Author/s:
Putica, A;Yurtbasi, M;Khanna, R

Title:
Integrating digital health technologies for ecological validity in computational psychiatry: challenges and solutions

Date:
2025-10-01

Citation:
Putica, A., Yurtbasi, M. & Khanna, R. (2025). Integrating digital health technologies for ecological validity in computational psychiatry: challenges and solutions. AI and Society, 40 (7), pp.5509-5525. https://doi.org/10.1007/s00146-025-02336-4.

AI & SOCIETY (2025) 40:5509–5525
https://doi.org/10.1007/s00146-025-02336-4

OPEN FORUM

# Integrating digital health technologies for ecological validity in computational psychiatry: challenges and solutions

**Andrea Putica**[1] · **Miriam Yurtbasi**[1] · **Rahul Khanna**[2]

Received: 25 December 2024 / Accepted: 24 March 2025 / Published online: 8 April 2025

## Abstract

Computational psychiatry offers promising opportunities for understanding and treating mental health disorders, yet achieving ecological validity—the accurate reflection of real-world experiences—remains a critical challenge. This perspective examines how digital health technologies can enhance ecological validity in computational psychiatry while addressing barriers in data collection, participant representation, validation, engagement, and methodological integration. We review key approaches, including digital phenotyping and adaptive design optimization, that enable more naturalistic data collection. However, achieving representative sampling and mitigating algorithmic biases remain unresolved challenges, particularly in AI-driven assessments. We discuss how expert-by-experience collaboration, systematic validation efforts, and structured open science practices can improve model generalizability and clinical applicability. Additionally, we explore the role of federated learning and edge computing in balancing privacy with robust, scalable model development. The paper concludes by integrating these challenges and solutions within a broader methodological framework, emphasizing the need for interdisciplinary approaches that bridge computational precision with real-world psychiatric care.

**Keywords** Computational psychiatry · Artificial intelligence · Ecological validity · Digital health technology · Mental health informatics

## 1 Introduction

Computational psychiatry, an emerging field at the intersection of neuroscience, psychiatry, and computational modeling, offers promising avenues to understanding, diagnosing, and treating mental health disorders (Capon et al. 2016; Nicholas et al. 2020; Roy 2017; Sanfelici et al. 2020; Torous et al. 2021). While originally focused on computational cognitive modeling of specific psychiatric processes (Huys et al. 2016), the field has expanded to encompass a broader range of data-driven approaches and digital methodologies applied to mental health, such as machine learning algorithms, digital phenotyping through passive sensing, generative models, large language models, and advanced statistical frameworks for multi-modal data integration (Friston 2023; Kolding et al. 2024). For example, Wong et al. (2024) developed a machine learning model that predicts future clozapine use in first-episode psychosis patients with moderate accuracy, potentially enabling earlier identification of treatment-resistant cases and more timely interventions, which could significantly improve long-term outcomes for these high-risk patients. By leveraging advanced mathematical models, machine learning algorithms, and big data analytics, computational psychiatry aims to provide more precise and personalized interventions for psychiatric conditions (Adams et al. 2016; Friston 2023; Huys et al. 2016). Despite these promising advances, significant challenges exist in translating computational psychiatry from research to practice.

Ecological validity in computational psychiatry refers to the extent to which computational models, tasks, algorithms, and derived insights accurately reflect and apply to the complexities of real-world mental health experiences and treatments (Benrimoh et al. 2023). This concept is particularly important as it bridges the gap between controlled

✉ Andrea Putica
a.putica@latrobe.edu.au

Miriam Yurtbasi
17871756@students.latrobe.edu.au

Rahul Khanna
rahul.khanna@unimelb.edu.au

[1] La Trobe University, Melbourne, Australia

[2] University of Melbourne, Melbourne, Australia

research environments and the diverse, dynamic real-world contexts in which mental health issues occur and are treated (Andrade 2018; Huang & Parikh 2024). The pursuit of ecological validity has emerged as a cornerstone of translating research into clinical practice (Benrimoh et al. 2023; Hitchcock et al. 2022).

The implications of this ecological validity gap raise fundamental methodological concerns about the translation of computational psychiatry to clinical practice. A meta-analysis by Salazar de Pablo et al. (2021) examined 600 individualized clinical prediction models published in psychiatry and found that only 10% underwent internal validation, with merely 5% subjected to external validation. This striking lack of real-world validation means we cannot yet reliably assess how these models perform across different clinical contexts, populations, and healthcare settings. Without robust validation, we risk developing sophisticated computational approaches that may not generalize beyond the specific research conditions in which they were developed.

The issue extends to fundamental data collection tools and their underlying assumptions about mental health measurement. Digital phenotyping methods face important reliability challenges that must be carefully addressed. Adler et al. (2024) demonstrated that seemingly objective behavioral markers like mobility patterns can indicate different mental health states across demographic and socioeconomic groups—higher mobility was associated with better mental health for employed individuals but worse mental health for those on disability. Similarly, Birk and Samuel (2022) highlight how digital phenotyping's emphasis on behavioral measurements may not fully capture the complex lived experiences of mental health across diverse populations. These challenges are not insurmountable barriers but rather critical considerations that must inform how we develop and implement these tools. As the following sections will demonstrate, addressing these reliability challenges through careful methodological design and implementation strategies is essential for bridging the gap between computational precision and clinical applicability.

Achieving ecological validity in computational psychiatry presents several significant challenges. First, the complexity and heterogeneity of mental health disorders make it difficult to create comprehensive models that account for all relevant factors. This is exemplified by the tendency of most approaches in computational psychiatry to focus on biological properties while failing to take psychosocial factors into consideration (Karvelis et al. 2022; Uusitalo et al. 2021). Second, the controlled environments in which many computational models are developed may only partially capture the nuances of real-world clinical settings. This limited external validation raises fundamental questions about the generalizability and practical utility of these models across diverse clinical landscapes (Fusar-Poli et al. 2022; Salazar de Pablo et al. 2021). Third, the rapid pace of technological advancement in this field can sometimes take priority over validating findings in diverse populations and contexts. This is evident in the concerns about the'third wave of biological psychiatry'(Karvelis et al. 2022), which risks ignoring sociocultural factors and changes in conscious experiences associated with mental disorders (McCradden et al. 2023), potentially leading to suboptimal treatments or optimal treatments for a select few. Additionally, the data analyzed in psychiatric clinical prediction models are typically complex derivatives of the original raw data, such as connectivity matrices from functional (fMRI) resonance imagining signal fluctuations, not enriched information on disease mechanisms and their causal links to targeted outcomes (Dima et al. 2022; Frangou et al. 2022). Despite these challenges, integrating computational approaches with digital health innovation can enhance ecological validity, leading to more accurate diagnostics, targeted treatments, and deeper insights into the bio-psycho-environmental factors affecting mental health.

This perspective article addresses a critical challenge in computational psychiatry: strengthening the ecological validity of research findings. We first review data collection approaches, examining adaptive design optimization and digital phenotyping methods. The paper then discusses participant-related factors, including recruitment strategies and the role of experts-by-experience in model development. Next, we explore issues of model validation and generalizability, highlighting challenges in external validation and algorithmic bias. We then examine how open science practices and privacy-preserving technologies may support clinical translation. Finally, we discuss digital health implementation considerations and how these methodological frameworks contribute to the development of computational approaches that meaningfully reflect real-world clinical contexts.

Table 1 summarizes the key challenges, current solutions, and future directions discussed throughout this perspective regarding ecological validity in computational psychiatry. The table organizes evidence across five areas central to this discussion: digital phenotyping and data collection, population representation, clinical validation, participant engagement, and methodological integration. The following sections examine these areas in detail, beginning with digital phenotyping methods that enable naturalistic data collection, then addressing population representation challenges, clinical validation approaches, strategies for maintaining participant engagement, and finally methodological frameworks that integrate these various elements. Throughout, we provide empirical evidence and practical considerations for implementing digital health technologies in computational psychiatry research and practice.

