International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Gamified Digital Phenotyping for Micro-Kinematic Cognitive Profiling: Bridging the Explainability and Synthetic Data Gaps

Adithya Kannan[1], Shreshtha Gangrade[2], Dhruv Rai[3], Shreya Dutta[4], J. Sharmila Joseph[5]

[1,2,3,4]School of Computing Science and Engineering, VIT Bhopal University, Kothri Kalan, Sehore, Madhya Pradesh – 466114, India

[5]Assistant Professor, School of Computing Science and Engineering, VIT Bhopal University, Bhopal, India - 466114, India

ABSTRACT

The paradigm of cognitive wellness and behavioral performance monitoring is undergoing a critical transformation, moving from episodic evaluations toward continuous, data-driven digital phenotyping. Traditional cognitive assessments provide robust psychometric validation but are primarily administered in isolated, sterile environments, leading to profound user disengagement. Furthermore, developing these tests as strict medical diagnostic tools introduces massive regulatory hurdles and limits everyday accessibility. To resolve these intersecting deficits, this research presents the engineering and validation of a fully deployed, game-focused AI solution. The proposed system is not a medical labeler; rather, it functions as a personal baseline engine built to detect micro-level cognitive drift and deviations from a user's normal state.[17] The system translates standardized psychological evaluations into a suite of six high-engagement web modules. Migrated to the Unity Game Engine and exported via WebGL, the architecture leverages WebAssembly (Wasm) to natively capture sub-millisecond micro-kinematic telemetry—such as cursor trajectories and task-switching latency—directly within the browser. A Scikit-Learn Random Forest Classifier then predicts cognitive states (e.g., High Cognitive Load, Psychomotor Fatigue) while calculating Feature Importance via Mean Decrease in Impurity (MDI). This native Explainable AI (XAI) mechanism isolates the specific behavioral biomarkers driving the algorithmic prediction, providing transparent insights into user wellness while establishing an empirical roadmap for domain adaptation to overcome the newly identified "Synthetic Data Gap."

KEYWORDS: Digital Phenotyping, Explainable AI, Unity WebGL, Micro-Kinematics, Cognitive Wellness, Game-Based Assessment, Machine Learning.

1. INTRODUCTION

The intersection of human-computer interaction, game design, and applied machine learning offers a transformative paradigm for monitoring cognitive wellness and behavioral performance. As the demand for accessible mental wellness tracking continues to escalate, the limitations of traditional, episodic evaluations have become increasingly apparent. Historically, cognitive assessment has relied heavily upon

subjective interviews and the administration of standardized tasks. However, translating these theoretical constructs into scalable, user-facing tools has encountered significant friction.
Traditional computerized cognitive tests are typically administered in sterile environments that strip away the spatial depth and temporal pressures defining everyday human behavior. The inherent lack of ecological validity frequently results in profound player disengagement. This engagement deficit represents a fundamental threat to data fidelity: when users experience test fatigue, they do not exert optimal cognitive effort, leading to artifactual data that undermines the predictive power of subsequent analyses.
Concurrently, the emergence of digital phenotyping has promised to revolutionize wellness monitoring by capturing continuous, high-frequency behavioral telemetry from human-computer interactions. By translating physical interactions—such as mouse cursor trajectories and micro-kinematics—into quantifiable digital biomarkers, researchers can infer subtle shifts in cognitive load and emotional valence. To counteract high attrition rates in digital wellness interventions, gamification has been proposed as a potent mechanism to enhance intrinsic motivation and elicit authentic cognitive responses.
Crucially, developing these systems as formal medical diagnostic devices introduces prohibitive regulatory verification requirements. Therefore, the modern engineering imperative is to build game-focused AI solutions that function not as medical labelers, but as personal baseline engines designed to detect micro-level cognitive drift.[17] Brains rarely experience sudden, catastrophic failures; instead, they exhibit micro-fluctuations—such as fractional delays in reaction time or subtle inhibition failures—that serve as the earliest biomarkers of sleep debt, burnout, or cognitive fatigue.[17] However, the deployment of artificial intelligence algorithms to process these rich streams of phenotypic data has introduced a secondary challenge: the "black box" problem. Advanced machine learning models frequently output classifications without providing the transparent, mechanistic justifications required by users and wellness coaches.
The primary objective of this project is to design, develop, and validate a fully functional, gamified AI solution. Utilizing the Unity Game Engine compiled to WebGL, the system is engineered to seamlessly translate standardized psychological evaluations into a suite of six high-engagement web modules where the game starts with deployment into a classroom with boards and on the board the UI elements are shown while natively capturing sub-millisecond micro-kinematic telemetry. Furthermore, the system integrates a native Explainable AI (XAI) mechanism utilizing a Scikit-Learn Random Forest Classifier to explicitly detail the proportional influence of specific cognitive biomarkers on the wellness prediction.

