Position: AI Will Transform Neuropsychology Through Mental Health Digital Twins for Dynamic Mental Health Care, Especially for ADHD

Neil Natarajan[1]
[1]University of Oxford

Sruthi Viswanathan[1]

Xavier Roberts-Gaal[2]
[2]Harvard University

Michelle Marie Martel[3]
[3]University of Kentucky

Abstract

Static solutions don't serve a dynamic mind. Thus, we advocate a shift from static mental health diagnostic assessments to continuous, artificial intelligence (AI)-driven assessment. Focusing on Attention-Deficit/Hyperactivity Disorder (ADHD) as a case study, we explore how generative AI has the potential to address current capacity constraints in neuropsychology, potentially enabling more personalized and longitudinal care pathways. In particular, AI can efficiently conduct frequent, low-level experience sampling from patients and facilitate diagnostic reconciliation across care pathways. We envision a future where mental health care benefits from continuous, rich, and patient-centered data sampling to dynamically adapt to individual patient needs and evolving conditions, thereby improving both accessibility and efficacy of treatment. We further propose the use of mental health digital twins (MHDTs) – continuously updated computational models that capture individual symptom dynamics and trajectories – as a transformative framework for personalized mental health care. We ground this framework in empirical evidence and map out the research agenda required to refine and operationalize it.

1 Introduction

Mental health conditions are dynamic; presentation and impact shift with treatment response, life events, and development [10]. Attention-Deficit/Hyperactivity Disorder (ADHD) illustrates this dynamism well: it often persists into adulthood and affects roughly 4% of adults worldwide, yet its symptomatology varies across contexts and over the lifespan [13]. Hyperactivity tends to dominate in preschool years, whereas inattention becomes central from middle childhood onward; DSM subtypes show limited temporal stability [21, 22]. Traditional point-in-time assessments struggle to capture these dynamics [11, 18]. We argue that recent advances in generative AI can directly address these constraints and enable patient-centered, longitudinal assessment. Throughout, we use ADHD as a running case to illustrate both the limitations of current practice and the potential of AI-assisted approaches.

This position paper makes three contributions: (i) it explains why episodic neuropsychological assessment is misaligned with dynamic conditions, using ADHD as exemplar; (ii) it proposes using mental health digital twin (MHDT) systems – a clinician-supervised, multimodal, continuously updated model – to support longitudinal diagnostic reconciliation in response; and (iii) it outlines a concrete research and operational agenda to validate, govern, and safely deploy such systems.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: The Second Workshop on GenAI for Health: Potential, Trust, and Policy Compliance.