**Table 1** Ecological validity in computational psychiatry: challenges, evidence-based solutions, and future directions

| Domain | Documented challenges | Evidence-based solutions | Future directions |
|---|---|---|---|
| Digital Phenotyping and data collection | Battery limitations: Continuous GPS tracking depletes smartphone batteries within 10 h, limiting long-term monitoring capabilities in natural settings (Pramanik et al. 2019) | Power-efficient monitoring: Adaptive GPS sampling that increases frequency only when movement is detected and decreases during sedentary periods, preserving battery life while maintaining data quality. This approach balances continuous behavioral tracking with device limitations (Cholaquidis et al. 2023) | Edge computing and federated learning for balancing privacy with robust model development |
| | Lack of standardized data formats: Inconsistent sensor data representation (sampling frequencies, file formats) and collection protocols across studies creates heterogeneity that limits cross-study comparison, validation, and meta-analytic integration of findings. This variability prevents effective pooling of data across studies and hinders reproducibility (Mendes et al. 2022) | Multi-sensor integration: Combining positioning, movement (accelerometer), virtual (screen usage), and ambient sensing (light, sound) data to create comprehensive behavioral profiles. Specific implementations use hierarchical sensor fusion (e.g., accelerometer data with Wi-Fi scanning to identify familiar locations, combined with ambient light/sound measurements) to improve contextual understanding of behavior | Integration of passive sensing technologies with ecological momentary assessments |
| | Traditional clinical assessments fail to capture dynamic behavioral patterns that occur in real-world settings, limiting ecological validity | Integrating location data with communication patterns (calls/texts) to distinguish between social withdrawal and mobility limitations (Mendes et al. 2022) | Development of multi-modal behavioral signatures that combine passive sensing with communication patterns to create clinically meaningful digital biomarkers of psychiatric states |
| | Integration with clinical frameworks: Need for digital phenotyping approaches to align with established clinical assessment paradigms | Clinical framework alignment: Longitudinal integration of passively collected data with established frameworks like National Institute of Mental Health Research Domain Criteria and Measurement-Based Care, enabling continuous feedback loops between research and practice | Development of standardized digital phenotyping protocols that directly map to clinical assessment domains and treatment decision points |
| Population representation | Algorithmic bias: Speech recognition systems show nearly double the error rate (35%) for Black speakers compared to white speakers (19%), potentially skewing assessment data and leading to systematic errors in symptom detection and classification (Bajorek 2019) | Cultural inclusivity: Implementing community-engaged research approaches with cultural adaptations to ensure tools are appropriate across diverse populations (Constant et al. 2022) | Systematic integration of patient-reported outcomes |
| | Diagnostic disparities: AI systems trained on majority populations systematically misclassify clinical markers in minority groups, leading to potentially inaccurate assessments (Davidson et al. 2019). For example, Black patients are disproportionately diagnosed with schizophrenia-related disorders while mood disorders are under-recognized in the same population (Olbert et al. 2018) | Representative sampling: Combining targeted outreach to underrepresented communities with traditional recruitment methods to ensure computational models are trained on diverse datasets | Enhanced frameworks for evaluating model performance across diverse populations |

**Table 1** (continued)

| Domain | Documented challenges | Evidence-based solutions | Future directions |
|---|---|---|---|
| | Assessment bias: AI-driven symptom assessments exhibit higher false-positive rates for psychosis in racial minorities, reinforcing disparities in access to appropriate treatments (Davidson et al. 2019) | Fairness-aware modeling: Implementing machine learning approaches that specifically detect and mitigate systematic biases across demographic groups, reducing misclassification rates in underrepresented populations | Development of standardized tools that automatically detect performance disparities across demographic groups and implement real-time adjustments to ensure equitable algorithmic performance, moving from reactive corrections to proactive fairness assurance in clinical applications |
| Clinical validation | Validation deficit: Of 600 clinical prediction models in psychiatry, only 10% underwent basic internal validation testing for reliability | Coordinated research frameworks: Implementation of standardized data collection and analysis protocols across multiple sites, as demonstrated by the ENIGMA neuroimaging consortium spanning 45 countries (Thomson et al., 2020), which has successfully established standardized processing pipelines for psychiatric neuroimaging | Development of standardized data sharing frameworks |
| | Generalizability gap: Only 5% of psychiatric prediction models were tested with external validation to confirm performance across different clinical settings (Salazar de Pablo et al. 2021) | Data quality standards: Adoption of FAIR principles (Findable, Accessible, Interoperable, Reusable) to ensure computational models can be independently validated and reproduced | Systematic validation approaches integrated with clinical expertise and supported by external validation incentives |
| | Distanced from source data: Most models analyze complex data derivatives (e.g., connectivity matrices) rather than raw measurements, potentially obscuring causal mechanisms (Dima et al. 2022) | Methodological transparency: Pre-registration of analysis protocols before data collection to reduce selective reporting and strengthen reproducibility (Poldrack et al. 2020) | Integration of raw data alongside processed derivatives to improve interpretation and causal inference |
| | Framework limitations: While standardized frameworks exist for neuroimaging (e.g., ENIGMA), equivalent frameworks for behavioral data, ecological momentary assessments, and digital phenotyping remain underdeveloped, limiting cross-study validation | Emerging collaborative initiatives: Initial development of multi-site data-sharing protocols for behavioral data integration, though less established than neuroimaging frameworks | Development of standardized validation frameworks for behavioral and ecological momentary assessment data |
| | Implementation barriers: Translation of computational models from research to clinical settings hindered by lack of standardized data formatting protocols and secure sharing mechanisms compatible with clinical workflows | Secure implementation frameworks: Development of privacy-preserving data sharing architectures and standardized processing pipelines suitable for clinical environments | Integrated clinical-research dataflows: Creation of seamless data exchange systems between research and clinical environments with built-in privacy safeguards and standardized metadata |
| | Incentive misalignment: Limited incentives for researchers to conduct external validation studies or implement open science practices, resulting in under-validated computational models | Early adoption leaders: Pioneering journals and funding bodies implementing data sharing mandates and reproducibility requirements for computational research, demonstrating feasibility of incentive restructuring | Policy alignment: Journal and funding agency requirements for external validation and open science compliance to encourage more rigorous model assessment |
| Participant engagement | Participation consistency: Sustained participant engagement required for collecting reliable longitudinal data, particularly challenging in psychiatric populations | Optimized interaction timing: Strategic delivery of notifications at 12:30 PM increased user engagement by 8.8%, with higher effects on weekends (Bidargaddi et al. 2018) | Context-aware systems that can dynamically adjust computational task delivery |

**Table 1** (continued)

| Domain | Documented challenges | Evidence-based solutions | Future directions |
|---|---|---|---|
| | Engagement barriers: User concerns about privacy, data security, and perceived burden significantly reduce willingness to participate in digital monitoring (Simblett et al. 2019) | User-centered design: Collaborating with service users and experts-by-experience to develop engagement strategies that address their concerns and preferences, ensuring digital tools align with real-world lived experiences and user needs. This co-design approach improves sustained participation and enhances the ecological validity of computational models | Integration of expert-by-experience insights (Vuqetrna et al. 2023) |
| | Assessment adherence: Inconsistent completion of clinical assessments compromises data quality and limits model effectiveness in real-world settings | Autonomy-supporting interfaces: Providing users with control over when and how they interact with digital health tools, reducing perceived burden while maintaining data quality | Personalized adherence systems that dynamically balance user autonomy with data quality requirements, using machine learning to identify optimal assessment timing and format for each individual while respecting user preference |
| Methodological integration | Methodological inconsistency: Lack of standardized data collection and processing procedures prevents effective comparison and integration of findings across studies | Parameter transparency: Implementation of standardized reporting frameworks for computational model parameters to enable replication and comparison across studies | Establishment of reference frameworks |
| | Resource limitations: Significant imbalance between available applications (31) and publicly accessible datasets (8), hindering independent validation and replication (Mendes et al. 2022) | Processing pipeline clarity: Detailed documentation of all preprocessing steps to ensure computational reproducibility and enable accurate replication (Poldrack et al. 2020) | Standardized approaches for evaluating computational models |
| | Real-time adaptation limitations: Traditional fixed-design methodologies lack flexibility to adjust research parameters during studies, compromising data quality and participant experience (Kwon et al. 2023) | Adaptive Design Optimization: Implementation of real-time data analysis throughout trials, enabling dynamic adjustment of study parameters while maintaining methodological rigor (Kwon et al. 2023) | Context-sensitive computational models that dynamically adjust based on real-time behavioral and physiological inputs, enabling personalized assessment and intervention |
| | Processing pipeline complexity: Inconsistent documentation of computational methods creates barriers to reproduction and clinical implementation | Technical implementations including standardized metadata schemas for sensor specifications and contextual information, automated anonymization pipelines, and Docker containers for ensuring computational reproducibility (Wilson et al. 2017; Boettiger 2015) | Standardized, modular processing pipelines with versioned components that can be audited, validated, and certified for clinical implementation across diverse healthcare settings |

*ENIGMA* Enhancing Neuro Imaging Genetics through Meta-Analysis, *FAIR* Findable, Accessible, Interoperable, Reusable

## 2 Digital phenotyping and data collection

The advancement of computational psychiatry requires methodological approaches that can effectively capture the complexity of mental health phenomena in real-world settings. Digital phenotyping has emerged as a promising strategy for capturing naturalistic behavioral patterns through passive data collection. By leveraging real-time behavioral tracking, this approach enhances ecological validity in computational psychiatry, offering empirically validated improvements in measurement precision and temporal resolution. To fully capture the complexity of mental health in real-world settings, digital phenotyping requires a multi-tiered data collection strategy that integrates both passive and active monitoring. Passive data collection through smartphone and wearable sensors enables continuous tracking of behavioral and physiological patterns, while active methods such as Ecological Momentary Assessments (EMA) provide contextual self-reports, complementing sensor-derived data and improving interpretability. This section explores key methodologies in digital data collection, including sensor-based tracking, adaptive GPS sampling, and intelligent trigger mechanisms that optimize data quality while reducing participant burden.

### 2.1 Digital phenotyping

Digital phenotyping through passive data collection—including smartphone sensors, wearable devices, and ambient monitoring systems—represents an emerging approach for capturing real-world behavior in psychiatric research and clinical practice (Oudin et al. 2023), offering significant potential for enhancing ecological validity. These methods directly address key limitations of more established retrospective assessments, which can be subject to recall bias and often fail to capture the temporal dynamics of psychiatric symptoms (Pacheco-Romero et al. 2024). By enabling real-time data collection in participants'natural environments, these approaches provide more ecologically valid data on patients'experiences, behaviors, and symptoms as they occur in real-world settings (Torous et al. 2021).