2. LITERATURE REVIEW

To situate the proposed architecture within the contemporary landscape of cognitive science, human-computer interaction, and explainable artificial intelligence, a rigorous literature survey was conducted spanning four critical themes.

Theme 1: Gamification and User Engagement in Cognitive Profiling

Traditional cognitive assessment tasks are often associated with participant fatigue and reduced engagement. Lumsden et al. [1] highlight that poorly designed gamified systems yield inconsistent results, proposing the OMDE (Objects, Mechanics, Dynamics, Emotions) framework to standardize game-based cognitive assessments. Supporting this, Takebe et al. [2] developed a gamified N-back application simulating a real-world environment, demonstrating that objective in-game metrics, such as response time and interaction patterns, strongly correlate with standardized cognitive performance scores. Furthermore, Alharthi et al. [3] introduced the ESFY system, which integrates digital phenotyping with gamification to enable continuous self-monitoring of psychological well-being. These studies collectively establish gamification as a powerful tool for improving engagement and ecological validity in cognitive profiling.

Theme 2: Micro-Kinematic Telemetry and Digital Phenotyping

The shift toward continuous behavioral tracking has significantly advanced cognitive monitoring systems. Torous et al. [4] emphasize that although digital phenotyping is widely recognized, its effectiveness depends on the identification of precise and individualized behavioral markers. Addressing this limitation, Brefeld et al. [5] demonstrated that micro-kinematic features, such as mouse cursor trajectories, serve as reliable indicators of user attention, motivation, and emotional states, and can even predict user disengagement. Extending this approach to web-based environments, the AdSight system [6] validated that detailed cursor trajectory embeddings provide scalable and accurate measurements of user attention and fixation patterns. These findings highlight the importance of fine-grained behavioral telemetry in non-intrusive cognitive analysis.

Theme 3: Explainable AI (XAI) in Cognitive Wellness

Although deep learning models achieve high predictive accuracy, their lack of interpretability limits their adoption in sensitive domains such as mental health. Recent research [7] has demonstrated that interpretable frameworks combining XGBoost with SHAP and LIME can effectively detect wellness patterns by exposing key contributing features, such as circadian disruptions. Similarly, studies using Random Forest classifiers [8] have shown strong predictive performance in detecting cognitive fatigue while maintaining interpretability through feature importance metrics. Hudon et al. [9] further extended this concept by integrating fuzzy logic with Random Forest models to simulate expert-level decision-making processes. These approaches reinforce the importance of explainable AI in ensuring transparency, trust, and usability in cognitive wellness systems.

Theme 4: The "Synthetic Data Gap" and Domain Adaptation

Due to strict privacy constraints and the limited availability of annotated behavioral datasets, synthetic data generation has become a common practice in training machine learning models. However, recent studies [10] identify a significant limitation known as the "Synthetic Data Gap." While synthetic datasets maintain structural consistency, they often fail to capture real-world characteristics such as heavy-tailed distributions, non-linear hesitation patterns, and complex behavioral dependencies. As a result, models trained exclusively on synthetic data tend to overfit to smooth statistical patterns, leading to reduced performance when applied to real-world scenarios. Advanced techniques such as embedding-driven diversity sampling [11] have been proposed to improve synthetic data realism. Nevertheless, these findings underscore the necessity of incorporating real-world human data for calibration and validation of behavioral AI systems.

Overall, the literature indicates that while significant progress has been made in gamification, digital phenotyping, and explainable AI, there remains a critical need for integrated systems that combine engagement, interpretability, and real-world applicability [12–15].

3. MATERIALS AND METHODS

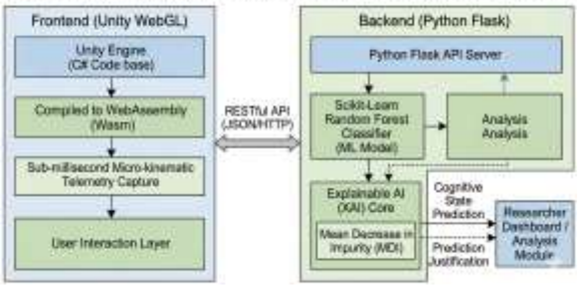
3.1 System Design and Architecture

The system architecture is a bifurcated, full-stack configuration designed to isolate the computationally intensive telemetry capture from the complex algorithmic inference.