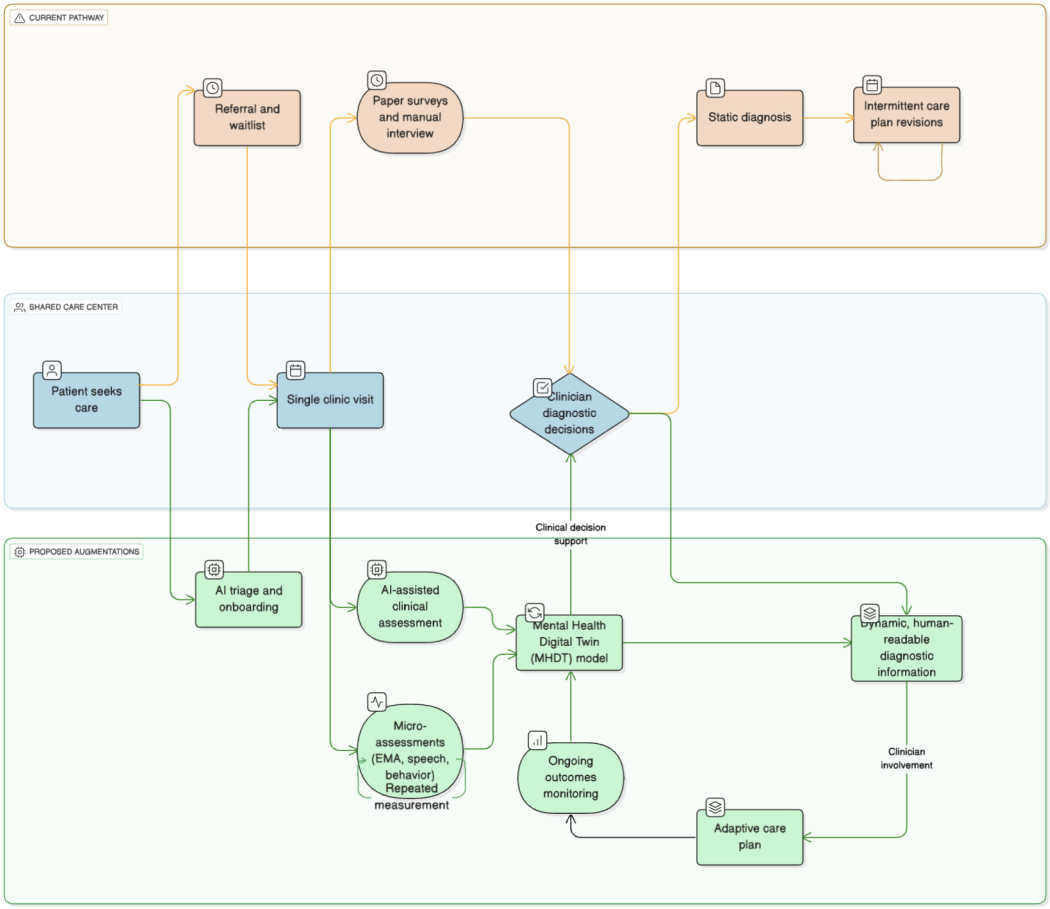

Figure 1: Comparison of traditional episodic ADHD diagnostic process versus continuous AI-assisted diagnostic companion approach. The traditional model (top) shows discrete assessment points separated by long intervals, while the continuous model (bottom) demonstrates ongoing data collection, dynamic modeling, and iterative diagnostic refinement.

2 Limitations of Current Neuropsychological Assessment

The evolution of many mental health conditions over time complicates single-occasion diagnostics. For ADHD, adult diagnosis often relies on retrospective reconstruction of childhood onset, frequently without reliable records [9]. Longitudinal cohort evidence identifies a fluctuant subgroup whose diagnostic status changes with age, consistent with relapsing–remitting patterns or sensitivity to developmental context [28]. Taken together with developmental shifts in symptom structure and instability of Diagnostic and Statistical Manual of Mental Disorders (DSM) subtypes [21, 22], this motivates longitudinal, dimensional assessment in place of single-point categorical assignment.

Current neuropsychological workflows typically compress evaluation into episodic interactions over short intervals [9, 11]. Such snapshots miss within-person variation across time and settings, creating risks of misclassification as new information renders prior labels obsolete without systematic reconciliation [3, 11]. In practice, many adult ADHD determinations are made via rating scales and brief clinical encounters [15], and re-assessment requires new visits, extending delays and costs [9, 11].

Structural constraints further limit ecological validity: primary care lacks capacity for longitudinal, context-rich evaluations, while specialty clinics face long waitlists [9, 20]. Office-bound assessments struggle to capture context effects (home, school, work) and time-varying presentations [26]. Consequently, transient states can be frozen into durable labels [21, 22].

3 Dynamic Approaches to Diagnosis

Ecological momentary assessment (EMA), a method for collecting real-time data on symptoms, behaviors, and experiences in participants' natural environments through repeated brief surveys delivered via smartphones or other devices, has emerged as a promising approach to continuous monitoring of mental health symptoms [19, 25]. However, purely EMA-based monitoring has practical limitations: it relies on self-report, creates burden and attrition risk, and may itself alter symptom reports (measurement reactivity) [19, 25]. EMA also under-samples nonlinguistic and physiologic correlates relevant to attention and affective regulation [25].

Going beyond EMA, AI systems can coordinate adaptive prompts and fuse EMA with other modalities through ongoing, clinician-supervised conversational loops. Unlike human clinicians, these systems can interact continuously, enabling longitudinal data collection before, during, and after the traditional diagnostic process [5]. Furthermore, generative AI systems have demonstrated capacity for real-time, personalized interactions that adapt to user behavior, increasing patient engagement and allowing for on-demand interactions based on the most recent changes in symptom presentation [29]. Indeed, these AI systems could prompt patients for interactions across various environments, gathering data difficult for clinicians to obtain directly, enabling more holistic diagnostic understanding. These capabilities present an opportunity to build high-touch diagnostic systems capable of capturing information over time, improving on retrospective analyses.

