Foundations for AI-Driven Enhancements in E-CBT

Najare Johnson
DREAM 2024 Final Paper
University of South Florida

Abstract

Suicide prevention remains a pressing challenge in mental health care, with barriers to timely intervention and personalized treatment often hindering effective outcomes. Electronic Cognitive Behavioral Therapy (E-CBT) platforms offer scalable and accessible mental health support but typically lack integrated mechanisms for real-time suicidality risk assessment without clinician involvement. This study investigates the relationship between patient engagement metrics—such as login frequency, time spent, and lesson reviews—and changes in depression scores on an E-CBT platform implemented within a West Virginia primary care network.

By analyzing correlations between engagement patterns and standardized depression measures (PHQ-9 and PHQ-2), the research highlights key insights into the optimization of digital mental health interventions. A comparative analysis of two groups—Full Coaching (Group A) and Reminder-Only interventions (Group B)—reveals the impact of varying levels of human support on treatment outcomes. Group A exhibited superior reductions in depression scores, emphasizing the potential benefits of integrated human support.

The findings demonstrate modest but significant correlations between engagement metrics and improvements in PHQ-2 scores, while PHQ-9 outcomes were less strongly associated. This suggests that brief depression screening tools may be more sensitive to changes driven by patient engagement and better equipped to distinguish changes for less acute patients since the PHQ-9 was only given to clients that reported high suicidal ideation scores on the PHQ-2.. The study also identifies program retention as a critical challenge, with a substantial attrition rate highlighting the need for strategies to sustain participation.

This research addresses gaps in understanding the dynamic relationship between user interaction and clinical outcomes in E-CBT platforms. It provides foundational insights for enhancing digital mental health interventions through AI-assisted tools and personalized engagement strategies. Future work will explore natural language processing, predictive modeling, and the integration of human-AI collaboration to advance the field of digital mental health care, with a focus on scalability and usability in underserved regions.

1. Introduction

Suicide prevention remains a critical challenge in mental health care, exacerbated by barriers to timely intervention and personalized treatment. While E-CBT(Electronic Cognitive Behavioral Therapy) platforms provide accessible mental health support, they often lack mechanisms to assess suicidality risk without the support of a licensed clinician.

Current approaches to depression screening and suicide risk assessment primarily rely on standardized tools such as the PHQ-9 and PHQ-2 (Patient Health Questionairre), which provide valuable but static snapshots of patient mental health status. These instruments, while validated and widely used, may not fully capture the dynamic nature of depression symptoms or effectively leverage additional data generated through patient platform interactions. The integration of engagement metrics with standardized assessments presents an opportunity to enhance our understanding of treatment effectiveness and patient progress.

This study analyzes patient interaction data from an E-CBT platform implemented in a West Virginia primary care network, examining correlations between program engagement metrics (including login frequency, time spent, and amount of lesson reviews) and changes in depression scores. By exploring these relationships, we aim to establish foundational insights into how patient engagement patterns might inform treatment optimization and allow for less specialized mental health professionals to assist with risk assessments. Our analysis focuses particularly on comparing outcomes between full coaching (Group A) support and reminder-only interventions (Group B), addressing the crucial question of how varying levels of human support influence treatment effectiveness.

The research addresses three key gaps in current digital mental health interventions: (1) the limited understanding of specific engagement metrics that correlate with clinical outcomes, and (2) the potential for leveraging patient-generated data to enhance treatment personalization. By examining these relationships in a real-world implementation, this study provides crucial insights for the development of more AI assisted digital mental health platforms for usability with mental health professionals of varying degree.

This investigation represents an initial step toward developing more sophisticated, data-driven approaches to digital mental health intervention. Our findings aim to inform future developments in automated assessment and personalized treatment delivery, while maintaining the accessibility and scalability that make digital interventions particularly valuable in underserved regions.