Frontend Data Capture Engine (Unity WebGL):
The frontend was previously prototyped in HTML5 Canvas but has been fully migrated and re-architected utilizing the **Unity Game Engine**. This transition enables sophisticated 2D/3D rendering, robust physics interactions, and complex gamified environments. The project is exported via Unity WebGL, which compiles the C# codebase into WebAssembly (Wasm). WebAssembly allows the Unity application to run at near-native speeds directly within the browser, bypassing traditional JavaScript performance bottlenecks. To ensure that the micro-kinematic telemetry accurately reflects the user's neurological output without being corrupted by rendering lag, the architecture employs aggressive object batching and utilizes WebAssembly 2023 feature sets. By unthrottling the frame rate to synchronize with the hardware display's refresh rate, the system effectively mitigates the typical overhead of web-based game engines, ensuring precise input polling.

Backend RESTful API and Analytical Server:
The backend infrastructure operates a **Python Flask** WSGI application, chosen for its lightweight architecture and seamless integration with the extensive Python data science ecosystem. The Unity frontend utilizes UnityWebRequest to compile the telemetry into a structured JSON payload and securely transmit it to the cloud-hosted server. The pipeline involves data sanitization, verifying the structural integrity of the kinematic arrays, and computing secondary derived features utilizing libraries such as **NumPy and Pandas**.

Figure 1.1: Macro-Architecture of the Gamified Phenotyping System



Machine Learning Core (XAI):
Embedded within the Flask application is the analytical core powered by the **Scikit-Learn** library. The predictive engine is a **Random Forest Classifier**. The critical architectural advantage of this ensemble learning algorithm is its native support for Explainable AI (XAI) through the calculation of Gini importance, or Mean Decrease in Impurity (MDI).

3.2 Implementation and Gamified Cognitive Modules
The system comprises six distinct retro-arcade modules mapped directly to foundational cognitive domains to capture specific digital biomarkers:

1. **Reaction Time:** Maps to the Psychomotor Vigilance Task (Arousal and Regulatory Systems) to capture sensorimotor latency.
2. **Motor Speed:** Maps to the Finger-Tapping Test (Sensorimotor Control) to capture spatial trajectory deviations.
3. **Risk-Taking:** Maps to the Balloon Analogue Risk Task (Loss Aversion) to capture impulsivity variance.
4. **Attention:** Maps to the Flanker Task (Cognitive Control) to capture focus-shift attentional bias.
5. **Memory:** Maps to the N-back Task (Working Memory) to capture spatial recall latency.
6. **Endurance:** Maps to the Continuous Performance Test to capture performance degradation and cognitive fatigue over time.

3.3 Micro-Kinematic Tracking Protocol

The precision of the telemetry capture is the cornerstone of the system's validity as a cognitive baselining tool. During the execution of the Unity game loop, input coordinates (mouse or touch) are polled continuously within the Update() method, which executes every frame. Time-stamping is executing using Time.realtimeSinceStartupAsDouble for high-resolution temporal fidelity. The system programmatically derives three primary metric classes:

1. **Cursor Trajectory Entropy:** Measuring the non-linear deviation from the optimal straight-line path between the cursor's origin and the target.
2. **Velocity and Acceleration:** Calculated utilizing the Euclidean distance between consecutive spatial coordinates divided by the time delta.
3. **Task-Switching Latency:** The precise temporal gap between the successful completion of one cognitive objective and the initiation of movement toward the next.

