Opportunities and Considerations for Augmented Intelligence in Measurement-Informed Care in Mental Health

SEPTEMBER 2024

MEADOWS
MENTAL HEALTH
POLICY INSTITUTE

MEADOWS
MENTAL HEALTH
POLICY INSTITUTE

Contents

Executive Summary

Measurement-informed care (MIC) informs clinical care by leveraging repeated, systematic use of validated measures to inform treatment decisions and monitor progress over time. Augmented intelligence (AI)—a conceptualization of artificial intelligence that focuses on how it enhances human intelligence rather than replacing it—is opening new opportunities to expand and potentially improve MIC for individuals with mental health conditions.[1] This includes opportunities to change analysis of traditional MIC data and incorporate new types of data, including digital phenotyping and natural language processing. AI applications for MIC for mental health have an emerging evidence base, and they will likely change current processes. How data are measured, assessed, and used in mental health contexts is constantly evolving – AI is the next step in the evolution of measurement and monitoring of mental health care. New ethics and governance frameworks are needed to facilitate the development, deployment, and implementation of AI for MIC mental health in a way that is evidence-based and preserves privacy, safety, equity, and fairness.

The Importance of Measurement-Informed Care for Mental Health

MIC is an important approach for mental health care that integrates the regular use of patient-reported outcome measures (PROMs) and other clinical outcome measures to inform treatment decisions and monitor progress over time.[2] Starting with the initial screening of patients with psychiatric disorders, MIC involves repeated, systematic use of validated measures during clinical encounters to inform decision-making about treatment, thereby supporting – not replacing – clinical judgement. Importantly, in MIC for mental health care, measurement informs but does not dictate care. Providers and patients use measurement, with multiple other factors, to develop an individualized treatment plan and continuously adjust treatment. In other words, MIC emphasizes that measurement is an important, though not exclusive, factor for mental health treatment decisions. Additionally, MIC emphasizes that measurement has limitations due to the need to incorporate qualitative data and account for various factors, such as the impact of race, gender, socioeconomic status, and other social determinants of health.[3]

MIC facilitates earlier detection, more timely treatment (less lag time), and improved outcomes for individuals with mental illnesses.[4] The evidence for MIC is well-established, showing potential outcome improvements in the range of 20% to 60%.[5] A meta-analysis of 51 randomized controlled trials found that trials that consistently used MIC for medication management and psychotherapy showed significantly improved patient outcomes.[6,7] Other studies have found MIC is associated with lower costs of care.[8] MIC has also been found to allow providers to better identify patients at risk for nonresponse or deterioration in functioning.[9]

Challenges in Measurement-Informed Care

Despite the benefits of MIC in mental health care, there are also limitations.[10,11] Fewer than 20% of mental health providers use measurement-informed care.[12] A number of barriers lead to a lack of uptake, such as the time and effort required by both patients and providers to complete and review measures, patients' symptoms (e.g., acute psychiatric symptoms, such as psychosis, preventing completion of patient-reported measures), and clinicians' attitude towards MIC (e.g., belief that standardized measures are not as accurate as clinical judgement).[13]

The subjective nature of measurement tools may lead to measurement errors.[14] In the pediatric setting, "gold standard screening surveys to evaluate mental health problems in children typically rely on reports given by caregivers, who tend to unintentionally under-report, and in some cases over-report, child symptomology."[15] A recent study found that "nearly 30% of adolescents who engaged in self-harm or suicide reported 'not at all' on the PHQ-9 item about suicidal thoughts."[16] Additionally, validated symptom scales, such as the Personal Health Questionnaire-9 (PHQ-9) and General Anxiety Disorder-7 (GAD-7), may downplay daily functioning, which is often of primary importance to patients but is not always the primary target for psychiatric care.[17] All of this points to a need to incorporate new types of data in measurement into traditional forms of measurement in mental health care.

How AI Can Support Measurement in Mental Health Care

Recent advancements in AI hold potential to bring new approaches to measurement in mental health care. There are new possibilities for using diverse data sources to support the screening, measurement, and monitoring in mental health care using data from both inside and outside clinical settings. These new sources of data may augment current MIC approaches.