The rest of this paper is organized as follows: Section 2 reviews current literature on digital mental health interventions and AI applications in mental health care, with particular attention to suicide risk assessment tools and E-CBT platforms. Section 3 details our methodology for analyzing patient interaction data and depression outcomes, including our approach to data preprocessing and statistical analysis. Section 4 presents our findings on the relationships between program engagement metrics and mental health outcomes, comparing results between full coaching and reminder-only interventions. Section 5 concludes with a discussion of our findings' implications and outlines directions for future research in AI-supported digital mental health interventions.

2. Related Work

Digital mental health interventions, particularly E-CBT platforms, have emerged as a promising solution to address the growing demand for mental health services. Several studies have explored their effectiveness and implementation challenges. Kambeitz-Ilankovic et al. (2022) demonstrated that computerized CBT could achieve comparable outcomes to face-to-face therapy for depression, though noting significant variability in engagement and completion rates. Their work highlighted the potential of digital interventions while acknowledging the critical challenge of patient retention.

Research on digital mental health platforms has consistently identified engagement as a key factor in treatment outcomes. Gan et al. (2021) and Boucher and Raiker (2024) found that user engagement metrics, such as login frequency and module completion, significantly predicted treatment success in digital mental health interventions. Similarly, Werntz et al. (2023) explored the role of human support in digital interventions, finding that coached interventions typically showed slightly better outcomes to fully automated approaches. Due to the scant nature of this research our approach will highlight the effects of the varied groups of full coaching versus reminder-only support.

Current approaches to digital suicide risk assessment heavily rely on standardized screening tools. The work of Na et al. (2018) demonstrated the predictive validity of the PHQ-9 in identifying suicide risk through electronic health records. However, their research also highlighted the assessment's limitations in capturing risk factors in certain subgroups. Levis et. al. (2020) advanced this field by exploring natural language processing applications in suicide risk assessment, though their work primarily focused on clinical notes rather than patient-generated content.

The integration of AI in mental health care represents a growing field of research. Liu et al. (2022) developed and evaluated an AI-driven chatbot for delivering CBT, showing promising results for mild to moderate depression. However, their system, like many others, relied primarily on structured interactions rather than analyzing free-text patient communications.

Research seldom emphasizes the importance of considering cultural and regional factors in digital mental health interventions, particularly in underserved areas like rural Appalachia. This research informs our approach to analyzing data from West Virginia participants and understanding engagement patterns in this specific context.

4. Results

Sample Characteristics and Data Preprocessing:

The study analyzed data from participants recruited through a West Virginia primary care facility and affiliated schools. Participants were adults (aged 18 and older) who provided informed consent for program engagement and self-reported outcomes. From an initial dataset of 787 participants, 297 completed both baseline and follow-up assessments over an average enrollment period of three months. The substantial reduction in sample size (62.3%) was primarily due to missing endpoint PHQ-9 scores ($n = 490$) and PHQ-2 scores ($n = 147$). These participants either elected to opt out of the program or did not engage; therefore, the suicide risk assessments provided at later intervals were not complete. Due to privacy considerations, participant data was stored using anonymous identifiers, precluding detailed demographic analysis.

Program Engagement and Depression Outcomes

Analysis of participant engagement revealed slight but statistically significant associations with changes in PHQ-2 scores. The number of program logins demonstrated a weak positive correlation with PHQ-2 score improvements ($r = 0.14$, $p = 0.015$), as did the number of completed reviews ($r = 0.16$, $p = 0.007$). Time spent in the program showed a non-significant trend toward improvement ($r = 0.10$, $p = 0.082$). However, none of the engagement metrics showed significant correlations with PHQ-9 score changes: number of logins ($r = 0.08$, $p = 0.186$), time spent ($r = 0.02$, $p = 0.726$), and completed reviews ($r = 0.01$, $p = 0.912$).