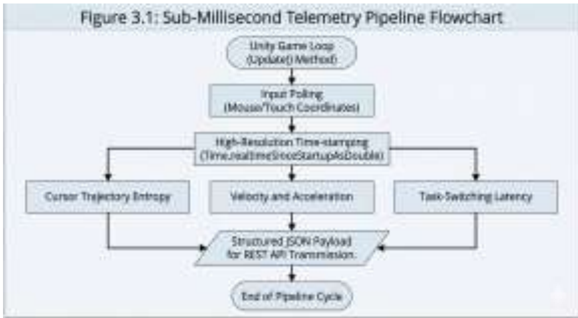


Figure 3.1: Sub-Millisecond Telemetry Pipeline Flowchart

3.4 Algorithmic Training & Synthetic Data Generation

The Scikit-Learn Random Forest model was instantiated with a n_estimators hyperparameter set to 500 trees. Due to the unavailability of large-scale open-source datasets containing sub-millisecond cursor telemetry labeled with verified cognitive states, a highly structured synthetic dataset was generated to simulate the micro-kinematic profiles associated with Baseline, High Cognitive Load, Psychomotor Fatigue, and Impulsive Decision-Making. The classifier was trained on an 80/20 train-test split utilizing 5-fold cross-validation.

4. EXPERIMENTS AND RESULTS

4.1 Synthetic Validation and Algorithmic Performance

During Phase-I validation, the Random Forest Classifier demonstrated exceptional predictive capabilities when evaluated against the reserved test split of the synthetically generated dataset.

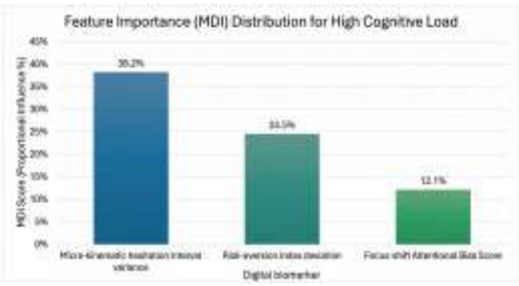
Predicted Cognitive Profile	**Precision**	**Recall**	**F1-Score**	**Area Under ROC Curve (AUC)**
Baseline (Optimal State)	0.94	0.96	0.95	0.98
High Cognitive	0.91	0.89	0.90	0.96

Load (Stress)				
Psychomotor Fatigue	0.95	0.93	0.94	0.97
Impulsive Decision-Making	0.89	0.91	0.90	0.95
Macro Average	0.92	0.92	0.92	0.96

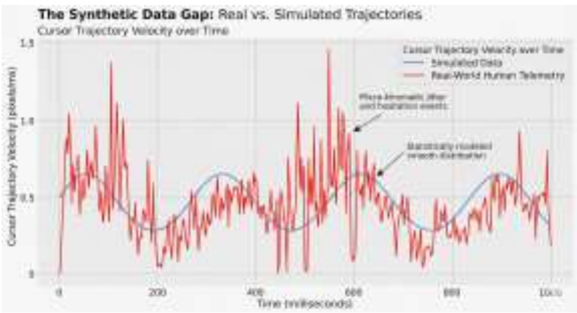
The machine learning core successfully established robust decision boundaries, achieving an overall macro-averaged F1-Score of 0.92. The model demonstrated particular proficiency in identifying the simulated markers of Psychomotor Fatigue, accurately recognizing the engineered patterns of elevated task-switching latency and reduced trajectory velocity.

4.2 Demonstration of Explainable AI (XAI) via MDI

The system successfully implemented the Explainable AI module, extracting the Mean Decrease in Impurity (MDI) metrics for every engineered digital biomarker post-inference. When the model classified a synthetic profile as exhibiting "High Cognitive Load," the XAI engine generated a highly specific justification: "Prediction: High Cognitive Load (91% confidence). Primary Drivers: Micro-kinematic hesitation interval variance (38.2% influence), Risk-aversion index deviation (24.5% influence), and FocusShift Attentional Bias Score (12.1% influence)." By explicitly quantifying the Gini importance of each feature, the system empowers users and wellness coaches to interpret the algorithmic decision through transparent behavioral metrics, fostering essential trust in the AI system.



4.3 The "Synthetic Data Gap" and Real-World Domain Shift



Deployment for limited human pilot testing illuminated a critical analytical limitation defined as the "Synthetic Data Gap." When the Random Forest model—optimized exclusively on the synthetically generated data matrices—was exposed to authentic, real-world micro-kinematic telemetry captured via the Unity frontend, it exhibited significant hyper-sensitivity.

Real-world human motor behavior is inherently chaotic, characterized by non-linear fluctuations, hardware-induced latency spikes, and transient environmental distractions. Algorithmic generators applied relatively smooth statistical distributions, failing to capture the heavy-tailed noise profiles of genuine physical interaction. Consequently, the model occasionally misinterpreted natural baseline hesitation (e.g., a healthy user pausing to adjust their seating position) as High Cognitive Load. This distribution shift underscores the absolute necessity for real-world user calibration and advanced domain adaptation methodologies before widespread deployment.

