A Hybrid AI and Rule-Based Analytics Framework Using Pre-Trained NLP Sentiment Models for Web-Based Mental Well-Being Self-Assessment

Monika M M[1], Bhanupriya V P[2], Diya S Thange[3], Ganavi R D[4] and Rohan D Joel[5]
[1]*Assistant Professor, Department of Computer Science and Engineering, Malnad College of Engineering, Hassan, Karnataka, India*
[2, 3, 4, 5]*Department of Computer Science and Engineering, Malnad College of Engineering, Hassan, Karnataka, India*

***Abstract*—Mental well-being issues such as stress, emotional distress, and mood imbalance are increasingly prevalent, yet many individuals lack access to simple, ethical, and transparent tools for early self-awareness and screening. Existing digital mental health solutions often rely on complex machine learning models, opaque decision-making, or clinical datasets, which may limit accessibility, interpretability, and responsible use for non-diagnostic purposes.**

This paper presents a web-based hybrid mental well-being self-check system that combines rule-based analytics with sentiment analysis using pre-trained Natural Language Processing (NLP) models to provide a structured, interpretable, and user-friendly screening framework. The system employs a multi-domain questionnaire covering mood, anxiety, stress, sleep, energy, and social well-being, where responses are processed through a rule-based scoring mechanism to compute category-wise and overall difficulty indices. Additionally, free-text emotional input provided by the user is analyzed using a pre-trained transformer-based sentiment model to capture linguistic emotional cues. These two components are fused to generate a hybrid assessment that enhances contextual understanding while maintaining transparency and ethical boundaries.

The proposed system delivers visual analytics, including tabular summaries and graphical representations, to support user comprehension and reflection. Experimental demonstrations indicate consistent and interpretable outputs suitable for awareness-level screening. The framework is not intended for diagnosis but serves as an accessible tool for early self-reflection and mental health awareness. The proposed approach highlights the effectiveness of combining rule-based systems with pre-trained NLP models to create responsible, scalable, and explainable digital mental health applications.

***Index Terms*—Artificial Intelligence, Digital Mental Health, Hybrid AI System, Mental Well-being Screening, Natural Language Processing, Rule-based Analytics, Sentiment Analysis, Web-based Self-Assessment**

I. INTRODUCTION

Mental health and emotional well-being have emerged as critical public health concerns worldwide. According to global health organizations, a significant proportion of individuals experience stress, anxiety, depression, or emotional distress at some stage of their lives, yet many do not seek timely support due to stigma, lack of awareness, limited access to professionals, or fear of medical labeling. Early self-reflection and awareness are therefore essential in identifying emotional difficulties before they escalate into severe mental health conditions [1], [2].

With the rapid growth of digital technologies, several online mental health platforms and mobile applications have been developed to provide screening, counseling, or diagnostic support. However, many existing solutions either rely on complex clinical assessments, require professional supervision, or directly attempt automated diagnosis using machine learning models. Such approaches raise ethical concerns, including misinterpretation of results, over-diagnosis, lack of transparency, and potential psychological harm to users. Additionally, fully data-driven machine learning systems often behave as "black boxes," making their outputs difficult for users to understand or trust [3]–[6].

Recent advancements in Natural Language Processing (NLP) and sentiment analysis have enabled systems to analyze free-text emotional expressions using pre-trained deep learning models. While these models demonstrate strong performance in identifying sentiment polarity, they are typically trained on general-purpose datasets and are not designed for clinical diagnosis. When used responsibly, however, such models can enhance non-clinical mental health screening by capturing emotional tone and subjective expression without replacing professional judgment [7]–[10].

Despite these advancements, there remains a clear need for ethical, transparent, and accessible self-check systems that focus on awareness rather than diagnosis. Such systems should provide understandable feedback, minimize risks associated with misclassification, and support users in recognizing when professional help may be beneficial. Rule-based approaches, when combined with AI techniques, offer an effective solution by maintaining interpretability while still benefiting from intelligent text analysis.

