-1104. [doi: 10.1037/amp0000654] [Medline: 33252947]
33. Rotenstein LS, Holmgren AJ, Thombley R, et al. Changes in clinician time expenditure and visit quantity with adoption of artificial intelligence-powered scribes: a multisite study. JAMA. Apr 28, 2026;335(16):1408-1417. [doi: 10.1001/jama.2026.2253] [Medline: 41920565]
34. Burns ML, Chen SY, Tsai CA, et al. Generative AI costs in large healthcare systems, an example in revenue cycle. NPJ Digit Med. Sep 30, 2025;8(1):579. [doi: 10.1038/s41746-025-01971-x] [Medline: 41028226]
35. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res. Jan 25, 2019;21(1):e11752. [doi: 10.2196/11752] [Medline: 30681966]
36. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. World Psychiatry. Feb 2024;23(1):139-149. [doi: 10.1002/wps.21183] [Medline: 38214614]
37. Sohn JS, Ha BG, Park S, et al. Systematic review and meta analysis of chatbots in the management of depressive and anxiety symptoms. NPJ Digit Med. Mar 25, 2026;9(1):377. [doi: 10.1038/s41746-026-02566-w] [Medline: 41882250]
38. Oermann MH, Owens JK, Carter-Templeton H, Peterson G, Bailey HE. Using artificial intelligence for scholarly writing. Am J Nurs. Nov 1, 2025;125(11):52-55. [doi: 10.1097/AJN.0000000000000179] [Medline: 41128585]
39. Leung TI, de Azevedo Cardoso T, Mavragani A, Eysenbach G. Best practices for using AI tools as an author, peer reviewer, or editor. J Med Internet Res. Aug 31, 2023;25:e51584. [doi: 10.2196/51584] [Medline: 37651164]
40. Torous J, Ledley KT, Gorban C, et al. Accelerating digital mental health: the society of digital psychiatry's three-pronged road map for education, digital navigators, and AI. JMIR Ment Health. Nov 27, 2025;12(1):e84501. [doi: 10.2196/84501] [Medline: 41313160]

Abbreviations

AI: artificial intelligence
LLM: large language model

Edited by Stefano Brini; This is a non–peer-reviewed article; submitted 09.Jun.2026; final revised version received 21.Jun.2026; accepted 24.Jun.2026; published 10.Jul.2026

Please cite as:
Birk MV, Kochhar S, Myrick K, Schueller SM, Torous J
Digital Mental Health Research Priorities, Revisited for the AI and Large Language Model Era
JMIR Ment Health 2026;13:e104118
URL: https://mental.jmir.org/2026/1/e104118
doi: 10.2196/104118