This concept may be pushed even further through the development of mental health digital twins (MHDTs), i.e., continuously updated, individualized computational models of a patient's mental states and likely trajectories [32]. While the MHDT remains largely theoretical [36], it provides a blueprint for future research and development. These AI models would track and predict patient presentation by integrating data from patient interactions, behavioral patterns, and physiological responses, moving beyond single-point diagnoses to continuous, personalized care [5]. By using these tools for longitudinal diagnosis in addition to care, clinicians may glean insights into evolving conditions beyond the scope of what point assessments, even ones assisted by AI systems, can capture [32].

An ADHD-focused diagnostic companion built on MHDT principles would integrate: (i) structured conversational history-taking aligned to DSM constructs and comorbidity screening [17, 36]; (ii) brief at-home cognitive and performance tasks (e.g., continuous performance test [CPT] variants) for longitudinal tracking of attention and inhibition [8]; (iii) short speech probes for prosodic and linguistic markers associated with ADHD [4, 38]; and (iv) optional wearable-derived behavioral and neurophysiologic signals when appropriate and consented [23]. Clinician-in-the-loop triage and reporting remain central to ensure fitness-to-practice and guard against over-pathologizing normal variability.

The MHDT would explicitly model symptom dynamics, secondary features (e.g., speech prosody and language markers, home-based cognitive performance tasks, or wearable signals), and context effects in latent space. This allows for forward simulation and counterfactual reasoning, enabling clinicians to test hypotheses about the impact of environmental structure or medication adjustments under uncertainty bounds. Evidence supporting each feature family includes prosodic and linguistic markers [4, 38], at-home cognitive tasks [8], wearable neurophysiology in mood prediction [23], and known EMA dynamics such as measurement reactivity [19, 25], under-sampling of nonlinguistic and physiologic correlates [25], or predictable cyclic modulators (e.g., menstrual cycle) [30]. In combination, these digital twins could transform the diagnostic reconciliation process by continuously analyzing new information and updating diagnoses, moving beyond repeated reconciliation to a fundamentally evolving understanding of patient conditions.

4 A Mental Health Digital Twin Research Agenda

Several challenges stand in the way of making MHDT a reality. Data quality and generalizability remain open questions, as signal-behavior relationships may vary across demographics and contexts [2]. Continuous monitoring introduces privacy, consent, and governance complexities beyond traditional records [34, 39]. Bias and fairness concerns must be proactively addressed, given evidence of differential performance across subgroups and potential racial bias in model-mediated recommendations [6, 35]. There is risk of diagnostic drift or over-pathologizing variability if update thresholds are not

conservative and well-calibrated [18]. Finally, integrating with existing workflows while avoiding alert fatigue, ensuring safety guardrails, and maintaining human accountability are essential [33, 39].

Research milestones:

- *Validated AI-delivered intake probes.* Build and evaluate AI applications for history-taking and standardized tasks (modalities as in Section 3), with longitudinal scoring to support clinician diagnosis. Validate via uplift-style trials comparing diagnostic accuracy and cost to clinician-only baselines [16]. Demonstrate non-inferiority or superiority while reducing assessment burden.
- *Multimodal MHDT modeling.* Develop calibrated fusion of EMA, speech, behavioral, and physiologic data with quantified uncertainty; detect signal and population drift; and maintain provenance across model updates [32].
- *Diagnostic reconciliation policies.* Specify conservative posterior-thresholds for label changes, minimum evidence windows, rate limits on update frequency, and confirmatory clinician reviews; quantify false-update risk and clinician workload under different policies [3, 33].

Operationalization priorities:

- *Informed, revocable consent and patient control.* Granular, ongoing consent with clear opt-in/out and patient portals to view, pause, or delete data where permissible [34, 39].
- *Inclusive calibration and drift safeguards.* Use conservative thresholds, confirmatory assessments before label changes, and rate-limit updates [18]. Calibrate to account for known EMA dynamics [19, 30], so updates reflect durable change rather than transient state fluctuations.
- *Data governance, protection, and response.* Data minimization, on-device processing when feasible, and encryption in transit/at rest [34]; role-restricted access, minimal retention, and lineage/versioning [39]; and transparent incident response aligned with GDPR/HIPAA requirements [31].
- *Transparency and interpretability.* Clinician-readable explanations, uncertainty estimates, and data/model provenance; user interface (UI) that supports shared decision-making via interpretable longitudinal visualizations [33, 39].
- *Human oversight and governance.* Explicit override pathways, audit trails, and multidisciplinary governance to review audits, approve model changes, and manage deprecation [3, 33, 39].
- *Bias and fairness management.* Pre-specified subgroup audits, equity-drift monitoring, stakeholder reviews, and mitigations for differential performance and racial bias [2, 6, 35].
- *Safety and escalation.* Escalation protocols for risk signals integrated to minimize alert fatigue while ensuring timely response [33, 39].