In this work, we propose a hybrid web-based mental well-being self-check framework that integrates rule-based analytics with NLP-based sentiment analysis using pre-trained transformer models. The system employs a structured questionnaire to evaluate multiple well-being dimensions and complements this assessment with sentiment analysis of user-provided emotional text. Rather than providing a diagnosis, the proposed framework generates difficulty indicators, visual analytics, and interpretive feedback aimed at self-awareness and early reflection. The key contribution of this study lies in demonstrating a practical, ethical, and interpretable hybrid approach that balances AI capabilities with rule-based decision logic, making it suitable for non-clinical mental health awareness applications.

II. PROBLEM STATEMENT

Mental health and emotional well-being challenges affect individuals across all age groups and socio-economic backgrounds, yet a large proportion of affected individuals do not seek timely support due to stigma, lack of awareness, limited access to mental health professionals, or fear of being clinically labeled [1], [2]. While early identification and self-reflection can play a crucial role in preventing the escalation of emotional distress into severe mental health conditions, accessible and user-friendly self-check mechanisms remain limited.

In recent years, numerous digital mental health platforms and mobile applications have emerged, offering services such as psychological screening, counseling support, and automated assessments [3], [4]. However, many of these systems rely heavily on clinically validated diagnostic scales, professional supervision, or fully data-driven machine learning models. Such approaches often introduce significant challenges, including reduced transparency, limited interpretability, and ethical concerns related to misclassification, over-diagnosis, and user anxiety [5], [6]. In particular, black-box machine learning models may produce outputs that are difficult for users to understand or trust, especially when applied to sensitive domains such as mental health.

Furthermore, existing AI-driven mental health tools frequently attempt to infer diagnostic outcomes directly from user data, despite being trained on generalized or non-clinical datasets [7], [8]. This raises concerns regarding reliability, ethical deployment, and responsible use, as such systems may unintentionally replace professional judgment or provide misleading conclusions. As a result, there is a growing need to clearly distinguish between clinical diagnostic tools and non-clinical awareness-oriented systems.

Therefore, the core problem addressed in this work is the absence of simple, transparent, and ethically designed mental well-being self-check systems that focus on awareness rather than diagnosis. Current solutions either over-depend on complex clinical methodologies or employ opaque AI models without sufficient interpretability. There is a lack of lightweight, web-based frameworks that integrate structured rule-based evaluation with AI-assisted sentiment understanding while maintaining user trust, ethical boundaries, and ease of access.

This work aims to address this gap by proposing a non-diagnostic, awareness-oriented hybrid framework that combines interpretable rule-based analytics with pre-trained NLP sentiment analysis models. The proposed system seeks to provide meaningful self-reflection indicators and visual feedback without making medical claims, thereby supporting early awareness while respecting ethical

and practical constraints of digital mental health applications.

III. RELATED WORK

Mental health screening and assessment have been extensively studied across clinical, digital, and computational domains. Traditional mental health assessment methods rely on clinician-administered questionnaires, interviews, and standardized diagnostic criteria. While clinically reliable, these approaches are often resource-intensive and inaccessible to a large portion of the population, particularly in low-resource settings [1], [2]. This has motivated the development of digital mental health tools aimed at improving accessibility and early identification of emotional distress.

Recent years have witnessed a rapid increase in mobile and web-based mental health applications designed to support screening, monitoring, and self-management. Studies have shown that such tools can enhance engagement and provide scalable support; however, many lack strong empirical validation and clinical oversight [4]. Furthermore, several applications attempt to infer mental health conditions directly using machine learning models, which raises ethical concerns related to misdiagnosis, lack of transparency, and potential psychological harm [10], [11], [15]. These limitations highlight the need for cautious deployment of AI-driven mental health technologies.

From a computational perspective, Natural Language Processing (NLP) and sentiment analysis have become prominent techniques for analyzing emotional content in textual data. Extensive surveys have demonstrated the effectiveness of sentiment and emotion detection methods in capturing affective states from text [5], [6]. Transformer-based architectures such as BERT [7] and its distilled variants like DistilBERT [8] have significantly improved performance in sentiment classification tasks and are widely adopted through frameworks such as HuggingFace Transformers [9]. However, these models are generally trained on large-scale, general-purpose corpora and are not designed for clinical diagnosis, limiting their direct applicability in mental healthcare contexts [10].