Contemporary evidence from systematic reviews and empirical investigations demonstrates the utility and feasibility of this approach. A systematic review of 35 studies (Cornet and Holden 2018) established the efficacy of passive sensing methodologies for symptom monitoring across major psychiatric conditions. These investigations validated the capacity of smartphone sensors to quantify behavioral endophenotypes through objective measurements of mobility, social interaction patterns, and sleep–wake cycles. Benrimoh et al. (2023) demonstrated how passive sensing can capture dynamic relationships between daily routines and depressive symptoms, showing that smartphone-based measurements provided more accurate representations of symptom fluctuations compared to traditional weekly assessments.

Expanding on these findings, Mendes et al. (2022) conducted a systematic review of 31 sensing applications and 8 public datasets, revealing important patterns in digital phenotyping approaches. Their review identified three predominant types of analyses across studies: classification ($n = 11$), prediction ($n = 6$), and correlation analysis ($n = 6$) of mental states/disorders, with some applications ($n = 9$) focused solely on raw data collection. The most frequently combined sensor data streams included GPS coordinates with call logs ($n = 20$), accelerometer data ($n = 17$), and screen activity ($n = 10$), indicating researchers'interest in capturing multiple behavioral dimensions. By integrating GPS data with call and text metadata, researchers can refine analyses of movement-based social behaviors, enabling differentiation between voluntary social withdrawal and mobility impairments due to psychiatric symptoms (Mendes et al. 2022). This approach has been particularly valuable in psychiatric research, where mobility tracking alone may not distinguish between individuals actively avoiding social interactions versus those experiencing physical limitations due to depressive or anxiety-related symptoms. By analyzing movement trajectories alongside communication patterns, researchers can better identify behavioral phenotypes indicative of clinical risk factors, such as isolation trends that precede depressive episodes. Notably, the integration of positioning, inertial, virtual and ambient sensor data enabled more comprehensive behavioral pattern detection compared to single-sensor approaches. However, the authors highlighted significant methodological challenges, including the lack of standardization in data representation (e.g., sampling frequencies, file formats) and collection protocols across studies. This heterogeneity limited opportunities for cross-study validation and meta-analytic integration of findings. The review also revealed a scarcity of publicly available datasets ($n = 8$) relative to sensing applications ($n = 31$), with most datasets having limited sample sizes ranging from 1 to 482 participants.

Despite these challenges, the longitudinal integration of passively collected real world outcomes data aligns with established clinical frameworks like the NIMH Research Domain Criteria (RDoC) and Measurement-Based Care (MBC) approaches in psychiatry. These frameworks emphasize dimensional assessment and continuous outcome monitoring, respectively, with passive sensing offering new opportunities for systematic data collection. By establishing continuous feedback loops between clinical practice and research, these data streams enable iterative refinement of computational models based on actual patient outcomes (Chambers 2020; Mohr et al. 2017). This dynamic approach helps ensure that models remain clinically relevant and adaptable to evolving patient needs and treatment patterns. While methodological challenges remain, including the need for rigorous validation studies and standardized collection protocols (Büscher et al. 2024), the field shows promise for establishing more ecologically valid approaches in psychiatric research and care. One of the key obstacles in digital phenotyping is the lack of standardization in data representation across studies. Variability in sensor data formats, sampling rates, and preprocessing pipelines limits cross-study validation and meta-analytic integration (Mendes et al. 2022). Without standard protocols, findings remain difficult to generalize across different populations and settings, reducing the potential clinical utility of digital phenotyping models.

## 2.2 Methods of data collection in digital phenotyping

The effectiveness of digital phenotyping hinges on optimizing data collection strategies that balance continuous passive sensing with targeted, context-sensitive active self-reports. These practical recommendations focus on deploying digital health technologies that capture real-world psychiatric phenomena while maintaining clinical utility and research rigor. Successful implementation of ecologically valid digital health tools in computational psychiatry requires a multi-tiered data collection architecture. The foundation rests on continuous passive monitoring through smartphone and wearable sensors (Insel 2017). Smartphone accelerometer data can capture physical activity patterns, screen time monitoring can indicate engagement and attention patterns, and call/text metadata can quantify social interaction frequency (Mohr et al. 2017). However, implementation requires careful consideration of both technical and practical constraints. For example, Recent analyses indicate that prolonged GPS use can rapidly deplete battery reserves, necessitating adaptive power management strategies to ensure feasibility in long-term digital phenotyping applications (Pramanik et al. 2019). A more ecologically valid approach involves adaptive GPS sampling, with frequency increasing when movement is detected and decreasing during sedentary periods (Cholaquidis et al. 2023).

Active data collection methods complement passive monitoring through carefully implemented EMAs (Wrzus & Neubauer 2023). Traditional EMAs often use fixed sampling schedules that can artificially structure participant behavior (Wrzus and Neubauer 2023). Modern digital health implementations can instead use"intelligent triggers"based on detected behavioral patterns (Giurgiu et al. 2020). For instance, Giurgiu et al. (2020) demonstrated an effective implementation of adaptive sampling in sedentary behavior monitoring. Their system used thigh-worn Move 3 accelerometers that communicated with smartphones via Bluetooth Low Energy to detect prolonged sedentary periods. Across three studies, they tested different sedentary duration thresholds (30 min in Study 1, 20 min in Studies 2 and 3) that would trigger contextual assessment prompts. The system achieved an overall accuracy of 82.77% in capturing sedentary bouts while managing participant burden through timeout periods (40 min in Study 1, 20 min in Studies 2 and 3). Compared to simulated random sampling every 120 min, their adaptive sampling approach captured substantially more prolonged sedentary episodes, with improvements ranging from 20.83% to 47.88% across studies. This method enabled researchers to collect richer contextual data about participants'sedentary behaviors, revealing that nearly 40% of prolonged sedentary bouts occurred during work, and in 57% of all bouts, participants were not alone.

Context-aware data collection enhances the precision of digital phenotyping by integrating multiple sensor streams. For example, ambient light and accelerometer data can differentiate between indoor and outdoor settings, while Bluetooth/Wi-Fi signals provide information about social proximity and location stability (Ghahramani et al. 2020). Combining these data sources with machine learning enables more accurate classification of behavioral contexts, refining the ecological validity of digital health assessments. However, raw sensor data alone often provides ambiguous context information. Successful implementations combine multiple data streams with machine-learning approaches for context classification (Toch et al. 2019). Recent implementations have demonstrated how hierarchical sensor fusion can improve real-world behavioral tracking by integrating multiple sensor streams. For example, combining accelerometer data to detect movement patterns, Wi-Fi scanning to identify familiar locations, and ambient light and sound measurements provide a more nuanced understanding of environmental context. For instance, a typical office context might be identified through the stable presence of workplace Wi-Fi networks, sustained periods of low movement interspersed with walking patterns, and consistent indoor light levels during working hours. This multi-modal approach significantly improves accuracy over single-sensor methods, as it can distinguish between superficially similar contexts (like differentiating between sitting at a desk at home versus at work) by considering the unique combination of environmental signatures. The system can then adapt its behavior accordingly, such as adjusting notification settings or monitoring stress levels in different contexts (Dautov et al. 2019).

As digital phenotyping scales in clinical applications, ensuring data privacy and computational efficiency becomes a critical challenge—especially given the increasing complexity of multi-sensor data integration. Future developments in this field aim to address these concerns through emerging approaches like edge computing and federated learning (Upreti et al. 2024). Edge computing enables real-time processing of sensor data directly on user devices, reducing reliance on centralized cloud storage and minimizing privacy risks. Federated learning further enhances data security by allowing machine learning models to be trained across multiple decentralized devices while keeping individual data local, mitigating risks associated with data centralization. By decentralizing data processing and minimizing privacy risks, these innovations enable the scalable deployment of digital phenotyping tools in real-world psychiatric settings. Federated learning and edge computing hold particular promise for developing personalized, adaptive interventions while ensuring data security in clinical applications (Xia et al. 2021; Li et al. 2020). While these technological solutions address important data collection and processing challenges, the ultimate validity of computational models depends critically on who is represented in the underlying datasets. Even the most sophisticated digital phenotyping approaches will yield limited clinical utility if they fail to include diverse populations or account for systematic differences in how psychiatric symptoms manifest across different demographic groups. The next section examines how participant-related factors, particularly issues of representation and inclusion, impact the development and implementation of computational psychiatry approaches.

## 3 Population representation

The successful development and implementation of computational psychiatry approaches critically depend on careful consideration of participant-related factors throughout the research process. As computational models become increasingly sophisticated, ensuring appropriate participant selection, engagement, and representation has emerged as a fundamental challenge that can significantly impact model validity and clinical utility (Salazar de Pablo et al. 2021). Key areas requiring particular attention include strategies for representative recruitment across diverse populations (Leslie et al. 2021) and meaningful integration of lived experience into model development (Abbasgholizadeh Rahimi et al. 2022). Understanding and effectively addressing these participant-related challenges is essential for developing computational models that can meaningfully advance psychiatric care while ensuring equitable benefits across different patient populations (Lee et al. 2021; Torous et al. 2021).

### 3.1 Diverse recruitment strategies

The development of robust computational models in psychiatry requires carefully considered strategies for acquiring high-quality training and validation data that truly represent target clinical populations (Gannedahl et al. 2024). While digital recruitment through social media platforms and online communities has emerged as a powerful tool for assembling large-scale datasets needed for computational modeling (Sanchez et al. 2020), the specific requirements of computational psychiatry present unique recruitment challenges that go beyond traditional clinical trial considerations.