5 A Vision for Personalized Mental Health Models Beyond Diagnostic Categories

Looking further into the future, the MHDT framework may ultimately transcend traditional diagnostic categories altogether. The fundamental purpose of psychiatric diagnosis is to predict treatment response and guide clinical decision-making. However, categorical diagnoses like ADHD, depression, or anxiety often fail to capture heterogeneity and comorbidity [28]. Evidence that categorical subtypes are unstable across development [21] and that symptom loadings shift with age [22] further supports dimensional, individualized modeling. A general "p" factor of psychopathology may explain cross-diagnostic covariance, motivating transdiagnostic modeling [1, 7].

MHDTs offer an alternative: individualized models that directly predict treatment responses, functional outcomes, and support needs. These models may capture each patient's constellation of symptoms, contexts, biological markers, and treatment history, avoiding brittle intermediary labels. For example, rather than "ADHD with comorbid anxiety," a profile would indicate likely response to behavioral interventions, expected benefit from stimulant medication under certain environmental conditions, and the need for accommodations.

This paradigm shift acknowledges the dimensional nature of mental health and better accounts for developmental trajectories and context. It directly optimizes for treatment outcomes and functional improvement rather than labels. It is also evidence-generating: routine prospective measurement coupled with causal inference can enrich both clinical care and research.

However, this vision faces substantial challenges beyond those discussed in Section 2. Medical systems and insurance frameworks are organized around diagnostic categories [40]; and regulatory approval often presumes categorical indications [37]. Clinicians are trained—and incentivized—to think in categorical terms using the DSM and the International Classification of Diseases (ICD) [12] and routine intake practices [27]. Patients, too, can experience diagnostic labels as meaningful identities, with both stigma and persistence of impressions even after retraction [14, 24]. Transitioning to a fully personalized, model-based approach will require both technical and institutional change.

The future of mental wellbeing lies in companions (AI or otherwise) that adapt to evolving psychological landscapes. Static, episodic diagnostic models are poorly matched to dynamic conditions and capacity constraints. A mental health digital twin can combine conversational history-taking, repeatable at-home probes, and selective passive sensing to support diagnostic reconciliation over time under clinician oversight. We propose a research agenda and a roadmap toward MHDTs that are interpretable, auditable, safe, and clinically useful.

Realizing this vision requires rigorous validation across populations, careful attention to privacy and governance, explicit bias and fairness auditing, human-centred design, and seamless integration into clinical workflows. If these milestones are met, continuous AI-supported diagnostics could reduce misclassification, shorten time-to-assessment, reduce cost, and enable more precise, measurement-based care. The ambition is not to replace clinicians, but to equip them with longitudinal, high-fidelity evidence that makes diagnosis and treatment more accurate and responsive to change. The ultimate promise of this position is a future where no mind is left without timely, precise, accessible, and dynamic care.

References

[1] David Adam. "p factor" could open an important window on core attributes of mental health maladies. *Proceedings of the National Academy of Sciences*, 120(41):e2316297120, October 2023. doi: 10.1073/pnas.2316297120.

[2] Daniel A. Adler, Caitlin A. Stamatis, Jonah Meyerhoff, David C. Mohr, Fei Wang, Gabriel J. Aranovich, Srijan Sen, and Tanzeem Choudhury. Measuring algorithmic bias to analyze the reliability of ai tools that predict depression risk using smartphone sensed-behavioral data. *npj Mental Health Research*, 3(1):17, April 2024. ISSN 2731-4251. doi: 10.1038/s44184-024-00057-y.

[3] Sandra Algarin Perneth, Gilberto Perez Rodriguez Garcia, Juan P. Brito, Tejal Gandhi, Carma L. Bylund, Ian G. Hargraves, and Naykky Singh Ospina. Developing a framework for understanding diagnostic reconciliation based on evidence review, stakeholder engagement, and practice evaluation. *Diagnosis (Berlin, Germany)*, 12(1):45–52, December 2024. ISSN 2194-8011. doi: 10.1515/dx-2024-0132.