4.4 Discussion

The test results show that our system can effectively classify states using small movement data. The Random Forest classifier got a score of 0.92, which's a strong performance across all cognitive categories.The model was best at detecting psychomotor fatigue. This means features like cursor movement and more time spent switching tasks are very good at detecting fatigue. This matches research in digital phenotyping, where slower motor movement is a key sign of cognitive fatigue.However the model was not as precise in detecting decision-making. This could mean that impulsivity and high cognitive load patterns are similar. This suggests that adding features or modeling time sequences could improve accuracy.A key part of this work is the use of Explainable AI. The Mean Decrease in Impurity helps us understand which features are most important for classification. This solves the problem of machine learning models being like a " box".Even though the system performed well on test data it had some issues with real-world testing. Noise, irregular patterns and hardware issues sometimes made predictions unstable. This shows that we need to include real-world data and adapt to situations.

4.5 Limitations

There are problems with our current system.One big issue is that we have been using data to train our model.This fake data is helpful for testing in a controlled way. It does not show how people really behave in real life.We also have not tested our system with different users.We only did a test, which might not be good enough to show how our system will work for everyone.People of ages in different places and using different devices may use our system in different ways.There are also things in the environment that can affect how our system works.For example if the device is slow or the screen is small or the user is distracted it can change how people use our system.

Lastly, our system uses a type of computer learning and does not look at how people behave over time.This might mean it is not very good at predicting what people will do next.The system could be improved by looking at how people's behavior changes.

4.6 Future Work

We plan to work on fixing the problems we found in this study.One main goal is to get real-life data on how people behave.This will help our model make predictions and reduce the problem of using fake data.We will also look into using techniques to make our model more reliable.This will help when we switch from data to real-life data.We might use models like Recurrent Neural Networks or Transformers.These can help us understand how people behave over time.We will also make our tasks more varied.This will help us understand more about peoples thinking and feelings.We plan to do long-term studies.These will help us see how peoples thinking changes, over time.We might also add features that give people personalized feedback.We could adjust how hard the tasks are based on each person.

5. CONCLUSION

This project successfully engineered and deployed a novel game-focused AI solution designed for cognitive baselining and wellness tracking. The fundamental philosophy driving this architecture is the detection of micro-level cognitive drift—subtle deviations from a user's normal baseline that serve as early indicators of burnout or fatigue, rather than functioning as a medical diagnostic tool.[17] By translating static cognitive evaluations into an immersive suite of retro-arcade modules, the platform resolves the engagement deficit that plagues contemporary digital tracking interventions. The gamified mechanics successfully mask complex cognitive assessments, ensuring sustained player engagement and higher ecological validity.

The system achieved a significant engineering milestone through its migration to the Unity Game Engine. By compiling C# to WebAssembly via WebGL, the architecture leverages high-performance Update() polling to capture sub-millisecond micro-kinematic digital biomarkers natively in the browser. Furthermore, the integration of a Scikit-Learn Random Forest Classifier profoundly advances Explainable Artificial Intelligence (XAI) within behavioral tracking. By natively extracting Feature Importance via Mean Decrease in Impurity (MDI), the system eradicates the algorithmic "black box" problem, providing users with a precise attribution of the specific cognitive biomarkers driving the AI's prediction.

Finally, the empirical discovery of the "Synthetic Data Gap" validates the extreme sensitivity of the telemetry engine while establishing a clear roadmap for future research. The finding that a model trained on simulated data becomes hyper-sensitive to the chaotic variance of authentic human interaction dictates that future iterations must employ comprehensive in-vivo pilot testing. Capturing ground-truth human telemetry will facilitate vital domain adaptation, ensuring the platform can accurately differentiate between benign environmental noise and genuine cognitive drift in real-world, game-based applications.