Recent work in Artificial Intelligence models for mental health assessment has elucidated the importance of representative recruitment in computational psychiatry. When algorithms were built using data predominantly from white, high-income populations, they showed significantly reduced accuracy when applied to racial and ethnic minorities, even for basic tasks like speech recognition and language processing (Buolamwini & Gebru 2018; Koenecke et al. 2020; Levkovich et al. 2024). For instance, automated speech-recognition systems, which many computational psychiatry applications rely on for data collection, show a 35% word-error rate for Black and African American speakers compared to 19% for non-Hispanic white speakers (Bajorek 2019). Natural language processing models trained on majority populations have demonstrated systematic misclassification and systematic over-identification of clinical markers for minority groups, leading to biased assessments (Davidson et al. 2019; Sap et al. 2019).

Beyond speech recognition errors, systematic biases in diagnostic algorithms have contributed to racial disparities in psychiatric classification. For example, studies have shown that Black patients are disproportionately diagnosed with schizophrenia-related disorders, while mood disorders are under-recognized in the same population, partly due to algorithmic reliance on linguistic and behavioral patterns derived from predominantly white cohorts (Olbert et al. 2018). Similarly, AI-driven symptom assessments have exhibited higher false-positive rates for psychosis in racial minorities, reinforcing disparities in access to appropriate treatments (Davidson et al. 2019). Addressing these biases requires both diversifying training datasets and incorporating fairness-aware machine learning approaches to reduce systematic misclassification.

The field has responded to these challenges through specialized recruitment frameworks in computational psychiatry. Community-engaged approaches incorporating cultural adaptation have shown particular promise in reaching underrepresented groups (Constant et al. 2022). The integration of targeted outreach strategies with traditional recruitment methods helps ensure broader population coverage while maintaining data quality for model development (Benrimoh et al. 2023). Beyond ensuring diversity in data collection, computational psychiatry models must also undergo rigorous validation to ensure their predictive accuracy across real-world patient populations.

However, validation efforts in computational psychiatry often fail to systematically assess model performance across demographic subgroups, leading to hidden biases in clinical decision-making. Without structured fairness evaluations, machine learning models may yield misleading predictions when applied to historically underrepresented populations (Seyyed-Kalantari et al. 2021). Recent approaches emphasize the need for subgroup-specific performance metrics and bias audits to ensure model generalizability across diverse patient populations (Sahin et al. 2024). These population representation challenges directly impact the clinical validity of computational models. A model can only be considered valid to the extent that it has demonstrated reliable performance across the diverse populations it aims to serve. The following section explores how open science practices and rigorous validation approaches can help ensure that computational psychiatry models not only include diverse populations but also undergo appropriate testing to confirm their clinical utility across different demographic and clinical groups.

## 4 Clinical validation

The translation of computational psychiatry from research to clinical practice critically depends on robust validation and reproducibility of findings, making open science practices increasingly essential. Elaborating on previously identified validation gaps in computational psychiatry—particularly, the lack of standardized neurobiological frameworks and insufficient normative data for benchmarking models—recent initiatives have emerged to address these fundamental challenges in model robustness assessment. These efforts specifically target the establishment of reference frameworks and standardized approaches for evaluating computational models in psychiatry.

The ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis) consortium demonstrates both the potential and current limitations of coordinated open science approaches in computational psychiatry. While this international collaboration spanning 45 countries has successfully implemented standardized protocols and mega-analyses in psychiatric neuroimaging (Thompson et al. 2020), its focus remains primarily on neuroimaging data. Through ENIGMA's efforts, researchers have established standardized processing pipelines and quality control metrics that have become field standards for neuroimaging-based studies. However, computational psychiatry requires integration of diverse data types beyond brain imaging, including behavioral, environmental, and longitudinal clinical measures. While ENIGMA's creation of the largest standardized neuroimaging database in psychiatry represents an important step forward, its scope highlights the need for more comprehensive open science frameworks that can accommodate the full range of data types and analytical approaches needed for computational psychiatry's advancement.

Despite the success of ENIGMA in standardizing neuroimaging workflows, the field lacks equivalent frameworks for behavioral and ecological momentary assessment data. Existing standardization efforts primarily focus on neuroimaging modalities, yet multi-site data-sharing protocols and validation pipelines for real-world behavioral data remain underdeveloped. Ensuring the generalizability of computational psychiatry models requires common validation pipelines that encompass passive sensor data, ecological momentary assessments, and multi-modal behavioral tracking. Establishing such frameworks is essential for improving model reliability and applicability in real-world psychiatric research. Ongoing open science efforts, such as large-scale collaborations integrating multi-modal data (e.g., behavioral tracking, digital phenotyping, and passive sensing), are necessary to establish comprehensive validation pipelines beyond neuroimaging. These initiatives, coupled with standardized data-sharing mechanisms, will be essential for ensuring model reproducibility and reliability across diverse psychiatric populations.

For computational psychiatry to achieve clinical impact, its underlying methodologies must be transparent and reproducible. This requires systematic approaches to data sharing and analysis that go beyond traditional research practices. The FAIR principles (Findable, Accessible, Interoperable, Reusable) established by Wilkinson et al. (2016) provide a foundational framework for scientific data management in the digital age. In response to these challenges, Poldrack et al. (2020) propose more targeted requirements, including mandatory pre-registration of analysis protocols, standardized reporting of model parameters, and comprehensive documentation of preprocessing steps. These guidelines are particularly important given the reproducibility challenges identified in systematic reviews of computational psychiatry research (Salazar de Pablo et al. 2021), where lack of methodological transparency has hindered clinical translation.

The implementation of open science practices in clinical settings presents several key challenges (Bertl et al. 2022). These include establishing standardized data formatting protocols, developing secure data-sharing mechanisms, and creating clear guidelines for clinical implementation. The successful translation of computational models into clinical practice requires addressing these infrastructure challenges while maintaining rigorous validation standards. Additionally, there is a critical need to establish clear frameworks for evaluating model performance across diverse clinical populations, particularly when implementing computational approaches in real-world settings.

The success of the ENIGMA consortium demonstrates how coordinated open science initiatives can transform research practices in computational psychiatry (Thompson et al. 2020). However, bridging the gap between research capabilities and clinical implementation remains a critical challenge. Despite the importance of external validation and adherence to FAIR principles, there is a lack of strong incentives for researchers to engage with these essential initiatives. Many computational models in psychiatry remain under-validated due to the absence of standardized publishing requirements or funding mandates prioritizing external validation efforts (Monteith et al. 2023). One potential solution would be for journals and funding agencies to require external validation and open science compliance as conditions for publication and grant eligibility. Such policies would encourage more rigorous model assessment, promote transparency, and ultimately enhance the clinical utility of computational psychiatry findings. While standardized protocols and collaborative frameworks provide a foundation, the field must now focus on developing rigorous validation pipelines that can support reliable clinical deployment (Fusar-Poli et al. 2022). While these validation frameworks and open science practices provide essential methodological foundations, their successful implementation ultimately depends on sustained participant engagement. Even the most rigorous validation protocols and data-sharing frameworks will fail to yield clinically useful insights if researchers cannot maintain consistent, high-quality data collection from diverse participants over time. The challenges of recruitment and retention in computational psychiatry studies are particularly pronounced given the intensive nature of data collection and the vulnerability of many target populations. The following section examines how participant engagement strategies can be optimized to support both robust validation efforts and ecological validity in computational psychiatry research.

## 5 Participant engagement

Sustained participant engagement in digital health implementations is crucial not only for ensuring high-quality data collection but also for improving the ecological validity and generalizability of computational psychiatry models. Without sustained engagement, model predictions risk being based on incomplete, biased, or unrepresentative datasets, limiting their clinical utility. The challenges of engagement and compliance have been well documented in electronic diary studies (Wen et al. 2017). Traditional approaches relying on generic reminders can become intrusive and alter natural behavior patterns, affecting both data quality and ecological validity. Personal sensing approaches in mental health require careful consideration of both passive data collection and active user participation strategies (Mohr et al. 2017). More sophisticated implementations are exploring personalized engagement approaches based on detected usage patterns. In a micro-randomized trial, Bidargaddi et al. (2018) examined how the timing of push notifications affected engagement with a mobile health app. Their study found that sending prompts at 12:30 PM resulted in the highest engagement overall, with an 8.8% increase in the likelihood of app use in the following 24 h. Effects were particularly strong on weekends at both 12:30 PM and 7:30 PM, with up to 11.8% increased engagement. The findings suggest that mid-day notifications, particularly during lunch hours, may be most effective for promoting app engagement. Beyond optimizing timing, emerging digital health systems are integrating adaptive engagement strategies that respond dynamically to individual behavior patterns. For instance, context-aware mobile interventions can modify the frequency and content of engagement prompts based on real-time user activity, such as adjusting reminders based on detected stress levels or modifying assessment intervals depending on prior response consistency (Mohr et al. 2017). Such context-sensitive approaches enhance both compliance and data integrity, reducing participant burden while ensuring robust longitudinal data collection. While participant engagement strategies focus on sustaining long-term user interaction with digital health tools, collaboration with experts-by-experience extends beyond engagement to actively shape computational psychiatry methodologies. This approach ensures that psychiatric models are not only user-friendly but also accurately reflect symptom dynamics in real-world contexts.