AI provides opportunities to collect and analyze measurements in new ways, such as through using digital phenotyping and leveraging new data sources both inside and outside traditional clinical settings. Below, we discuss ways AI is being used to augment traditional measurement in mental health care, including:

1. Analysis of traditional MIC data
2. Incorporation of new types of data, such as digital phenotyping and natural language processing

Using AI to Collect & Analyze Traditional Mental Health Data in New Ways

As noted above, despite its benefits, fewer than 20% of mental health care providers report using MIC.[18,19] Technology can make adopting MIC easier by automating part or all of the process of administering and scoring measures. AI can be applied to draw comparisons to past data in real-time, and, using analytics to aggregate patient information, enables understanding of clinic and population-level data to inform practice.[20] Some health systems have invested in deep learning technology, a type of artificial intelligence, to predict which patients are going to respond to a treatment and adjust treatment plans accordingly.[21]

AI can also provide new ways to collect measurements and assessments in ways that may potentially be more engaging for individuals. AI chatbots can provide a new mechanism for people to fill out intake questionnaires and screening tools. Additionally, AI chatbots can be used to automate reminders and identify trends in form responses. For example, Headspace combines AI to automate gold-standard clinical assessments, appointment reminders, and aspects of coaching to free up time for clinical providers to meet with clients.[22] Specifically, Headspace uses AI to analyze trends in clinical assessments over time, personalize suggestions for self-care materials in between suggestions, and suggest prompts that coaches can utilize or modify during text-based coaching sessions, which provides a hybrid human-chatbot approach to coaching. However, despite the increased interest in chatbots for mental health, existing options have limitations, including struggling to respond to user input, low engagement and high dropout rates from participants.[23,24,25,26]

Using AI to Augment Traditional MIC Data with New Kinds of Data

In addition to enhancing the ways we collect and analyze current forms of mental health measurement data, AI can enable measurements from new sources of data inside and outside clinical settings, such as social media data, information consumed on the internet, voice and facial recognition, electronic health record (EHR) data and clinical notes, and data from wearables, which come from our interactions with smartphones (e.g., analyzing patient's verbal patterns, tone, and word choice).[27,28]

In particular, the increasing use of telemedicine for mental health visits has produced new sources of data, including videos from patient visits, audio recordings, eye movements, etc. There are also massive amounts of untapped data outside medical systems, such as lifestyle, nutrition, and environmental data. Wearable devices are increasingly able to measure these untapped data sources that have a considerable influence on better understanding mental health.[29] Applying AI to these previously non- or under-utilized mental health data sources enables greater insights to be available for MIC.

Digital Phenotyping

Digital phenotyping, an emerging area of research, "is an innovative approach concerning the extraction and analysis of information about behavior, cognition, and mood through digital tools" collected on devices people now commonly possess, such as smartphones and wearables.[30] AI enables the processing of the data from these digital devices. Digital phenotyping aims to monitor patients with mental health conditions to detect changes or predict relapse. The validation of digital phenotyping methods is a growing area of research.[31] Digital phenotyping may advance MIC by providing new or additional measurements to inform decision-making about treatment and support clinical judgement.

Digital phenotyping can be based on active or passive data collection. Passive data collection may entail information continuously collected from smartphone usage (e.g., GPS and movement patterns, keystrokes, etc.), driving behavior, biometric data from wearable devices, or social media activity.[32,33] Active data collection may involve specific assessments at home or in a clinical setting, such as voice analysis, facial expression analysis, or eye-tracking testing.[34]

Digital phenotyping methods using AI are being developed and validated for mental health conditions. A recent systematic review found physiological and behavioral data collected through digital phenotyping methods that use mobility, location, phone use, call log, heart rate, sleep, head movements, facial and vocal characteristics, social rhythms, conversations, number of steps, screen use, text message log, etc., to detect and predict changes in symptoms of patients with schizophrenia, mood disorders, and anxiety disorders to allow for contact and intervention before an adverse event can occur.[35] Digital phenotyping data are analyzed to predict early signs of symptom relapse using skin sensors or symptom exacerbations by tracking movement.[36]

Digital phenotyping can complement existing sources of measurement and provide value in new types of measurements. A growing body of evidence has found that geolocation-derived data, including number of locations visited, distance traveled, and time spent at prespecified locations, are associated with symptoms of bipolar disorder and schizophrenia.[37] Additionally, geolocation-derived digital phenotyping was more strongly associated with symptoms of bipolar disorder and schizophrenia than other information, including self-reported symptoms.[38] Thus, smartphone-based geolocation-derived digital phenotyping may augment MIC by relying on digital indicators like geolocation sensors to complement self-reported information. This may improve outcomes through enhancing precision and consistency in disease assessment, monitoring, and care provision. More research is needed to better understand how digital phenotyping incorporated into MIC for mental health may affect patient engagement and treatment efficacy.