Intervention Group Comparison

Participants were assigned to either Full Coaching (Group A) or Reminders Only (Group B) conditions. Group A demonstrated greater improvements in depression scores compared to Group B across both measures. The average reduction in PHQ-9 scores was more apparent for Group A (-2.45 points) versus Group B (-2.05 points). Similarly, PHQ-2 scores showed larger reductions in Group A (-1.51 points) compared to Group B (-1.19 points).

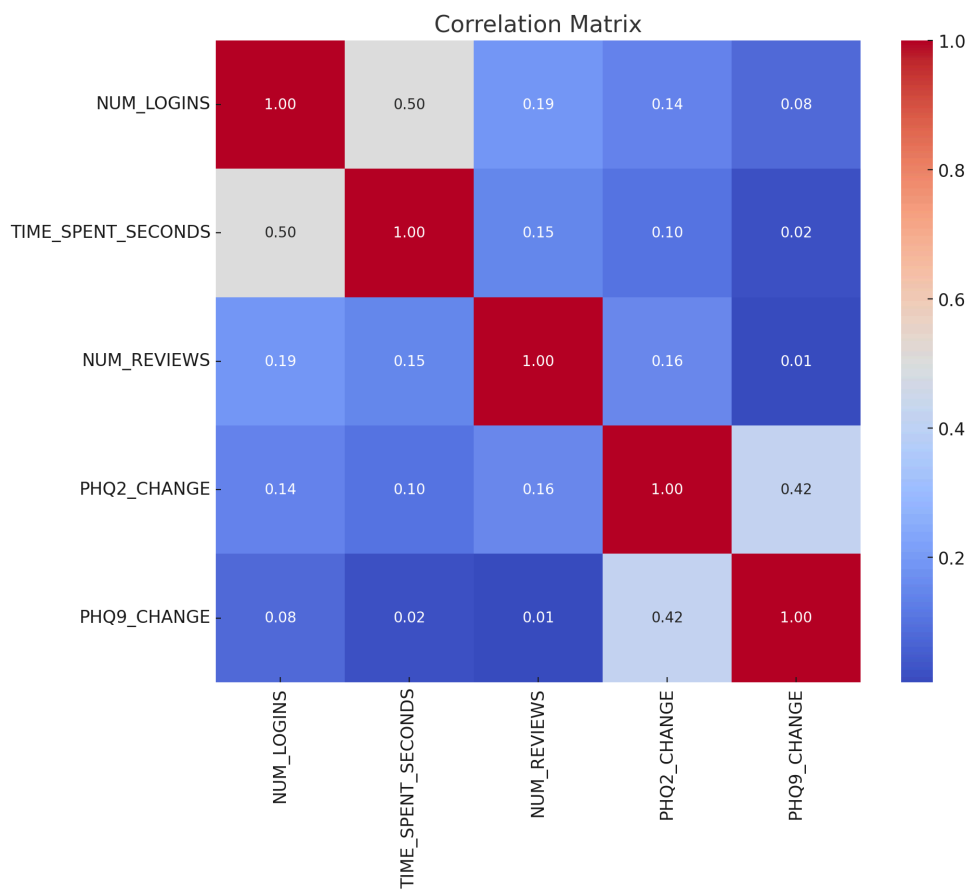
Predictive Modeling and Score Relationships