Critically, the development of engagement strategies benefits from incorporating lived experience perspectives from the outset (Zidaru et al. 2021). Simblett et al. (2019) found through qualitative interviews with service users that concerns about privacy, data usage, and perceived burden significantly influenced willingness to engage with digital health tools. Service users emphasized the importance of having control over when and how they interact with technology-based assessments. This aligns with findings from Bucci et al. (2019), who demonstrated that co-designing engagement approaches with service users led to higher sustained participation rates—participants reported feeling more invested in the process when they understood how their data contributed to their health outcomes. Ensuring engagement in digital health research requires not only well-designed interventions but also adaptive methodologies that respond dynamically to participant behaviors. Integrating expert-by-experience perspectives further enhances engagement strategies by ensuring that computational tools align with the lived realities of those they aim to support—necessary for computational psychiatry's advancement.

### 5.1 Collaboration with experts-by-experience

The development of clinically valid computational models requires careful consideration of how psychiatric symptoms manifest in real-world contexts, with increasing recognition that computational psychiatry must account for both temporal dynamics and contextual factors (Hitchcock et al. 2022; Treadway 2023). While computational approaches offer powerful analytical tools, their clinical utility ultimately depends on how well they capture individual experiences of mental illness. Recent evidence suggests that standardized computational tasks often fail to adequately reflect the complex, dynamic nature of psychiatric symptoms as they occur in daily life, particularly when models prioritize algorithmic efficiency over ecological validity (Brown et al. 2020).

Patient-Reported Outcome Measures (PROMs) have emerged as one effective formalization of this collaboration, systematically capturing patients'views on symptoms, treatment effects, and overall health status (Øvretveit et al. 2017). These standardized measures help ensure that computational models incorporate outcomes that are meaningful to patients'lived experiences while maintaining scientific rigor. This approach represents a shift from purely nomothetic research designs toward methods that better account for individual variation in symptom expression and experience (Brown et al. 2020; Treadway 2023).

The implementation of expert-by-experience collaboration requires systematic evaluation of computational paradigms against lived experience, including identification of person-specific symptom dynamics and documentation of contextual factors that trigger or modify symptoms (Hitchcock et al. 2022). The practical impact of this approach was demonstrated in a groundbreaking pilot study by Vuqetrna et al. (2023), where people with Functional Neurological Disorder actively participated in model development. Their collaboration improved both task design and model validity through the development of computational tasks specifically tailored to reflect lived experiences. This integration of lived experience has established a crucial methodological framework in computational psychiatry, advancing the field's translational potential (Treadway 2023) while ensuring models meaningfully capture the complexity of psychiatric presentations in real-world clinical contexts (Hitchcock et al. 2022). Sustaining engagement in digital health research requires not only well-designed interventions but also adaptive methodologies that integrate participant feedback in real time. While effective participant engagement provides the foundation for quality data collection, these engagement strategies must be supported by robust methodological frameworks that can systematically optimize study designs and adapt to participant needs. Adaptive methodologies, such as ADO (Adaptive Design Optimization), further strengthen the integration of expert-by-experience insights by enabling real-time adjustments to computational models based on ongoing participant responses and behaviors (Kwon et al. 2023). The following section examines these methodological frameworks in detail, exploring how innovations in study design, computational approaches, and open science practices create the technical infrastructure necessary to support both participant engagement and ecological validity in computational psychiatry.

## 6 Methodological integration

Computational psychiatry depends on robust methodological frameworks that optimize data collection, maintain scientific rigor, and enhance real-world applicability. Previous sections have highlighted challenges in ensuring participant engagement, clinical validation, and population representativeness—all of which require methodological advancements to address effectively. This section explores ADO as a strategy for refining study protocols, computational frameworks that support real-time decision-making, and open science principles that promote transparency and reproducibility. Together, these innovations ensure that digital health methodologies remain both scalable and scientifically rigorous.

### 6.1 Adaptive design optimization

Adaptive design optimization (ADO) enhances ecological validity in computational psychiatry research and assessments by offering a more flexible and responsive framework that better mirrors real-world conditions (Kwon et al. 2023). Unlike traditional fixed-design methodologies, ADO incorporates real-time data analysis throughout the trial process, allowing researchers to dynamically adjust study parameters based on interim findings while maintaining trial integrity. For example, cognitive task parameters can be modified based on participant performance, ensuring assessments remain optimally informative across diverse patient populations.

ADO's value in computational psychiatry lies in its ability to maximize information gain while minimizing participant burden. Research demonstrates this efficiency through delay discounting assessments—measures of how people value immediate versus future rewards that are particularly relevant in addiction research. Studies show that ADO approaches can be 3–5 times more precise and 3–8 times more efficient than standard fixed-sequence methods while maintaining excellent reliability even in clinical populations where attention difficulties or symptom severity might otherwise compromise assessment quality (Ahn et al. 2020). Rather than using fixed sequences of questions, ADO can quickly identify key decision points (such as when someone switches preference from immediate to delayed rewards) and focus subsequent questions around these points, enabling more precise measurement of decision-making patterns with fewer trials.

Furthermore, ADO enables researchers to refine intervention protocols in response to emerging data and participant feedback. This dynamic adaptation extends beyond individual assessments to address broader methodological challenges in computational psychiatry. For example, ADO can help detect and counter algorithmic biases in computational models by continuously adjusting trial parameters in response to subgroup-specific response patterns. By dynamically analyzing real-time participant data, ADO can identify systematic differences in task performance, treatment response, or model predictions across demographic groups, allowing researchers to implement corrective adjustments that improve model fairness and generalizability. When such biases are detected, researchers can adjust trial parameters to explore alternative strategies or parameter settings that may better serve these populations (Thorlund et al. 2018). This responsiveness not only enhances the ecological validity of the study but also increases the likelihood that the interventions being tested will be both feasible and effective in real-world settings.

### 6.2 Bridging implementation with methodological foundations

Digital health platforms must not only collect ecologically valid data but also support the computational modeling approaches central to the field (Gillan & Rutledge 2021). The integration of ADOs with continuous passive monitoring requires sophisticated technical architectures that can dynamically adjust computational task delivery based on real-world behavioral patterns (Pooseh et al. 2018). Similarly, the incorporation of expert-by-experience insights discussed earlier demands flexible platforms that can adapt to diverse symptom presentations while maintaining computational rigor (D'Alfonso 2020).

### 6.3 Open science and reproducibility in computational psychiatry

The integration of open science practices into digital health implementations requires specific technical infrastructure and standardized protocols. Successful implementation of open science principles in computational psychiatry's digital health tools involves both data-sharing architectures and reproducible analysis pipelines (Botvinik-Nezer et al. 2020; Poldrack et al. 2020). Digital health platforms supporting open science should implement FAIR data principles through specific technical features (Wilkinson et al. 2016; Wise et al. 2019). Data storage systems need standardized metadata schemas that capture both technical parameters (sensor specifications, sampling rates, processing pipelines) and contextual information (study protocols, participant demographics, environmental factors). For example, accelerometer data should include not just raw measurements but also device model, wearing position, and calibration parameters (Willetts et al. 2018). Standardized data-sharing protocols enable researchers from different institutions to systematically verify findings and build upon each other's work while maintaining appropriate privacy safeguards (Alter and Gonzalez 2018).

In implementing these technical requirements, privacy-preserving data sharing represents a particular challenge in digital health implementations (Price and Cohen 2019). Successful approaches implement multi-tiered access controls where different data components have varying sharing levels (Price and Cohen 2019). Raw sensor data might be restricted due to privacy concerns, while derived features (daily activity levels, sleep patterns, social interaction frequency) can be more widely shared. Implementation of automated anonymization pipelines is crucial—these should not only remove direct identifiers but also assess and mitigate re-identification risk from combined sensor data (Arbuckle and Emam 2020). For example, GPS data requires careful processing to obscure home locations while maintaining sufficient granularity for mobility analysis.

Alongside privacy considerations, the implementation of reproducible processing pipelines presents unique challenges in digital health (Beaulieu-Jones and Greene 2017). Sensor data processing often involves multiple steps—filtering, feature extraction, context classification—each potentially introducing variability (Qiu et al. 2022). Successful implementations maintain detailed provenance records for all processing steps, including software versions, parameter settings, and environmental conditions (Wilson et al. 2017). Docker containers, which are standardized units of software that package up code and all its dependencies, can encapsulate complete processing environments, ensuring computational reproducibility across different systems (Boettiger 2015; Nüst et al. 2020). For instance, accelerometer processing pipelines should document all filtering parameters, activity classification thresholds, and any adaptive processing decisions. By ensuring computational reproducibility through standardized containers (such as Docker) and detailed version-controlled documentation, these approaches provide a scalable framework for multi-site collaboration and model validation in computational psychiatry (Karas et al. 2019; van Hees et al. 2018).

As these methodological frameworks continue to evolve, they create new opportunities for advancing both the scientific rigor and real-world applicability of computational psychiatry. Expanding upon the data collection approaches, population representation strategies, validation frameworks, and engagement methods discussed above, several emerging technological innovations offer promising directions for future development. These advances directly address many of the challenges identified throughout this perspective, from enhancing ecological validity to improving privacy preservation and model generalizability.