Natural Language Processing

Natural language processing (NLP) is a field of AI that focuses on computer processing language and text data. While NLP can be used for digital phenotyping, it is worth devoting a separate section in this paper to NLP due to the increasing research interest in this branch of AI. NLP includes tracking the use of language in conversation formats like chats, emails, social media posts, text from therapy sessions, intake interviews, EHR data, and clinical notes to enable proactive mental health care and early diagnosis.[39,40] Due to the proliferation of text and language-data sources (e.g., widespread use of texting and social media) and the technical advancements occurring in NLP and its subset of large language models (LLMs; e.g., Chat GPT), it is likely its use will become increasingly common in mental health measurement.[41]

NLP works by detecting linguistic patterns that have been linked to mental health conditions and emotional state patterns.[42,43] This can be done by analyzing the choice and frequency of words (e.g., frequence of first-person singular pronouns)[44], linguistic style of written responses, length of speech, and frequency of communication.[45] A 2022 review of 399 studies found that NLP is increasingly being applied to mental health research across types of mental illnesses (e.g., suicide, depression), data sources (e.g., social media, interviews, EHR), and types of AI (e.g., supervised learning, deep learning).[46] While NLP for mental health conditions is still in the early stages of clinical application, there are many ways it could ultimately be integrated into clinical care. For instance, NLP may be used more widely in the future to analyze messaging-based platforms used between clinical visits, social media posts, or linguistic patterns during a therapy session.[47] Importantly, research shows that NLP can be used to detect suicide significantly more accurately than human raters, and its use in suicide prediction may grow.[48]

It is likely that NLP and LLMs will increasingly change how data captured from electronic health records (EHRs) are used to screen and monitor patients. Tellingly, many hospital systems have begun to partner with tech companies to integrate AI into their EHRs. Oracle has integrated AI into its Cerner EHR,[49] Epic has partnered with Microsoft to integrate AI into their EHR, [50] Mayo Clinic is working with Sandbox AQ,[51] Google is working with Meditech,[52] and Abridge is working with EPIC and Kaiser Permanente.[53]

There are many ways that NLP technology is starting to be tested and used to analyze data in EHRs, including clinical measures, clinician notes, comorbid conditions, and sociodemographic factors, to predict symptoms of severe mental illness and suicidal ideation and attempts.[54] For example, the Department of Veterans Affairs (VA) and National Institute of Mental Health scientists developed an expansive suicide mortality risk prediction algorithm using Veterans Health Administration (VHA) electronic health records, enabling the VA to provide a more targeted, enhanced outreach and care program for veterans identified at high risk of suicide.[55] More recently, the VA added the use of NLP to tap into unstructured EHR data, such as clinical notes, to enhance the accuracy of this risk prediction algorithm, resulting in an additional 19% accuracy. Thus, it showed that NLP-supplemented predicative models improve the benefits of the predictive model overall.[56]

Considerations in the Widespread Use of Emerging AI Models in Mental Healthcare

Despite the potential AI advancements to improve population mental health, these approaches bring newfound risk and ethical considerations. For example, hallucinations or the incorrect or misleading results produced by generative AI may have unique harms in the context of mental health. Another significant consideration when using AI in mental healthcare is privacy. As more data about our everyday lives are available, it is important to consider both what may be sensitive data about people's mental health conditions and what other details about their lives may be used as data.

Additionally, the potential for bias in AI may be particularly problematic in the mental health care setting. One pathway for bias is the data used to train and test AI systems. Data may lack demographic diversity or not be representative of the end users (e.g., data set is primarily based on people with light skin). Bias in data can also result from bias or discrimination in our society that is reflected in data patterns resulting from historical discrimination or barriers to care for certain groups.[57]

Specific ethical considerations related to the domains described above are discussed below.