[4] Juan Barrios, Elena Poznyak, Jessica Lee Samson, Halima Rafi, Simon Gabay, Florian Cafiero, and Martin Debbané. Detecting adhd through natural language processing and stylometric analysis of adolescent narratives. *Frontiers in Child and Adolescent Psychiatry*, 4, May 2025. ISSN 2813-4540. doi: 10.3389/frcha.2025.1519753. URL `https://www.frontiersin.org/journals/child-and-adolescent-psychiatry/articles/10.3389/frcha.2025.1519753/full`.

[5] Archana Reddy Bongurala, Dhaval Save, and Ankit Virmani. Revolutionizing attention deficit hyperactivity disorder with artificial intelligence. *PLOS Mental Health*, 1(6):e0000181, November 2024. ISSN 2837-8156. doi: 10.1371/journal.pmen.0000181.

[6] Ayoub Bouguettaya, Elizabeth M. Stuart, and Elias Aboujaoude. Racial bias in ai-mediated psychiatric diagnosis and treatment: a qualitative comparison of four large language models. *npj Digital Medicine*, 8(1):332, June 2025. ISSN 2398-6352. doi: 10.1038/s41746-025-01746-4.

[7] Avshalom Caspi, Renate M. Houts, Daniel W. Belsky, Sidra J. Goldman-Mellor, HonaLee Harrington, Salomon Israel, Madeline H. Meier, Sandhya Ramrakha, Idan Shalev, Richie Poulton, and Terrie E. Moffitt. The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical psychological science: a journal of the Association for Psychological Science*, 2(2):119–137, March 2014. ISSN 2167-7026. doi: 10.1177/2167702613497473.

[8] K. Cedergren, S. Östlund, J. Åsberg Johnels, E. Billstedt, and M. Johnson. Monitoring medication response in adhd: what can continuous performance tests tell us? *European Archives of Psychiatry and Clinical Neuroscience*, 272(2):291–299, 2022. ISSN 0940-1334. doi: 10.1007/s00406-021-01319-y.

[9] Larry Culpepper and Gregory Mattingly. Challenges in identifying and managing attention-deficit/hyperactivity disorder in adults in the primary care setting: A review of the literature. *Primary Care Companion to The Journal of Clinical Psychiatry*, 12(6):PCC.10r00951, 2010. ISSN 1523-5998. doi: 10.4088/PCC.10r00951pur.

[10] Hermien Dijk and Jochen Mierau. Mental health over the life course: Evidence for a u-shape? *Health Economics*, 32(1):155–174, January 2023. ISSN 1057-9230. doi: 10.1002/hec.4614.

[11] Helen Donovan, Emma Ellis, Laura Cole, Emma Townsend, and Alejandra Cases. Reducing time to complete neuropsychological assessments within a memory assessment service and evaluating the wider impact. *BMJ Open Quality*, 9(3), July 2020. ISSN 2399-6641. doi: 10.1136/bmjoq-2019-000767. URL `https://bmjopenquality.bmj.com/content/9/3/e000767`.

[12] Michael B. First and Drew Westen. Classification for clinical practice: how to make icd and dsm better able to serve clinicians. *International Review of Psychiatry (Abingdon, England)*, 19(5):473–481, October 2007. ISSN 0954-0261. doi: 10.1080/09540260701563429.

[13] Barbara Franke, Giorgia Michelini, Philip Asherson, Tobias Banaschewski, Andrea Bilbow, Jan K. Buitelaar, Bru Cormand, Stephen V. Faraone, Ylva Ginsberg, Jan Haavik, Jonna Kuntsi, Henrik Larsson, Klaus-Peter Lesch, J. Antoni Ramos-Quiroga, János M. Réthelyi, Marta Ribases, and Andreas Reif. Live fast, die young? a review on the developmental trajectories of adhd across the lifespan. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, 28(10):1059–1088, October 2018. ISSN 1873-7862. doi: 10.1016/j.euroneuro.2018.08.001.

[14] Linda Garand, Jennifer H. Lingler, Kyaien O. Conner, and Mary Amanda Dew. Diagnostic labels, stigma, and participation in research related to dementia and mild cognitive impairment. *Research in gerontological nursing*, 2(2):112–121, April 2009. ISSN 1940-4921. doi: 10.3928/19404921-20090401-04.