To address interpretability and reliability concerns, rule-based expert systems have long been employed in healthcare decision support. Such systems use predefined logical rules to guide decision-making and are valued for their transparency and explainability [13], [14]. In mental health applications, rule-based approaches enable clear mapping between user responses and system outputs, reducing ambiguity and increasing user trust. Nonetheless, purely rule-based systems may fail to capture nuanced emotional expression, particularly in unstructured text data.

Several studies have suggested that hybrid approaches combining rule-based logic with machine learning techniques can balance interpretability and analytical capability [3], [11]. By integrating structured assessment methods with NLP-based sentiment analysis, hybrid systems can provide richer insights while avoiding the risks associated with fully automated diagnosis. Despite this potential, relatively few existing systems focus explicitly on non-clinical, awareness-oriented mental well-being self-check frameworks that prioritize ethical considerations and user transparency.

A. *Key Gap Identified*

While prior work has explored digital mental health tools, NLP-based sentiment analysis, and rule-based decision support independently, there is a lack of lightweight, ethical, and interpretable hybrid frameworks designed specifically for non-diagnostic mental well-being self-assessment. This gap motivates the development of the proposed hybrid system, which integrates rule-based analytics with pre-trained NLP sentiment models to support mental health awareness rather than clinical diagnosis.

IV. SYSTEM ARCHITECTURE

The proposed system follows a web-based, modular architecture designed to support ethical, non-clinical mental well-being self-checks. The architecture integrates rule-based analytics with NLP-based sentiment analysis in a hybrid framework, ensuring transparency, interpretability, and accessibility. The system is implemented using a lightweight web framework and pre-trained NLP models, making it suitable for real-time user interaction without complex infrastructure.

A. Architecture Components

1. Web Interface (Presentation Layer)

The web interface provides the primary interaction point between the user and the system. It consists of multiple pages including consent, questionnaire, free-text input, results, and history. This layer is responsible for collecting structured questionnaire responses and unstructured textual input, and for presenting analytical results and visualizations in an understandable manner.

2. Questionnaire Module (Rule-Based Input Module)
This module presents a structured set of multiple-choice questions covering key mental well-being dimensions such as mood, stress, anxiety, sleep, energy, and social interaction. User responses are numerically encoded and passed to the scoring engine. The rule-based nature of this module ensures interpretability and avoids opaque decision-making, aligning with established principles of transparent decision support systems [13], [14].

.3. NLP Sentiment Analysis Module
The NLP module processes user-provided free-text emotional descriptions. It employs a pre-trained transformer-based sentiment analysis model (e.g., DistilBERT via HuggingFace Transformers) to infer sentiment polarity and confidence scores. As the model is pre-trained on general-purpose datasets, it is used strictly for non-clinical sentiment interpretation, not diagnosis. This approach leverages the strengths of modern NLP while maintaining ethical boundaries [5]–[9].

4. Hybrid Scoring Engine
The hybrid scoring engine serves as the core analytical component of the system. It combines:

- Rule-based percentage scores derived from questionnaire responses, and
- Sentiment-derived difficulty indicators from the NLP module

These outputs are aggregated to compute both:

- a rule-based overall difficulty score, and
- a hybrid overall difficulty score that incorporates textual sentiment.

This hybrid design balances interpretability with expressive power and aligns with prior research advocating combined rule-based and machine learning approaches in healthcare analytics [3], [11].

5. Visualization and Reporting Module
The visualization module transforms computed scores into graphical representations such as bar charts, radar charts, and trend plots. These visualizations help users intuitively understand their mental well-being profile and changes over time. Results are accompanied by non-diagnostic interpretive text and disclaimers to reinforce ethical usage.