Digital Phenotyping

In the case of digital phenotyping, much of the data collected through wearables are dependent on skin sensors. However, few studies have examined the impact of skin tone on the accuracy of the data derived from consumer-grade wearables. Facial recognition technologies have been shown to perform less accurately on people with darker skin and may perform less well on people with developmental disabilities who may have limited facial expressions to inform the models.[58,59]

While wearables, smartphones, and social media generate troves of new data that can be used to help understand mental health status, data related to mental health are often sensitive, which raises questions about patient/consumer privacy. In mental health apps more broadly, sensitive data have been leaked and sold to third-party platforms, such as Meta, for advertisement purposes. This can lead to many imaginable consequences (e.g., alcohol advertisements to someone who has discussed substance abuse issues).[60] Privacy laws protecting consumers are not sufficient. The Health Insurance Portability and Accountability Act's (HIPAA) does not cover most mental health apps, and greater consumer protection is needed.[61]

Finally, when making predictions in health care, we often do not have the precise data to predict the actual outcomes or future mental health states we are concerned with and instead use proxy measures. Previous research has shown the risk of bias in AI to be the highest when proxy measures are used.[62] It is important to think critically about the symptom measures used in digital phenotyping, specifically if they predict risk accurately across all populations.

Natural Language Processing (NLP)

AI and data-driven algorithms, including NLP, may propagate existing and historical health disparities. A review of the uses of NLP in mental health identified significant biases with respect to religion, race, gender, nationality, sexuality, and age.[63] It is important to assess and mitigate bias in NLP for mental health through approaches such as tailoring NLP debiasing techniques for mental health, integrating statistical fairness approaches, and developing cross-disciplinary teams, can mitigate these potential biases. For instance, during the evaluating phase of AI algorithms, it is important to determine if the AI algorithm, such as analyzing voice recognition for mental health prediction, is equally accurate across language, varying accents, cultures, and abilities. NLP algorithms and training datasets for mental health applications used clinically or commercially should also be transparently standardized and audited.[64]

Recently, a READI (readiness for AI deployment and implementation) framework has been developed to evaluate and report if AI applications are ready for clinical deployment in mental health contexts based on safety, confidentiality/privacy, equity, effectiveness, and implementation concerns.[65] The READI framework provides an approach to evaluating new AI technologies both before large-scale clinical deployment and regularly after deployment. As an example, many AI-based chatbots for mental health treatment may meet many READI criteria, including safety and privacy/confidentiality, but they do not meet other criteria, like effectiveness or engagement.[66] As AI continues to develop rapidly, increased use and adoption of frameworks like READI will be important to ensure that risks are mitigated for AI approaches for MIC for mental health.

Conclusion

Mental health leaders are increasingly realizing that MIC needs to be at the core of mental health treatment. Advances in AI enable new MIC mental health capabilities, including a shift towards digital biomarkers, greater insights from existing and untapped data, and new applications for mental health chatbots. These applications can improve care quality, lower costs of care, improve outcomes, and accelerate accountability for mental health care providers and consumers. However, more research is needed to develop and validate AI for MIC for mental health, especially in applications within the clinical setting. New ethical guidelines are needed to facilitate better development, adoption, and deployment of AI for MIC for mental health. Equity must be an essential component of any technological advancement designed to improve the ease, accuracy, and quality of MIC for mental health.

Terminology

Active data collection: In digital phenotyping, active data collection includes data from specific assessments at home or within a clinical setting.

Artificial Intelligence: An umbrella term for the development of computer systems and algorithms that can perform tasks commonly requiring human intelligence

Augmented Intelligence: A conceptualization of artificial intelligence that focuses on AI's assistive role, emphasizing that its design enhances human intelligence rather than replaces it[67]

Chatbot: A computer program designed to simulate conversation with human users.

Digital phenotyping: An innovative approach concerning the extraction and analysis of information about behavior, cognition, and mood through digital tools collected on devices people now commonly possess, particularly smartphones and wearables[68]

Large language models (LLMs): Technologies built on AI that technologies can read, summarize, and generate text. LLMs can perform tasks, such as serving as conversational agents (chatbots), translating language, or writing text. LLMs have a wide range of abilities.

Measurement-informed care (MIC): An approach that integrates the regular use of patient-reported outcome measures and other clinical outcome measures to inform treatment decisions and monitor progress over time.[69]

Natural language processing (NLP): A subfield of AI that enables machines to process, comprehend, interpret, and generate human language[70]

Passive data collection: In digital phenotyping, passive data collection includes data continuously collected from smartphone usage, driving behavior, wearable devices, social media activity, etc. without a specific assessment administered to collect the data. Passive data collection may entail information continuously collected from smartphone usage (e.g., GPS and movement patterns, keystrokes, etc.), driving behavior, biometric data from wearable devices, or social media activity.