## 7 Future innovations in computational psychiatry

Building on the evidence and implementation strategies discussed above, particularly regarding digital phenotyping and context-aware data collection, several key technological developments on the horizon could further transform how we achieve ecological validity in computational psychiatry. These emerging approaches directly address many of the challenges identified throughout this paper, from improving context awareness to enhancing privacy preservation. Advanced sensor integration through upcoming smartphone and wearable technologies offers opportunities for more sophisticated context detection and behavioral monitoring. For instance, smartphone-based sensing combined with ecological momentary assessments can provide rich longitudinal data about behavioral patterns, mental health states, and their relationship to environmental factors, as demonstrated by studies tracking changes in student activity patterns, social interactions, and psychological well-being during periods of significant stress.

Extending the ADOs and hierarchical Bayesian approaches discussed earlier, edge computing—which processes data directly on local devices rather than remote servers—presents opportunities for more sophisticated real-time processing of behavioral data (Rubino et al. 2021). This could enable adaptive computational models to run directly on participants'devices, allowing for more responsive and personalized task adjustment while reducing privacy concerns associated with continuous data transmission (Covi et al. 2021). Such capabilities could also support more complex context detection algorithms that combine multiple sensor streams in real-time, enabling more nuanced environmental classification. For example, combining accelerometer, GPS, ambient noise, and Bluetooth proximity data could enable real-time detection of social interactions in crowded environments, allowing computational tasks assessing social reward processing to be triggered specifically during naturalistic social encounters rather than at random intervals, thereby increasing ecological validity of social cognition measurements.

Expanding on the open science principles and diverse recruitment strategies outlined above, federated learning (a machine learning approach where models are trained across multiple decentralized devices while keeping the data local) offers promising solutions for balancing scientific collaboration with privacy requirements in healthcare (Li et al. 2020), particularly addressing the representativeness challenges discussed in our participant factors section. This approach enables collaborative model development where institutions share only encrypted gradient information rather than raw data while keeping sensitive information localized (Price and Cohen 2019). Future implementations could enable cross-site model training and validation across diverse populations while maintaining privacy, potentially leading to more robust computational models that generalize across different clinical contexts.

These future developments will require careful consideration of ethical implications, particularly around privacy and agency—themes that resonate throughout our earlier discussion of participant factors, expert-by-experience collaboration, and implementation challenges. As implementations become more sophisticated in their ability to detect and respond to behavioral patterns, maintaining appropriate boundaries and participant control will become increasingly crucial, reflecting our earlier emphasis on integrating lived experience perspectives. The field must work toward implementations that enhance clinical utility while respecting participant autonomy and privacy, ultimately serving our core goal of achieving ecological validity in computational psychiatry.

## 8 Conclusion

The advancement of computational psychiatry through digital health technologies presents both significant opportunities and methodological challenges for achieving ecological validity. Our perspective has highlighted how innovative approaches in data collection and trial design are transforming our ability to capture and understand psychiatric phenomena in real-world settings. Adaptive design optimization, when combined with digital phenotyping and passive data collection, offers promising solutions for enhancing ecological validity while maintaining methodological rigor.

The integration of experts-by-experience in model development has emerged as crucial for ensuring that computational approaches meaningfully capture the lived experience of psychiatric conditions. This collaboration has proven particularly valuable in tailoring digital phenotyping approaches and ecological momentary assessments to reflect real-world symptom manifestations. Meanwhile, the adoption of open science practices, especially in digital health implementations, provides essential frameworks for enhancing model reliability and generalizability across different clinical settings.

Critical challenges remain, particularly in addressing algorithmic biases and ensuring equitable benefits across diverse populations. Current evidence indicates significant disparities in model performance across demographic groups, highlighting the need for more representative data collection and validation approaches. The implementation of privacy-preserving data-sharing architectures and standardized processing pipelines represents a crucial step toward more robust and generalizable computational models.

As digital health technologies continue to evolve, maintaining ecological validity while ensuring methodological rigor will remain fundamental to successful clinical translation. The field's future development depends on carefully balancing technological innovation with practical clinical utility, ensuring that computational approaches enhance rather than complicate psychiatric care. Success in this domain requires sustained attention to participant factors, systematic validation approaches, and integration with established clinical expertise, ultimately working toward computational psychiatry tools that can meaningfully improve patient outcomes in real-world settings.

**Author contributions** A.P and M.Y wrote the main manuscript R.K provided input and review All 3 authors reviewed the manuscript

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions.

**Data availability** No datasets were generated or analysed during the current study.

## References

Abbasgholizadeh Rahimi S, Cwintal M, Huang Y, Ghadiri P, Grad R, Poenaru D, Gore G, Zomahoun HTV, Légaré F, Pluye P (2022) Application of artificial intelligence in shared decision making: scoping review. JMIR Med Inform 10(8):e36199. https://doi.org/10.2196/36199

Adams RA, Huys QJM, Roiser JP (2016) Computational psychiatry: towards a mathematically informed understanding of mental illness. J Neurol Neurosurg Psychiatry 87(1):53–63. https://doi.org/10.1136/jnnp-2015-310737

Adler DA, Stamatis CA, Meyerhoff J, Mohr DC, Wang F, Aranovich GJ, Sen S, Choudhury T (2024) 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. https://doi.org/10.1038/s44184-024-00057-y

Ahn W-Y, Gu H, Shen Y, Haines N, Hahn HA, Teater JE, Myung JI, Pitt MA (2020) Rapid, precise, and reliable measurement of delay discounting using a Bayesian learning algorithm. Sci Rep 10(1):12091. https://doi.org/10.1038/s41598-020-68587-x

Alter G, Gonzalez R (2018) Responsible practices for data sharing. Am Psychol 73(2):146–156. https://doi.org/10.1037/amp0000258

Andrade C (2018) Internal, external, and ecological validity in research design, conduct, and evaluation. Indian J Psychol Med 40(5):498–499. https://doi.org/10.4103/IJPSYM.IJPSYM_334_18

Arbuckle L and Emam KE (2020) Building an Anonymization Pipeline: Creating Safe Data. O'Reilly Media, Inc

Bajorek JP (2019) Voice recognition still has significant race and gender biases. Harvard Business Review, 10, 1–4. https://www.researchgate.net/profile/Joan-Bajorek/publication/333004900_Voice_Recognition_Still_Has_Significant_Race_and_Gender_Biases/links/5cd85abf458515712ea673ac/Voice-Recognition-Still-Has-Significant-Race-and-Gender-Biases.pdf

Beaulieu-Jones BK, Greene CS (2017) Reproducibility of computational workflows is automated using continuous analysis. Nat Biotechnol 35(4):342–346. https://doi.org/10.1038/nbt.3780

Benrimoh D, Fisher V, Mourgues C, Sheldon AD, Smith R, Powers AR (2023) Barriers and solutions to the adoption of translational tools for computational psychiatry. Mol Psychiatry 28(6):2189–2196. https://doi.org/10.1038/s41380-023-02114-y

Bertl M, Ross P, Draheim D (2022) A survey on AI and decision support systems in psychiatry: uncovering a dilemma. Expert Syst Appl 202:117464. https://doi.org/10.1016/j.eswa.2022.117464

Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, Maaieh H, Strecher V (2018) To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 6(11):e10123. https://doi.org/10.2196/10123

Birk RH, Samuel G (2022) Digital phenotyping for mental health: reviewing the challenges of using data to monitor and predict mental health problems. Curr Psychiatry Rep 24(10):523–528. https://doi.org/10.1007/s11920-022-01358-9

Boettiger C (2015) An introduction to Docker for reproducible research. SIGOPS Oper Syst Rev 49(1):71–79. https://doi.org/10.1145/2723872.2723882

Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, Kirchler M, Iwanir R, Mumford JA, Adcock RA, Avesani P, Baczkowski BM, Bajracharya A, Bakst L, Ball S, Barilari M, Bault N, Beaton D, Beitner J, Schonberg T (2020) Variability in the analysis of a single neuroimaging dataset by many teams. Nature 582(7810):84–88. https://doi.org/10.1038/s41586-020-2314-9

Brown VM, Chen J, Gillan CM, Price RB (2020) Improving the reliability of computational analyses: model-based planning and its relationship with compulsivity. Biol Psych Cogn Neurosci Neuroimag 5(6):601–609. https://doi.org/10.1016/j.bpsc.2019.12.019

Bucci S, Berry N, Morris R, Berry K, Haddock G, Lewis S, Edge D (2019) They are not hard-to-reach clients. We have just got hard-to-reach services. Staff views of digital health tools in specialist mental health services. Front Psych. https://doi.org/10.3389/fpsyt.2019.00344

Buolamwini J and Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on Fairness, Accountability and Transparency, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline&ref=akusion-ci-shi-dai-bizinesumedeia

Büscher R, Winkler T, Mocellin J, Homan S, Josifovski N, Ciharova M, van Breda W, Kwon S, Larsen ME, Torous J, Firth J, Sander LB (2024) A systematic review on passive sensing for the prediction of suicidal thoughts and behaviors. Npj Mental Health Research 3(1):42. https://doi.org/10.1038/s44184-024-00089-4

Capon H, Hall W, Fry C, Carter A (2016) Realising the technological promise of smartphones in addiction research and treatment: an ethical review. Int J Drug Policy 36:47–57. https://doi.org/10.1016/j.drugpo.2016.05.013