6. Data Storage Module
Assessment results are stored in a lightweight persistent format (CSV) to enable historical analysis and trend visualization. This design choice avoids unnecessary database complexity while supporting longitudinal self-reflection. Stored data does not include sensitive identifiers, supporting privacy-aware design principles [15].

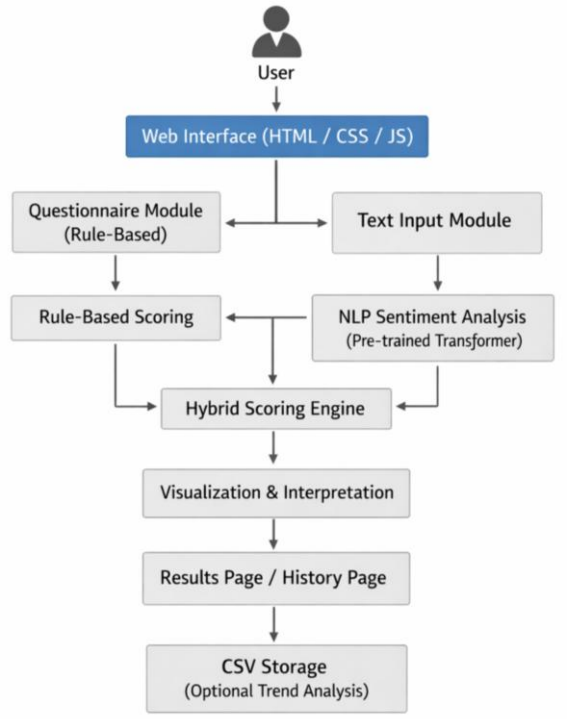

Fig. 1. System Architecture of the Proposed Hybrid Mental Well-being Self-Check Framework.

The architecture emphasizes modularity, ethical design, and interpretability by clearly separating the rule-based questionnaire logic from the NLP-based sentiment analysis module, while integrating both through a hybrid scoring engine to support non-clinical mental health awareness.

V. METHODOLOGY

The proposed system adopts a hybrid methodology that integrates rule-based analytics with Natural Language Processing (NLP)-based sentiment analysis to support non-clinical mental well-being self-assessment. The methodology is designed to ensure interpretability, ethical usage, and accessibility while avoiding automated diagnosis. The overall process consists of five main components: rule-based questionnaire design, scoring and threshold logic, NLP sentiment analysis, hybrid score computation, and ethical safeguards.

A. Rule-Based Questionnaire Design

The foundation of the proposed framework is a structured questionnaire designed to capture self-reported indicators of mental well-being across multiple dimensions, including mood, anxiety, stress, sleep, energy, and social functioning. Each dimension is assessed using a fixed set of closed-ended questions with standardized response options ranging from "Never" to "Almost always." This design follows principles commonly used in mental health screening tools, where structured self-reporting enables consistent data collection while remaining easy for users to understand and complete [1], [2].

The questionnaire is implemented as a rule-based module, ensuring that each response is directly mapped to a predefined numerical value. This approach prioritizes transparency and interpretability, allowing users and evaluators to clearly understand how inputs influence the resulting scores.

B. Scoring and Threshold Logic

Each questionnaire response is assigned a numerical score based on response frequency. Scores within each mental well-being category are aggregated and normalized to generate percentage-based difficulty indicators. Threshold-based rules are then applied to classify each category into qualitative levels such as low, moderate, or high concern.

Rule-based thresholding has been widely used in healthcare decision support systems due to its explainability and deterministic behavior [13], [14]. Unlike black-box machine learning models, this logic ensures that identical inputs always produce identical outputs, reducing ambiguity and enhancing user trust.

C. NLP Sentiment Analysis Using Pre-Trained Models

To complement structured questionnaire data, the system incorporates NLP-based sentiment analysis of free-text user input describing emotional experiences. A pre-trained transformer-based sentiment analysis model is employed to identify the polarity and confidence level of emotional expression. Transformer architectures such as BERT and its distilled variants have demonstrated strong performance in sentiment classification tasks and are widely used in modern NLP systems [7], [8], [9].