Phenotype: The observable characteristics of an individual

References

[1] American Medical Association. *Augmented intelligence in medicine*. (2024). American Medical Association. https://www.ama-assn.org/practice-management/digital/augmented-intelligence-medicine
[2] Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of Routine Feedback to Clinicians on Mental Health Outcomes of Youths: Results of a Randomized Trial. *Psychiatric Services*, *62*(12), 1423–1429. https://doi.org/10.1176/appi.ps.002052011
[3] National Council for Mental Wellbeing. *National Council Introduces Measurement-informed Care*. (2024). National Council for Mental Wellbeing. https://www.thenationalcouncil.org/news/national-council-introduces-measurement-informed-care/
[4] Jensen-Doss, A., Douglas, S., Phillips, D. A., Gencdur, O., Zalman, A., & Gomez, N. E. (2020). Measurement-based care as a practice improvement tool: Clinical and organizational applications in youth mental health. *Evidence-Based Practice in Child and Adolescent Mental Health*, *5*(3), 233–250. https://doi.org/10.1080/23794925.2020.1784062
[5] Alter, C.L., Mathias, A., Zahniser, J., Shah, S., Schoenbaum, M., Harbin, H.T., McLaughlin, R, & Sieger-Walls, J. (2021). Measurement-Based Care in the Treatment of Mental Health and Substance Use Disorders. Dallas, TX: Meadows Mental Health Policy Institute. mmhpi.org
[6] Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement-based care. *Psychiatric Services*, *68*(2), 179–188. https://doi.org/10.1176/appi.ps.201500439
[7] Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing Measurement-Based Care in Behavioral Health: A Review. *JAMA Psychiatry*, *76*(3), 324. https://doi.org/10.1001/jamapsychiatry.2018.3329
[8] Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing Measurement-Based Care in Behavioral Health: A Review. *JAMA Psychiatry*, *76*(3), 324. https://doi.org/10.1001/jamapsychiatry.2018.3329
[9] Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An Investigation of Self-Assessment Bias in Mental Health Providers. *Psychological Reports*, *110*(2), 639–644. https://doi.org/10.2466/02.07.17.PR0.110.2.639-644
[10] Fortney, John, Rebecca Sladek, & Jürgen Unützer. (2015). *Fixing behavioral health care in America: A national call for integrating & coordinating specialty behavioral health care with the medical system*. The Kennedy Forum. https://www.patrickjkennedy.net/app/uploads/2020/05/KennedyForum-BehavioralHealth_FINAL_3.pdf
[11] Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing measurement-based care in behavioral health: A review. *JAMA Psychiatry*, *76*(3), 324. https://doi.org/10.1001/jamapsychiatry.2018.3329
[12] Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing Measurement-Based Care in Behavioral Health: A Review. *JAMA Psychiatry*, *76*(3), 324. https://doi.org/10.1001/jamapsychiatry.2018.3329
[13] Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing Measurement-Based Care in Behavioral Health: A Review. *JAMA Psychiatry*, *76*(3), 324. https://doi.org/10.1001/jamapsychiatry.2018.3329
[14] National Council for Mental Wellbeing. *National Council Introduces Measurement-informed Care*. (2024). National Council for Mental Wellbeing. https://www.thenationalcouncil.org/news/national-council-introduces-measurement-informed-care/
[15] O'Leary, A., Lahey, T., Lovato, J., Loftness, B., Douglas, A., Skelton, J., Cohen, J. G., Copeland, W. E., McGinnis, R. S., & McGinnis, E. W. (2024). Using Wearable Digital Devices to Screen Children for Mental Health Conditions: Ethical Promises and Challenges. *Sensors (Basel, Switzerland)*, *24*(10), 3214. https://doi.org/10.3390/s24103214
[16] Flores, J. P., Kahn, G., Penfold, R. B., Stuart, E. A., Ahmedani, B. K., Beck, A., Boggs, J. M., Coleman, K. J., Daida, Y. G., Lynch, F. L., Richards, J. E., Rossom, R. C., Simon, G. E., & Wilcox, H. C. (2024). Adolescents Who Do Not Endorse Risk via the Patient Health Questionnaire Before Self-Harm or Suicide. *JAMA Psychiatry*, *81*(7), 717. https://doi.org/10.1001/jamapsychiatry.2024.0603