[15] Lily Hechtman, Lisa R French, Monica Mongia, and Mariya V Cherkasova. Diagnosing adhd in adults: limitations to dsm-iv and dsm-v proposals and challenges ahead. *Neuropsychiatry*, 1 (6):579–590, December 2011. ISSN 1758-2008. doi: 10.2217/npy.11.65.

[16] Maciej Jaskowski and Szymon Jaroszewicz. Uplift modeling for clinical trial data.

[17] Yi Jiang, Qingyang Shen, Shuzhong Lai, Shunyu Qi, Qian Zheng, Lin Yao, Yueming Wang, and Gang Pan. Copiloting diagnosis of autism in real clinical scenarios via llms. (arXiv:2410.05684), October 2024. doi: 10.48550/arXiv.2410.05684. URL `http://arxiv.org/abs/2410.05684`. arXiv:2410.05684 [cs].

[18] Luise Kazda, Katy Bell, Rae Thomas, Kevin McGeechan, Rebecca Sims, and Alexandra Barratt. Overdiagnosis of attention-deficit/hyperactivity disorder in children and adolescents: A systematic scoping review. *JAMA Network Open*, 4(4):e215335, April 2021. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2021.5335.

[19] Traci M. Kennedy, Brooke S. G. Molina, and Sarah L. Pedersen. Change in adolescents' perceived adhd symptoms across 17 days of ecological momentary assessment. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, 53(3):397–412, 2024. ISSN 1537-4424. doi: 10.1080/15374416.2022.2096043.

[20] Leonard F. Koziol and Michael C. Stevens. Neuropsychological assessment and the paradox of adhd. *Applied Neuropsychology: Child*, 1(2):79–89, July 2012. ISSN 2162-2965. doi: 10.1080/21622965.2012.694764.

[21] Benjamin B. Lahey, William E. Pelham, Jan Loney, Steve S. Lee, and Erik Willcutt. Instability of the dsm-iv subtypes of adhd from preschool through elementary school. *Archives of General Psychiatry*, 62(8):896–902, August 2005. ISSN 0003-990X. doi: 10.1001/archpsyc.62.8.896.

[22] Michelle M. Martel, Alexander von Eye, and Joel Nigg. Developmental differences in structure of attention-deficit/hyperactivity disorder (adhd) between childhood and adulthood. *International Journal of Behavioral Development*, 36(4):279–292, July 2012. ISSN 0165-0254. doi: 10.1177/0165025412444077.

[23] Sean H. Merritt and Paul J. Zak. Continuous remote monitoring of neurophysiologic immersion accurately predicts mood. *Frontiers in Digital Health*, 6, August 2024. ISSN 2673-253X. doi: 10.3389/fdgth.2024.1397557. URL `https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2024.1397557/full`.

[24] Amy J. Mickelberg, Bradley Walker, Ullrich K. H. Ecker, and Nicolas Fay. Helpful or harmful? the effect of a diagnostic label and its later retraction on person impressions. *Acta Psychologica*, 248:104420, August 2024. ISSN 0001-6918. doi: 10.1016/j.actpsy.2024.104420.

[25] Carolina Miguelez-Fernandez, Santiago J. de Leon, Itziar Baltasar-Tello, Inmaculada Peñuelas-Calvo, María Luisa Barrigon, Alba Sedano Capdevila, David Delgado-Gómez, Enrique Baca-García, and Juan J. Carballo. Evaluating attention-deficit/hyperactivity disorder using ecological momentary assessment: a systematic review. *Attention Deficit and Hyperactivity Disorders*, 10 (4):247–265, December 2018. ISSN 1866-6647. doi: 10.1007/s12402-018-0261-1.

[26] Hande Musullulu. Evaluating attention deficit and hyperactivity disorder (adhd): a review of current methods and issues. *Frontiers in Psychology*, 16, February 2025. ISSN 1664-1078. doi: 10.3389/fpsyg.2025.1466088. URL `https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2025.1466088/full`.

[27] Ora Nakash, Maayan Nagar, and Yaniv Kanat-Maymon. Clinical use of the dsm categorical diagnostic system during the mental health intake session. *The Journal of Clinical Psychiatry*, 76(7):e862–869, July 2015. ISSN 1555-2101. doi: 10.4088/JCP.14m09214.