Chambers DA (2020) Sharpening our focus on designing for dissemination: lessons from the SPRINT program and potential next steps for the field. Translat Behav Med 10(6):1416–1418. https://doi.org/10.1093/tbm/ibz102

Cholaquidis A, Fraiman R and Hernandez-Banadik M (2023) Home range estimation under a restricted sampling scheme (No. arXiv:2106.02035). arXiv. https://doi.org/10.48550/arXiv.2106.02035

Constant A, Badcock P, Friston K, Kirmayer LJ (2022) Integrating evolutionary, cultural, and computational psychiatry: a multilevel systemic approach. Front Psychiat. https://doi.org/10.3389/fpsyt.2022.763380

Cornet VP, Holden RJ (2018) Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 77:120–132. https://doi.org/10.1016/j.jbi.2017.12.008

Covi E, Donati E, Liang X, Kappel D, Heidari H, Payvand M, Wang W (2021) Adaptive extreme edge computing for wearable devices. Front Neurosci. https://doi.org/10.3389/fnins.2021.611300

D'Alfonso S (2020) AI in mental health. Cyberpsychology 36:112–117. https://doi.org/10.1016/j.copsyc.2020.04.005

Dautov R, Distefano S, Buyya R (2019) Hierarchical data fusion for smart healthcare. J Big Data 6(1):19. https://doi.org/10.1186/s40537-019-0183-6

Davidson T, Bhattacharya D and Weber I (2019) Racial Bias in Hate Speech and Abusive Language Detection Datasets (No. arXiv:1905.12516). arXiv. https://doi.org/10.48550/arXiv.1905.12516

Dima D, Modabbernia A, Papachristou E, Doucet GE, Agartz I, Aghajani M, Akudjedu TN, Albajes-Eizagirre A, Alnæs D, Alpert KI, Andersson M, Andreasen NC, Andreassen OA, Asherson P, Banaschewski T, Bargallo N, Baumeister S, Baur-Streubel R, Bertolino A, Karolinska Schizophrenia Project (KaSP) (2022) Subcortical volumes across the lifespan: data from 18,605 healthy individuals aged 3–90 years. Hum Brain Mapp 43(1):452–469. https://doi.org/10.1002/hbm.25320

Frangou S, Modabbernia A, Williams SCR, Papachristou E, Doucet GE, Agartz I, Aghajani M, Akudjedu TN, Albajes-Eizagirre A, Alnæs D, Alpert KI, Andersson M, Andreasen NC, Andreassen OA, Asherson P, Banaschewski T, Bargallo N, Baumeister S, Baur-Streubel R, Dima D (2022) Cortical thickness across the lifespan: data from 17,075 healthy individuals aged 3–90 years. Hum Brain Mapp 43(1):431–451. https://doi.org/10.1002/hbm.25364

Friston K (2023) Computational psychiatry: from synapses to sentience. Mol Psychiatry 28(1):256–268. https://doi.org/10.1038/s41380-022-01743-z

Fusar-Poli P, Manchia M, Koutsouleris N, Leslie D, Woopen C, Calkins ME, Dunn M, Tourneau CL, Mannikko M, Mollema T, Oliver D, Rietschel M, Reininghaus EZ, Squassina A, Valmaggia L, Kessing LV, Vieta E, Correll CU, Arango C, Andreassen OA (2022) Ethical considerations for precision psychiatry: a roadmap for research and clinical practice. Eur Neuropsychopharmacol 63:17–34. https://doi.org/10.1016/j.euroneuro.2022.08.001

Gannedahl A, Björkholm C, Leval A, Cars T, Hellner C, Lundberg J (2024) Association between neighborhood socioeconomic status and healthcare utilization among individuals with a first diagnosis of major depressive disorder in primary care in the Stockholm region. J Affect Disord. https://doi.org/10.1016/j.jad.2024.06.074

Ghahramani M, Zhou M, Wang G (2020) Urban sensing based on mobile phone data: approaches, applications, and challenges. IEEE/CAA J Automat Sin 7(3):627–637. https://doi.org/10.1109/JAS.2020.1003120

Gillan CM and Rutledge RB (2021) Smartphones and the Neuroscience of Mental Health. In Annual Review of Neuroscience (Vol. 44, Issue Volume 44, 2021, pp. 129–151). Annual Reviews. https://doi.org/10.1146/annurev-neuro-101220-014053

Giurgiu M, Niermann C, Ebner-Priemer U, Kanning M (2020) Accuracy of sedentary behavior-triggered ecological momentary assessment for collecting contextual information: development and feasibility study. JMIR Mhealth Uhealth 8(9):e17852. https://doi.org/10.2196/17852

Hitchcock PF, Fried EI, and Frank MJ (2022) Computational psychiatry needs time and context. In Annual Review of Psychology (Vol. 73, Issue Volume 73, 2022, pp. 243–270). Annual Reviews. https://doi.org/10.1146/annurev-psych-021621-124910

Huang MY, and Parikh H (2024) Towards Generalizing Inferences from Trials to Target Populations. Harvard Data Science Review, 6(4). https://assets.pubpub.org/jcjvqwck/Huang%20&%20Parikh%20(2024)_Just%20Accepted-01723833367992.pdf

Huys QJM, Maia TV, Frank MJ (2016) Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci 19(3):404–413. https://doi.org/10.1038/nn.4238

Insel TR (2017) Digital phenotyping: technology for a new science of behavior. JAMA 318(13):1215–1216. https://doi.org/10.1001/jama.2017.11295

Karas M, Bai J, Strączkiewicz M, Harezlak J, Glynn NW, Harris T, Zipunnikov V, Crainiceanu C, Urbanek JK (2019) Accelerometry data in health research: challenges and opportunities. Stat Biosci 11(2):210–237. https://doi.org/10.1007/s12561-018-9227-2

Karvelis P, Paulus M, and Diaconescu A (2022) Individual differences in computational psychiatry: a review of current challenges. OSF https://doi.org/10.31234/osf.io/bvjzn

Koenecke A, Nam A, Lake E, Nudell J, Quartey M, Mengesha Z, Toups C, Rickford JR, Jurafsky D, Goel S (2020) Racial disparities in automated speech recognition. Proc Natl Acad Sci 117(14):7684–7689. https://doi.org/10.1073/pnas.1915768117

Kolding S, Lundin RM, Hansen L, Østergaard SD (2024) Use of generative artificial intelligence (AI) in psychiatry and mental health care: a systematic review. Acta Neuropsychiatrica. https://doi.org/10.1017/neu.2024.50

Kwon M, Lee SH, Ahn W-Y (2023) Adaptive design optimization as a promising tool for reliable and efficient computational fingerprinting. Reliab Neurocogn Meas Mental Health 8(8):798–804. https://doi.org/10.1016/j.bpsc.2022.12.003

Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim H-C, Paulus MP, Krystal JH, Jeste DV (2021) Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. Biol Psychiat Cogn Neurosci Neuroimag 6(9):856–864. https://doi.org/10.1016/j.bpsc.2021.02.001

Leslie D, Mazumder A, Peppin A, Wolters MK, Hagerty A (2021) Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? BMJ 372:n304. https://doi.org/10.1136/bmj.n304

Levkovich I, Shinan-Altman S, and Elyoseph Z (2024) Can large language models be sensitive to culture suicide risk assessment? Research Square https://doi.org/10.21203/rs.3.rs-4066705/v1

Li Z, Sharma V, Mohanty SP (2020) Preserving data privacy via federated learning: challenges and solutions. IEEE Cons Electr Mag 9(3):8–16. https://doi.org/10.1109/MCE.2019.2959108

McCradden M, Hui K, Buchman DZ (2023) Evidence, ethics and the promise of artificial intelligence in psychiatry. J Med Ethics 49(8):573. https://doi.org/10.1136/jme-2022-108447

Mendes JPM, Moura IR, Van de Ven P, Viana D, Silva FJS, Coutinho LR, Teixeira S, Rodrigues JJPC, Teles AS (2022) Sensing apps and public data sets for digital phenotyping of mental health: systematic review. J Med Internet Res 24(2):e28735. https://doi.org/10.2196/28735

Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM (2017) Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 19(5):e153. https://doi.org/10.2196/jmir.7725

Monteith S, Glenn T, Geddes JR, Achtyes ED, Whybrow PC, Bauer M (2023) Challenges and ethical considerations to successfully implement artificial intelligence in clinical medicine and neuroscience: a narrative review. Pharmacopsychiatry 56(6):209–213. https://doi.org/10.1055/a-2142-9325

Nicholas J, Onie S, Larsen ME (2020) Ethics and privacy in social media research for mental health. Curr Psychiatry Rep 22(12):84. https://doi.org/10.1007/s11920-020-01205-9

Nüst D, Sochat V, Marwick B, Eglen SJ, Head T, Hirst T, Evans BD (2020) Ten simple rules for writing Dockerfiles for reproducible data science. PLoS Comput Biol 16(11):e1008316. https://doi.org/10.1371/journal.pcbi.1008316

Olbert CM, Nagendra A, Buck B (2018) Meta-analysis of Black vs. White racial disparity in schizophrenia diagnosis in the United States: do structured assessments attenuate racial disparities? J Abnor Psychol 127(1):104–115. https://doi.org/10.1037/abn0000309

Oudin A, Maatoug R, Bourla A, Ferreri F, Bonnot O, Millet B, Schoeller F, Mouchabac S, Adrien V (2023) Digital phenotyping: data-driven psychiatry to redefine mental health. J Med Internet Res 25(1):e44502. https://doi.org/10.2196/44502