[17] Cohen, A., Naslund, J., Lane, E., Bhan, A., Rozatkar, A., Mehta, U. M., Vaidyam, A., Byun, A. (Jin S., Barnett, I., & Torous, J. (2024). Digital phenotyping data and anomaly detection methods to assess changes in mood and anxiety symptoms across a transdiagnostic clinical sample. *Acta Psychiatrica Scandinavica*, acps.13712. https://doi.org/10.1111/acps.13712
[18] Fortney, John, Rebecca Sladek, & Jürgen Unützer. (2015). *Fixing behavioral health care in America: A national call for integrating & coordinating specialty behavioral health care with the medical system*. The Kennedy Forum. https://www.patrickjkennedy.net/app/uploads/2020/05/KennedyForum-BehavioralHealth_FINAL_3.pdf
[19] Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing measurement-based care in behavioral health: A review. *JAMA Psychiatry*, *76*(3), 324. https://doi.org/10.1001/jamapsychiatry.2018.3329
[20] Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing Measurement-Based Care in Behavioral Health: A Review. *JAMA Psychiatry*, *76*(3), 324. https://doi.org/10.1001/jamapsychiatry.2018.3329
[21] *Increasing Measurement-Based Assessment and Care for People with Serious Mental Illness*. (2024). Meadows Mental Health Policy Institute. https://mmhpi.org/project/increasing-measurement-based-assessment-and-care-for-people-with-serious-mental-illness/
[22] *Our Care Model | Headspace for Organizations, Formerly Ginger*. (n.d.). Headspace for Organizations. Retrieved August 22, 2024, from https://organizations.headspace.com/care-model
[23] Chan, W. W., Fitzsimmons-Craft, E. E., Smith, A. C., Firebaugh, M.-L., Fowler, L. A., DePietro, B., Topooco, N., Wilfley, D. E., Taylor, C. B., & Jacobson, N. C. (2022). The Challenges in Designing a Prevention Chatbot for Eating Disorders: Observational Study. *JMIR Formative Research*, *6*(1), e28003. https://doi.org/10.2196/28003
[24] Lim, S. M., Shiau, C. W. C., Cheng, L. J., & Lau, Y. (2022). Chatbot-Delivered Psychotherapy for Adults With Depressive and Anxiety Symptoms: A Systematic Review and Meta-Regression. *Behavior Therapy*, *53*(2), 334–347. https://doi.org/10.1016/j.beth.2021.09.007
[25] Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. *Journal of Medical Internet Research*, *21*(9), e14567. https://doi.org/10.2196/14567
[26] Torous, J., Nicholas, J., Larsen, M. E., Firth, J., & Christensen, H. (2018). Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. *BMJ Ment Health*, *21*(3), 116–119. https://doi.org/10.1136/eb-2018-102891
[27] Kellogg, K. C., & Sadeh-Sharvit, S. (2022). Pragmatic AI-augmentation in mental healthcare: Key technologies, potential benefits, and real-world challenges and solutions for frontline clinicians. *Frontiers in Psychiatry*, *13*, 990370. https://doi.org/10.3389/fpsyt.2022.990370
[28] Sadeh-Sharvit, S., & Hollon, S. D. (2020). Leveraging the Power of Nondisruptive Technologies to Optimize Mental Health Treatment: Case Study. *JMIR Mental Health*, *7*(11), e20646. https://doi.org/10.2196/20646
[29] Johnson, K. B., Wei, W., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2021). Precision Medicine, AI, and the Future of Personalized Health Care. *Clinical and Translational Science*, *14*(1), 86–93. https://doi.org/10.1111/cts.12884
[30] Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, *25*, e46778. https://doi.org/10.2196/46778
[31] Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, *25*, e46778. https://doi.org/10.2196/46778
[32] Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, *25*, e46778. https://doi.org/10.2196/46778
[33] Fraccaro, P., Beukenhorst, A., Sperrin, M., Harper, S., Palmier-Claus, J., Lewis, S., Van Der Veer, S. N., & Peek, N. (2019). Digital biomarkers from geolocation data in bipolar disorder and schizophrenia: A systematic review. *Journal of the American Medical Informatics Association*, *26*(11), 1412–1420. https://doi.org/10.1093/jamia/ocz043
[34] Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, *25*, e46778. https://doi.org/10.2196/46778