The use of a pre-trained model eliminates the need for task-specific training data while leveraging state-of-the-art language representations. However, the sentiment output is explicitly treated as a supportive indicator rather than a diagnostic signal, acknowledging that such models are trained on general-purpose datasets and are not intended for clinical inference [5], [10].

D. Hybrid Score Computation

The final system output is generated through a hybrid scoring mechanism that combines rule-based questionnaire results with NLP-derived sentiment indicators. Questionnaire-based difficulty scores and text-based sentiment difficulty scores are normalized and aggregated to compute an overall hybrid mental well-being index.

Hybrid approaches that combine symbolic rule-based reasoning with data-driven AI techniques have been shown to balance interpretability and analytical capability in healthcare applications [3], [11]. In the proposed framework, rule-based analytics serve as the primary assessment mechanism, while NLP sentiment analysis provides supplementary contextual insight, ensuring that the system remains explainable and ethically constrained.

E. Ethical Safeguards

Ethical considerations are integral to the proposed methodology. The system explicitly avoids diagnostic claims and includes clear disclaimers indicating that results are intended solely for self-awareness and reflection. User consent is obtained prior to assessment, and no clinical labels are assigned at any stage of processing.

Additionally, the framework incorporates safeguards against misuse, such as detection of inconsistent

response patterns and transparency in score computation. These measures align with ethical guidelines for digital mental health tools, which emphasize user safety, interpretability, and responsible AI deployment [4], [15].

VI. IMPLEMENTATION

The proposed mental well-being self-check system is implemented as a lightweight web-based application to demonstrate practical feasibility and ease of deployment. The backend of the system is developed using Python, chosen for its extensive ecosystem of libraries for web development, data processing, and artificial intelligence. The Flask microframework is employed to manage HTTP routing, session handling, and interaction between frontend and backend components, enabling a modular and maintainable architecture.

The Natural Language Processing (NLP) component utilizes the HuggingFace Transformers library, which provides access to state-of-the-art pre-trained transformer models. Specifically, a sentiment analysis pipeline based on a distilled transformer model is used to analyze free-text emotional input. The model is loaded once at application startup and used in inference mode, ensuring efficient CPU-based execution without the need for model retraining [7]–[9]. This design choice aligns with ethical and practical constraints, as the model is not adapted for clinical diagnosis.

User responses from the structured questionnaire and derived scores are stored in a CSV file, providing a simple and transparent data persistence mechanism suitable for prototyping and experimental evaluation. This approach avoids the overhead of database management while enabling longitudinal analysis of results. The frontend is implemented using HTML and CSS for layout and styling, while Chart.js is used to generate interactive bar, radar, and trend visualizations directly in the browser. These visual elements support user interpretability and align with the system's awareness-oriented objectives.

VII. EXPERIMENTAL EVALUATION

An experimental evaluation was conducted to demonstrate the functional behavior and practical applicability of the proposed hybrid framework. As the system is designed for non-clinical awareness rather than diagnostic accuracy, the evaluation focuses on behavioral validation rather than medical performance metrics.

To assess the sentiment analysis component, a small controlled experiment was performed using ten manually curated text samples representing clearly positive and negative emotional expressions. Each sample was assigned an expected sentiment label by human judgment and then processed by the pre-trained sentiment model. In this limited experiment, the model correctly classified all samples, demonstrating consistent sentiment detection for unambiguous emotional text. This experiment is intended solely to illustrate model behavior and not to claim generalizable accuracy, consistent with findings reported in prior studies on sentiment analysis limitations in mental health contexts [5], [6], [10].

In addition to sentiment testing, multiple end-to-end system runs were conducted using different questionnaire response patterns and free-text inputs. The outputs included category-wise difficulty percentages, overall rule-based and hybrid scores, and corresponding visualizations. These results confirmed that the hybrid scoring mechanism correctly integrates structured questionnaire data with NLP-derived sentiment indicators, producing interpretable outputs aligned with user inputs.