Øvretveit J, Zubkoff L, Nelson EC, Frampton S, Knudsen JL, Zimlichman E (2017) Using patient-reported outcome measurement to improve patient care. Int J Qual Health Care 29(6):874–879. https://doi.org/10.1093/intqhc/mzx108

Pacheco-Romero AM, Vázquez C, Peinado V, Contreras A, Trucharte A, Valiente C (2024) Not all psychological symptoms are remembered the same: recall biases of symptom severity in the general population. J Appl Res Mem Cogn No Paginat Spec-No Pag Spec. https://doi.org/10.1037/mac0000189

Poldrack RA, Huckins G, Varoquaux G (2020) Establishment of best practices for evidence for prediction: a review. JAMA Psychiat 77(5):534–540. https://doi.org/10.1001/jamapsychiatry.2019.3671

Pooseh S, Bernhardt N, Guevara A, Huys QJM, Smolka MN (2018) Value-based decision-making battery: a Bayesian adaptive approach to assess impulsive and risky behavior. Behav Res Methods 50(1):236–249. https://doi.org/10.3758/s13428-017-0866-x

Pramanik PKD, Sinhababu N, Mukherjee B, Padmanaban S, Maity A, Upadhyaya BK, Holm-Nielsen JB, Choudhury P (2019) Power consumption analysis, measurement, management, and issues: a state-of-the-art review of smartphone battery and energy usage. IEEE Access 7:182113–182172. https://doi.org/10.1109/ACCESS.2019.2958684

Price WN, Cohen IG (2019) Privacy in the age of medical big data. Nat Med 25(1):37–43. https://doi.org/10.1038/s41591-018-0272-7

Qiu S, Zhao H, Jiang N, Wang Z, Liu L, An Y, Zhao H, Miao X, Liu R, Fortino G (2022) Multi-sensor information fusion based on machine learning for real applications in human activity recognition: state-of-the-art and research challenges. Information Fusion 80:241–265. https://doi.org/10.1016/j.inffus.2021.11.006

Roy AL (2017) Innovation or violation? Leveraging mobile technology to conduct socially responsible community research. Am J Community Psychol 60(3–4):385–390. https://doi.org/10.1002/ajcp.12187

Rubino A, Livanelioglu C, Qiao N, Payvand M, Indiveri G (2021) Ultra-low-power FDSOI neural circuits for extreme-edge neuromorphic intelligence. IEEE Trans Circuits Syst I Regul Pap 68(1):45–56. https://doi.org/10.1109/TCSI.2020.3035575

Sahin D, Kambeitz-Ilankovic L, Wood S, Dwyer D, Upthegrove R, Salokangas R, Borgwardt S, Brambilla P, Meisenzahl E, Ruhrmann S, Schultze-Lutter F, Lencer R, Bertolino A, Pantelis C, Koutsouleris N, Kambeitz J (2024) Algorithmic fairness in precision psychiatry: analysis of prediction models in individuals at clinical high risk for psychosis. Br J Psychiat 224(2):55–65. https://doi.org/10.1192/bjp.2023.141

Salazar de Pablo G, Studerus E, Vaquerizo-Serrano J, Irving J, Catalan A, Oliver D, Baldwin H, Danese A, Fazel S, Steyerberg EW, Stahl D, Fusar-Poli P (2021) Implementing precision psychiatry: a systematic review of individualized prediction models for clinical practice. Schizophr Bull 47(2):284–297. https://doi.org/10.1093/schbul/sbaa120

Sanchez C, Grzenda A, Varias A, Widge AS, Carpenter LL, McDonald WM, Nemeroff CB, Kalin NH, Martin G, Tohen M, Filippou-Frye M, Ramsey D, Linos E, Mangurian C, Rodriguez CI (2020) Social media recruitment for mental health research: a systematic review. Compr Psychiatry 103:152197. https://doi.org/10.1016/j.comppsych.2020.152197

Sanfelici R, Dwyer DB, Antonucci LA, Koutsouleris N (2020) Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: a meta-analytic view on the state of the art. Psych Risk Synd 88(4):349–360. https://doi.org/10.1016/j.biopsych.2020.02.009

Sap M, Card D, Gabriel S, Choi Y, and Smith NA (2019) The risk of racial bias in hate speech detection. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1668–1678. https://aclanthology.org/P19-1163/?utm_campaign=piqcy&utm_medium=email&utm_source=Revue%20newsletter

Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M (2021) Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med 27(12):2176–2182. https://doi.org/10.1038/s41591-021-01595-0

Simblett S, Matcham F, Siddi S, Bulgari V, di San B, Pietro C, Hortas López J, Ferrão J, Polhemus A, Haro JM, de Girolamo G, Gamble P, Eriksson H, Hotopf M, Wykes T (2019) Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR Mhealth Uhealth 7(1):e11325. https://doi.org/10.2196/11325

Thompson PM, Jahanshad N, Ching CRK, Salminen LE, Thomopoulos SI, Bright J, Baune BT, Bertolín S, Bralten J, Bruin WB, Bülow R, Chen J, Chye Y, Dannlowski U, de Kovel CGF, Donohoe G, Eyler LT, Faraone SV, Favre P (2020) ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. Translat Psychiat 10(1):100. https://doi.org/10.1038/s41398-020-0705-1

Thorlund K, Haggstrom J, Park JJ, Mills EJ (2018) Key design considerations for adaptive clinical trials: a primer for clinicians. BMJ 360:k698. https://doi.org/10.1136/bmj.k698

Toch E, Lerner B, Ben-Zion E, Ben-Gal I (2019) Analyzing large-scale human mobility data: a survey of machine learning methods and applications. Knowl Inf Syst 58(3):501–523. https://doi.org/10.1007/s10115-018-1186-x

Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, Carvalho AF, Keshavan M, Linardon J, Firth J (2021) The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 20(3):318–335. https://doi.org/10.1002/wps.20883

Treadway MT (2023) Computational psychiatry and the lived experience of mental illness. Nat Rev Psychol 2(2):67–68. https://doi.org/10.1038/s44159-022-00142-z

Upreti R, Lind PG, Elmokashfi A, Yazidi A (2024) Trustworthy machine learning in the context of security and privacy. Int J Inf Secur 23(3):2287–2314. https://doi.org/10.1007/s10207-024-00813-3

Uusitalo S, Tuominen J, Arstila V (2021) Mapping out the philosophical questions of AI and clinical practice in diagnosing and treating mental disorders. J Eval Clin Pract 27(3):478–484. https://doi.org/10.1111/jep.13485

van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimäki M, Frayling TM, Pack AI, Bucan M, Trenell MI, Mazzotti DR, Gehrman PR, Singh-Manoux BA, Weedon MN (2018) Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep 8(1):12975. https://doi.org/10.1038/s41598-018-31266-z

Vuqetrna E, Selai C, Golder D, Moutoussis M (2023) 21 How can we collaborate with experts-by-example to improve the ecological validity of computational psychiatry? A pilot study where the experts-by-experience are people with functional neurological disorder (FND) and the overall aim is to understand what leads patients unilaterally to disengage from psychotherapy. J Neurol Neurosurg Psychiatry 94(12):e2. https://doi.org/10.1136/JNNP-2023-BNPA.37

Wen CKF, Schneider S, Stone AA, Spruijt-Metz D (2017) Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 19(4):e132. https://doi.org/10.2196/jmir.6641

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3(1):160018. https://doi.org/10.1038/sdata.2016.18

Willetts M, Hollowell S, Aslett L, Holmes C, Doherty A (2018) Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. Sci Rep 8(1):7961. https://doi.org/10.1038/s41598-018-26174-1

Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK (2017) Good enough practices in scientific computing. PLoS Comput Biol 13(6):e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Wise J, de Barron AG, Splendiani A, Balali-Mood B, Vasant D, Little E, Mellino G, Harrow I, Smith I, Taubert J, van Bochove K, Romacker M, Walgemoed P, Jimenez RC, Winnenburg R, Plasterer T, Gupta V, Hedley V (2019) Implementation and relevance of FAIR data principles in biopharmaceutical R&D. Drug Discovery Today 24(4):933–938. https://doi.org/10.1016/j.drudis.2019.01.008

Wong TY, Luo H, Tang J, Moore TM, Gur RC, Suen YN, Hui CLM, Lee EHM, Chang WC, Yan WC, Chui E, Poon LT, Lo A, Cheung K-M, Kan CK, Chen EYH, Chan SKW (2024) Development of an individualized risk calculator of treatment resistance in patients with first-episode psychosis (TRipCal) using automated machine learning: a 12-year follow-up study with clozapine prescription as a proxy indicator. Transl Psychiatry 14(1):50. https://doi.org/10.1038/s41398-024-02754-w

Wrzus C, Neubauer AB (2023) Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 30(3):825–846. https://doi.org/10.1177/10731911211067538

Xia Q, Ye W, Tao Z, Wu J, Li Q (2021) A survey of federated learning for edge computing: research problems and solutions. High-Confid Comp 1(1):100008. https://doi.org/10.1016/j.hcc.2021.100008

Zidaru T, Morrow EM, Stockley R (2021) Ensuring patient and public involvement in the transition to AI-assisted mental health care: a systematic scoping review and agenda for design justice. Health Expect 24(4):1072–1124. https://doi.org/10.1111/hex.13299