---

[35] Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, *25*, e46778. https://doi.org/10.2196/46778
[36] Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, *25*, e46778. https://doi.org/10.2196/46778
[37] Fraccaro, P., Beukenhorst, A., Sperrin, M., Harper, S., Palmier-Claus, J., Lewis, S., Van Der Veer, S. N., & Peek, N. (2019). Digital biomarkers from geolocation data in bipolar disorder and schizophrenia: A systematic review. *Journal of the American Medical Informatics Association*, *26*(11), 1412–1420. https://doi.org/10.1093/jamia/ocz043
[38] Fraccaro, P., Beukenhorst, A., Sperrin, M., Harper, S., Palmier-Claus, J., Lewis, S., Van Der Veer, S. N., & Peek, N. (2019). Digital biomarkers from geolocation data in bipolar disorder and schizophrenia: A systematic review. *Journal of the American Medical Informatics Association*, *26*(11), 1412–1420. https://doi.org/10.1093/jamia/ocz043
[39] Minerva, F., & Giubilini, A. (2023). Is AI the Future of Mental Healthcare? *Topoi*, *42*(3), 809–817. https://doi.org/10.1007/s11245-023-09932-3
[40] Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *Npj Digital Medicine*, *5*(1), 46. https://doi.org/10.1038/s41746-022-00589-7
[41] Stade, E. C., Wiltsy Stirman, S., Ungar, L., Boland, C., Schwartz, H. A., Yaden, D. B., Sedoc, J., DeRubeis, R. J., Willer, R., Kim, J. P., & Eichstaedt, J. C. (2024). *Toward Responsible Development and Evaluation of LLMs in Psychotherapy | Stanford HAI*. HAI Stanford University. https://hai.stanford.edu/toward-responsible-development-and-evaluation-llms-psychotherapy
[42] Minerva, F., & Giubilini, A. (2023). Is AI the Future of Mental Healthcare? *Topoi*, *42*(3), 809–817. https://doi.org/10.1007/s11245-023-09932-3
[43] Timmons, A. C., Duong, J. B., Fiallo, N. S., Lee, T., Vo, H. P. Q., Ahle, M. W., Comer, J. S., Brewer, L. C., Frazier, S. L., & Chaspari, T. (2023). A Call to Action on Assessing and Mitigating Bias in Artificial Intelligence Applications for Mental Health. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *18*(5), 1062–1096. https://doi.org/10.1177/17456916221134490
[44] Edwards, T., & Holtzman, N. S. (2017). A meta-analysis of correlations between depression and first personsingular pronoun use. *Journal of Research in Personality*, *68*, 63–68. https://doi.org/10.1016/j.jrp.2017.02.005
[45] Yahya, N. H., & Abdul Rahim, H. (2023). Linguistic markers of depression: Insights from english-language tweets before and during the COVID-19 pandemic. *Language and Health*, *1*(2), 36–50. https://doi.org/10.1016/j.laheal.2023.10.001
[46] Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *Npj Digital Medicine*, *5*(1), 46. https://doi.org/10.1038/s41746-022-00589-7
[47] Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *Npj Digital Medicine*, *5*(1), 46. https://doi.org/10.1038/s41746-022-00589-7
[48] Young, J., Bishop, S., Humphrey, C., & Pavlacic, J. M. (2023). A review of natural language processing in the identification of suicidal behavior. *Journal of Affective Disorders Reports*, *12*, 100507. https://doi.org/10.1016/j.jadr.2023.100507
[49] Miliard, M. (2023, September 19). *Oracle Cerner adds generative AI to its EHR platforms*. Healthcare IT News. https://www.healthcareitnews.com/news/oracle-cerner-adds-generative-ai-its-ehr-platforms
[50] *Turner, B. (2023, April 17). Epic, Microsoft bring GPT-4 to EHRs. Modern Healthcare. https://www.modernhealthcare.com/digital-health/himss-2023-epic-microsoft-bring-openais-gpt-4-ehrs*
[51] Landi, H. (2024, August 8). *Mayo Clinic, SandboxAQ partner to test out AI-based MCG imaging*. https://www.fiercehealthcare.com/ai-and-machine-learning/mayo-clinic-taps-startup-sandboxaq-study-new-tech-quantum-sensing-and-ai
[52] Adams, K. (2023, September 10). *How Meditech Plans to Integrate Google's Generative AI Into Its EHR*. MedCity News. https://medcitynews.com/2023/09/google-meditech-generative-ai-technology-healthcare/