All experiments were executed using CPU-based inference, demonstrating that the system can operate efficiently without specialized hardware. However, the evaluation has notable limitations: the absence of clinically validated datasets, the small size of the sentiment test set, and the reliance on self-reported user input. These limitations are acknowledged to avoid overinterpretation of results and to maintain ethical integrity [10], [11].

VIII. RESULTS AND DISCUSSION

The proposed hybrid mental well-being self-check system was evaluated through multiple test runs using sample questionnaire responses and free-text emotional inputs. The system generates category-wise difficulty percentages across six well-being dimensions—Mood, Anxiety, Stress, Sleep, Energy, and Social—and integrates these with an NLP-based

sentiment difficulty score derived from user-provided textual input.

A. Sample Outputs

The system outputs include a structured results table displaying difficulty percentages and corresponding concern levels (Low, Moderate, High) for each category. Additionally, two overall indicators are computed: (i) an overall rule-based difficulty score derived solely from questionnaire responses, and (ii) a hybrid difficulty score that incorporates both questionnaire-based analytics and NLP sentiment analysis.

B. Category-wise Difficulty Analysis

Experimental outputs demonstrate that the system effectively differentiates between well-being domains. For example, high difficulty scores in Mood, Stress, and Energy categories were observed in cases where users reported persistent emotional distress, while Sleep-related difficulty remained low when users indicated minimal sleep disruption. This domain-level granularity enables targeted self-reflection rather than a single aggregated label.

C. Hybrid vs. Rule-Based Comparison

A consistent observation across test cases was that the hybrid difficulty score slightly differed from the rule-based score, reflecting the influence of free-text sentiment analysis. In emotionally expressive inputs containing strong negative affect, the NLP component increased the overall difficulty indicator, demonstrating the added value of unstructured emotional expression beyond fixed-choice responses. This comparison highlights the complementary nature of combining rule-based logic with NLP-based analysis.

D. Interpretation and Discussion

Importantly, the system does not attempt to diagnose any mental health condition. Instead, results are presented as non-clinical difficulty indicators accompanied by interpretive guidance encouraging self-awareness and, where appropriate, professional consultation. Visualization through bar and radar charts further enhances interpretability by allowing users to visually inspect their emotional well-being profile.

IX. ETHICAL CONSIDERATIONS AND LIMITATIONS

The development and deployment of digital mental health technologies require careful ethical consideration due to the sensitive nature of psychological data and the potential impact on users' well-being. The proposed system is explicitly designed as a non-diagnostic, awareness-oriented self-check tool, and several safeguards are incorporated to minimize ethical risks.

A. Non-Diagnostic Design

The system does not provide clinical diagnoses, medical labels, or treatment recommendations. All outputs are framed as difficulty indicators for self-reflection, and clear disclaimers are presented to users stating that the system does not replace professional mental health care. This design choice reduces the risk of misinterpretation, over-diagnosis, and psychological harm.

B. Self-Reported Data Limitations

The system relies entirely on self-reported questionnaire responses and user-provided text input. Such data may be influenced by response bias, misunderstanding of questions, or intentional misreporting. Consequently, the generated results reflect perceived emotional states rather than objectively verified mental health conditions.

C. *Pre-Trained NLP Model Constraints*

The sentiment analysis component utilizes a pre-trained transformer-based model trained on general-purpose text corpora. While effective at identifying sentiment polarity, the model is not trained on clinical mental health datasets and may misinterpret sarcasm, ambiguous language, or culturally specific expressions. For this reason, NLP outputs are used only as supportive indicators within a hybrid framework rather than as standalone predictors.

D. User Privacy and Consent

User interaction with the system is preceded by an explicit consent step outlining the purpose, limitations, and non-clinical nature of the tool. No personally identifiable information is required, and stored results are limited to anonymized analytical data. These measures align with ethical best practices

for digital mental health applications and support responsible use.

X. FUTURE SCOPE

While the proposed hybrid mental well-being self-check framework demonstrates practical feasibility and ethical applicability for awareness-oriented screening, several directions remain open for future enhancement. First, large-scale validation studies involving diverse user populations could be conducted to assess robustness and usability across different demographic and cultural contexts. Such validation, when performed under ethical supervision, would strengthen confidence in the framework's applicability as a non-clinical screening aid.