[28] Luke J. Norman, Jolie Price, Kwangmi Ahn, Gustavo Sudre, Wendy Sharp, and Philip Shaw. Longitudinal trajectories of childhood and adolescent attention deficit hyperactivity disorder diagnoses in three cohorts. *eClinicalMedicine*, 60:102021, June 2023. ISSN 2589-5370. doi: 10.1016/j.eclinm.2023.102021.

[29] Graziella Orrù, Giulia Melis, and Giuseppe Sartori. Large language models and psychiatry. *International Journal of Law and Psychiatry*, 101:102086, July 2025. ISSN 0160-2527. doi: 10.1016/j.ijlp.2025.102086.

[30] Bethan Roberts, Tory Eisenlohr-Moul, and Michelle M. Martel. Reproductive steroids and adhd symptoms across the menstrual cycle. *Psychoneuroendocrinology*, 88:105–114, February 2018. ISSN 1873-3360. doi: 10.1016/j.psyneuen.2017.11.015.

[31] Wasim Fathima Shah. Preserving privacy and security: A comparative study of health data regulations - gdpr vs. hipaa. *International Journal for Research in Applied Science and Engineering Technology*, 11(8):2189–2199, August 2023. ISSN 23219653. doi: 10.22214/ijraset.2023.55551.

[32] Michael Spitzer, Itai Dattner, and Sigal Zilcha-Mano. Digital twins and the future of precision mental health. *Frontiers in Psychiatry*, 14:1082598, March 2023. ISSN 1664-0640. doi: 10.3389/fpsyt.2023.1082598.

[33] Reed T. Sutton, David Pincock, Daniel C. Baumgart, Daniel C. Sadowski, Richard N. Fedorak, and Karen I. Kroeker. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digital Medicine*, 3:17, February 2020. ISSN 2398-6352. doi: 10.1038/s41746-020-0221-y.

[34] Si Ying Tan, Jennifer Sumner, Yuchen Wang, and Alexander Wenjun Yip. A systematic review of the impacts of remote patient monitoring (rpm) interventions on safety, adherence, quality-of-life and cost-related outcomes. *npj Digital Medicine*, 7(1):192, July 2024. ISSN 2398-6352. doi: 10.1038/s41746-024-01182-w.

[35] Adela C. Timmons, Jacqueline B. Duong, Natalia Simo Fiallo, Theodore Lee, Huong Phuc Quynh Vo, Matthew W. Ahle, Jonathan S. Comer, LaPrincess C. Brewer, Stacy L. Frazier, and Theodora Chaspari. 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, September 2023. ISSN 1745-6916. doi: 10.1177/17456916221134490.

[36] Tao Tu, Mike Schaekermann, Anil Palepu, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang, Brenna Li, Mohamed Amin, Yong Cheng, Elahe Vedadi, Nenad Tomasev, Shekoofeh Azizi, Karan Singhal, Le Hou, Albert Webson, Kavita Kulkarni, S. Sara Mahdavi, Christopher Semturs, Juraj Gottweis, Joelle Barral, Katherine Chou, Greg S. Corrado, Yossi Matias, Alan Karthikesalingam, and Vivek Natarajan. Towards conversational diagnostic artificial intelligence. *Nature*, 642(8067):442–450, June 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-08866-7.

[37] Gail A. Van Norman. Drugs, devices, and the fda: Part 2. *JACC: Basic to Translational Science*, 1(4):277–287, June 2016. ISSN 2452-302X. doi: 10.1016/j.jacbts.2016.03.009.

[38] Georg G. von Polier, Eike Ahlers, Julia Volkening, Jörg Langner, Kaustubh R. Patil, Simon B. Eickhoff, Florian Helmhold, Agnieszka Ewa Krautz, and Daina Langner. Exploring voice as a digital phenotype in adults with adhd. *Scientific Reports*, 15(1):18076, May 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-01989-x.

[39] Ellison B. Weiner, Irene Dankwa-Mullan, William A. Nelson, and Saeed Hassanpour. Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice. *PLOS Digital Health*, 4(4):e0000810, April 2025. ISSN 2767-3170. doi: 10.1371/journal.pdig.0000810.

[40] Zejia A. Yu and Michelle B. Gorgone. *Pay-for-Performance and Value-Based Care*. StatPearls Publishing, Treasure Island (FL), 2025. URL http://www.ncbi.nlm.nih.gov/books/NBK607995/.