[53] Landi, H. (2024, August 14). *Kaiser Permanente taps Abridge's AI tool for thousands of docs*. https://www.fiercehealthcare.com/health-tech/kaiser-permanente-rolls-out-abridges-gen-ai-clinical-tech-across-40-hospitals-60
[54] Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C., & Jeste, D. V. (2019). Artificial Intelligence for Mental Health and Mental Illnesses: An Overview. *Current Psychiatry Reports*, *21*(11), 116. https://doi.org/10.1007/s11920-019-1094-0
[55] McCarthy, J. F., Cooper, S. A., Dent, K. R., Eagan, A. E., Matarazzo, B. B., Hannemann, C. M., Reger, M. A., Landes, S. J., Trafton, J. A., Schoenbaum, M., & Katz, I. R. (2021). Evaluation of the Recovery Engagement and Coordination for Health–Veterans Enhanced Treatment Suicide Risk Modeling Clinical Program in the Veterans Health Administration. *JAMA Network Open*, *4*(10), e2129900. https://doi.org/10.1001/jamanetworkopen.2021.29900
[56] Levis, M., Levy, J., Dent, K. R., Dufort, V., Gobbel, G. T., Watts, B. V., & Shiner, B. (2023). Leveraging Natural Language Processing to Improve Electronic Health Record Suicide Risk Prediction for Veterans Health Administration Users. *The Journal of Clinical Psychiatry*, *84*(4), 22m14568. https://doi.org/10.4088/JCP.22m14568
[57] Faber, S. C., Khanna Roy, A., Michaels, T. I., & Williams, M. T. (2023). The Weaponization of Medicine: Early Psychosis in the Black Community and the Need for Racially Informed Mental Healthcare. *Frontiers in Psychiatry*, *14*, 1098292. https://doi.org/10.3389/fpsyt.2023.1098292
[58] Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html
[59] Martinez-Martin, N. (2019). What Are Important Ethical Implications of Using Facial Recognition Technology in Health Care? *AMA Journal of Ethics*, *21*(2), E180-187. https://doi.org/10.1001/amajethics.2019.180
[60] Park, E., & West, D. (n.d.). *Why mental health apps need to take privacy more seriously*. Brookings. Retrieved May 21, 2024, from https://www.brookings.edu/articles/why-mental-health-apps-need-to-take-privacy-more-seriously/
[61] Stade, B., Eichstaedt, johannes C., Kim, J. P., & Stirman, S. W. (2024). *Readiness Evaluation for AI Deployment and Implementation for Mental Health: A Review and Framework*. OSF. https://doi.org/10.31234/osf.io/8zqhw
[62] Obermeyer, Z., Rebecca Nissan, Michael Stern, Stephanie Eaneff, Emily Joy Bembeneck, & Sendhil Mullainathan. (2021). *Algorithmic Bias Playbook*. Chicago Booth. https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias/playbook
[63] Straw, I., & Callison-Burch, C. (2020). Artificial Intelligence in mental health and the biases of language based models. *PLoS ONE, 15*(12), e0240376. https://doi.org/10.1371/journal.pone.0240376
[64] Caliskan, A. (2021). *Detecting and mitigating bias in natural language processing*. Brookings. https://www.brookings.edu/articles/detecting-and-mitigating-bias-in-natural-language-processing/
[65] Stade, E., Stirman, S., Ungar, L., Boland, C., Schwartz, H.A., Yaden, D.B., Sedoc, J., DeRubeis, R.J., Willer, R., Kim, J.P., & Eichstaedt, J.C. (2024). *Toward Responsible Development and Evaluation of LLMs in Psychotherapy.* Stanford University. https://hai.stanford.edu/toward-responsible-development-and-evaluation-llms-psychotherapy. https://hai.stanford.edu/toward-responsible-development-and-evaluation-llms-psychotherapy
[66] Stade, E., Stirman, S., Ungar, L., Boland, C., Schwartz, H.A., Yaden, D.B., Sedoc, J., DeRubeis, R.J., Willer, R., Kim, J.P., & Eichstaedt, J.C. (2024). *Toward Responsible Development and Evaluation of LLMs in Psychotherapy.* Stanford University. https://hai.stanford.edu/toward-responsible-development-and-evaluation-llms-psychotherapy
[67] American Medical Association. *Augmented intelligence in medicine*. (2024). American Medical Association. https://www.ama-assn.org/practice-management/digital/augmented-intelligence-medicine
[68] Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, *25*, e46778. https://doi.org/10.2196/46778
[69] Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of Routine Feedback to Clinicians on Mental Health Outcomes of Youths: Results of a Randomized Trial. *Psychiatric Services*, *62*(12), 1423–1429. https://doi.org/10.1176/appi.ps.002052011
[70] Thakkar, A., Gupta, A., & De Sousa, A. (2024). Artificial intelligence in positive mental health: A narrative review. *Frontiers in Digital Health*, *6*, 1280235. https://doi.org/10.3389/fdgth.2024.1280235