Second, the use of domain-specific NLP models trained on mental health–related text corpora may improve the sensitivity of sentiment and emotion detection while preserving ethical constraints. These models could better capture nuanced emotional expressions that generic sentiment models may overlook. Third, long-term trend analysis can be incorporated by leveraging longitudinal user data to visualize changes in well-being over time, enabling users to reflect on progress or deterioration patterns in a responsible manner.

Finally, the framework can be extended to mobile and cloud-based deployments to increase accessibility and scalability. Integration with mobile platforms would allow wider adoption while maintaining the system's core principles of transparency, non-diagnostic intent, and user consent.

XI. CONCLUSION

This work addressed the growing need for simple, transparent, and ethical mental well-being self-check tools that prioritize awareness over diagnosis. A hybrid web-based framework was proposed, integrating rule-based questionnaire analytics with NLP-based sentiment analysis using pre-trained transformer models. By combining structured self-report assessment with free-text emotional analysis, the system provides interpretable difficulty indicators and visual feedback without attempting clinical diagnosis.

The proposed approach demonstrates that hybrid AI systems can effectively balance interpretability and analytical capability, mitigating ethical concerns commonly associated with fully automated machine learning–based mental health tools. Experimental evaluation confirmed the feasibility of the approach using lightweight, CPU-based inference and pre-trained models, while emphasizing the non-clinical nature of the results.

Overall, the framework contributes a practical, responsible, and extensible solution for mental well-being self-reflection. When used appropriately, such systems can support early awareness, encourage help-seeking behavior, and complement—rather than replace—professional mental health care.

REFERENCES

[1] World Health Organization, *Mental Health: Strengthening Our Response*, WHO, Geneva, Switzerland, 2023. [Online]. Available: https://www.who.int

[2] World Health Organization, *Depression and Other Common Mental Disorders: Global Health Estimates*, WHO, Geneva, Switzerland, 2017.

[3] T. B. Murdoch and A. S. Detsky, "The inevitable application of big data to health care," *JAMA*, vol. 309, no. 13, pp. 1351–1352, 2013.

[4] J. Torous, A. Firth, J. Huckvale, and K. Larsen, "Mental health mobile applications: From innovation to evidence," *World Psychiatry*, vol. 17, no. 3, pp. 279–280, 2018.

[5] P. C. Roy, S. Bandyopadhyay, and A. Chakraborty, "Emotion detection from text: A survey," *IEEE Trans. Affective Computing*, vol. 11, no. 2, pp. 232–246, 2020.

[6] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 102–107, 2017.

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, Minneapolis, MN, USA, 2019, pp. 4171–4186.

[8] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT:

Smaller, faster, cheaper and lighter," in *Proc. NeurIPS Workshop*, Vancouver, Canada, 2019.

[9] T. Wolf *et al.*, "Transformers: State-of-the-art natural language processing," in *Proc. EMNLP: System Demonstrations*, Online, 2020, pp. 38–45.

[10] A. Miner, L. Haque, J. A. Fries, and J. J. Swendsen, "Assessing the accuracy of social media–based mental health predictions," *npj Digital Medicine*, vol. 2, no. 1, pp. 1–8, 2019.

[11] S. Shatte, D. Hutchinson, and S. Teague, "Machine learning in mental health: A scoping review," *Psychological Medicine*, vol. 49, no. 9, pp. 1426–1448, 2019.

[12] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., Cambridge, MA, USA: MIT Press, 2018.

[13] M. Shortliffe and J. Cimino, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, 4th ed., New York, NY, USA: Springer, 2014.

[14] E. H. Shortliffe, "Rule-based decision support systems," *IEEE Computer*, vol. 35, no. 11, pp. 55–63, Nov. 2002.

[15] J. Huckvale, J. Torous, and E. Larsen, "Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation," *JAMA Network Open*, vol. 2, no. 4, pp. 1–10, 2019.