REFERENCES

1. Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games, 4(2), e11.
2. Takebe, T., et al. (2025). Gamified N-back app metrics and clinical validation for cognitive impairment. JMA Journal, 10.31662/jmaj.2024-0217.
3. Alharthi, S. A., et al. (2026). ESFY: A Gamified Digital Phenotyping System for Self-Monitoring of Psychological Distress. Proceedings of the ACM on Human-Computer Interaction.
4. Immersive VR and ecological validity in Cognitive Instruments. (2024). MDPI Applied Sciences.
5. Torous, J., et al. (2024). Digital phenotyping in psychiatry: bridging the paradigms of healthcare and data. Journal of Medical Internet Research, 26, e59826.
6. Brefeld, U., et al. (2025). Mouse cursor trajectories and predictive performance in DMHIs. Journal of Medical Internet Research.
7. AdSight: a method leveraging mouse cursor trajectories to quantify user attention. (2025). arXiv preprint arXiv:2505.01451v2.
8. Digital biomarkers for early detection of neurodegenerative diseases. (2024). MDPI Diagnostics.
9. Explainable AI for Depression Detection and Severity Classification From Activity Data. (2025). JMIR Mental Health, 12, e72038.

10. Explainable machine learning framework for predicting depression risk based on behavioral features. (2025). NPJ Digital Medicine.
11. Hudon, A., et al. (2025). A hybrid fuzzy logic-Random Forest model to predict psychiatric treatment order outcomes: an interpretable tool for legal decision support. Frontiers in Artificial Intelligence, 8, 1606250.
12. Interpretable machine learning pipeline for NMDAR encephalitis prognosis. (2025). JMIR Medical Informatics.
13. Ayon, S. S., et al. (2026). Explainable AI framework for improved Thalassemia mental health classification and feature selection. PLoS ONE.
14. The "Synthetic Data Gap" in clinical dialogue generation and domain adaptation. (2025). arXiv preprint arXiv:2504.21800v4.
15. Embedding-Driven Diversity Sampling to Improve Few-Shot Synthetic Data Generation. (2025). ACM Transactions on Computing for Healthcare.
16. Unity Technologies (2025). Unity WebGL Performance Considerations. Unity Documentation, WebGL API guidelines.

Works cited

1. Human-Computer Interaction in Digital Mental Health - MDPI, accessed on March 12, 2026, https://www.mdpi.com/2227-9709/9/1/14
2. Computational Psychiatry for Computers - PMC - NIH, accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC7691174/
3. Modern views of machine learning for precision psychiatry - PMC, accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC9676543/
4. A Gamification Framework for Cognitive Assessment and Cognitive Training: Qualitative Study - PMC, accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC8170558/
5. Examination of Eye-Tracking, Head-Gaze, and Controller-Based Ray-Casting in TMT-VR: Performance and Usability Across Adulthood - MDPI, accessed on March 12, 2026, https://www.mdpi.com/2414-4088/9/8/76
6. A Gamification Framework for Cognitive Assessment and Cognitive Training: Qualitative Study - JMIR Serious Games, accessed on March 12, 2026, https://games.jmir.org/2021/2/e21900/
7. Evaluating the User Experience and Usability of Game-Based Cognitive Assessments for Older People: Systematic Review - JMIR Aging, accessed on March 12, 2026, https://aging.jmir.org/2025/1/e65252
8. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality - PMC, accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12079407/
9. Key Features of Digital Phenotyping for Monitoring Mental Disorders: Systematic Review, accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12588392/
10. Digital devices and continuous telemetry: opportunities for aligning psychiatry and neuroscience - PMC, accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC6224592/
11. UNIVERSITY OF PADOVA Department of General Psychology Master's Degree in Cognitive Neuroscience and Clinical Neuropsychology F, accessed on March 12, 2026, https://thesis.unipd.it/retrieve/622efb9b-1b79-403b-af87-19c588ee95e4/Damrongsit_Krittpoom.pdf

12. The promise and challenges of computer mouse trajectories in ..., accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12017972/
13. Serious Games and Gamification for Mental Health: Current Status and Promising Directions - PMC, accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC5222787/
14. EEG-Based Engagement Monitoring in Cognitive Games - MDPI, accessed on March 12, 2026, https://www.mdpi.com/1424-8220/25/7/2072
15. Explainable AI for Depression Detection and Severity Classification From Activity Data: Development and Evaluation Study of an Interpretable Framework - JMIR Mental Health, accessed on March 12, 2026, https://mental.jmir.org/2025/1/e72038
16. Explainable machine learning for mental health prediction from social media behavior: a nested cross-validation study with SHAP and LIME interpretability - PMC, accessed on March 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12909650/
17. PRECISE Mini: A Browser-Based System for Detecting Micro-Level Cognitive Drift - Medium, accessed on March 12, 2026, https://medium.com/@khoja.aliza/precise-mini-a-browser-based-system-for-detecting-micro-level-cognitive-drift-046617a4da31