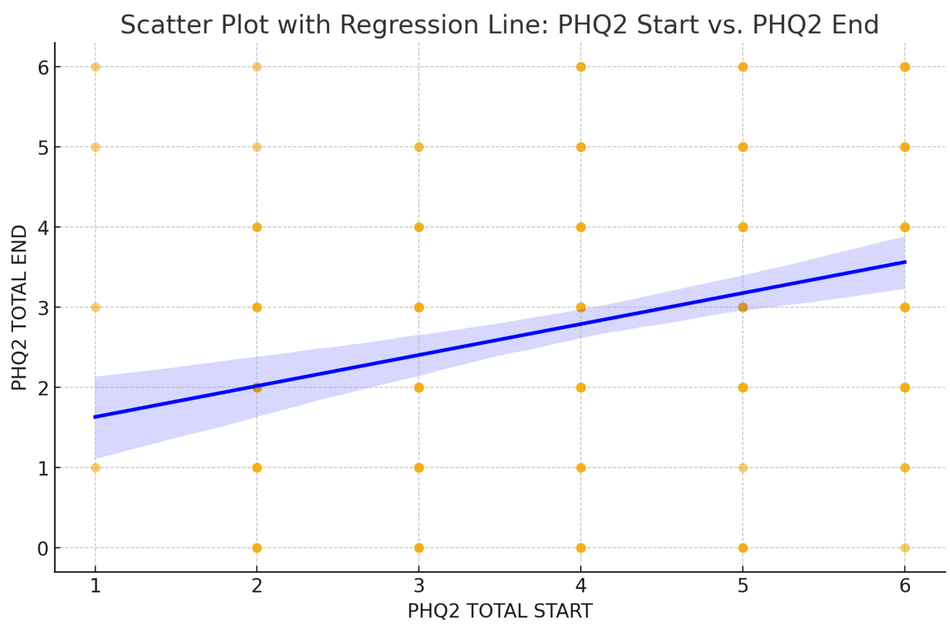
Linear regression analysis revealed significant relationships between initial and endpoint scores for both risk assessment measures. The PHQ-9 scores showed a strong correlation between baseline and endpoint measurements ($p \approx 7.34e\text{-}22$), indicating that initial severity was predictive of final scores. Similarly, PHQ-2 scores demonstrated a significant relationship between baseline and endpoint measurements ($p \approx 3.20e\text{-}08$), though with substantial individual variability in both measures.

A multiple regression model incorporating the initial PHQ-9 scores, the initial PHQ-2 scores, and intervention group assignment was developed to predict endpoint PHQ-9 scores. The model's performance metrics and detailed regression coefficients demonstrate the relative contribution of each predictor variable to final depression scores, though individual response patterns showed considerable variation (Figure 1).

Program Retention The completion rate of 37.7% (297/787) suggests substantial attrition over the three-month intervention period. This retention rate provides important context for interpreting the effectiveness of the intervention and highlights the challenges of maintaining participant engagement in digital mental health programs.

These results suggest that while both intervention conditions were associated with improvements in depression scores, the Full Coaching condition (Group A) demonstrated superior outcomes. The significant correlations between program engagement metrics and PHQ-2 score changes, albeit modest, indicate that higher levels of participation may contribute to better outcomes, particularly as measured by brief depression screening tools.





5. Conclusions and Future Work

Future Works

Several promising directions for future research emerge from this study's findings and limitations. First, the development and validation of AI-driven natural language processing models specifically tailored to analyze patient-generated free text in E-CBT platforms represents a critical next step. These models should be designed to detect subtle linguistic patterns that may indicate changes in suicide risk while accounting for cultural and regional variations specific to populations like those in West Virginia.

The integration of the Integrated Motivational Volitional (IMV) model into automated analysis systems presents another important avenue for research. Future studies should examine how to systematically incorporate this theoretical framework into NLP algorithms while maintaining transparency and clinical interpretability. This integration should include methods to identify and track motivational and volitional phase markers in patient communications.

Future research should also explore the development of explainable AI systems that can provide clinicians with clear rationales for risk assessments and treatment recommendations. This work should focus on creating interpretable models that maintain transparency while protecting patient privacy and confidentiality.

Additional research is needed to investigate the optimal integration of automated systems with human clinical judgment. This includes studying:

- The impact of AI-supported decisions on clinical workflows
- Methods for effectively presenting AI-generated insights to clinicians
- Strategies for maintaining appropriate human oversight while maximizing the benefits of automation

These future directions aim to address the current limitations identified in our study while advancing the field toward more effective, equitable, and ethical implementation of AI in mental health care.

Conclusion

This study addressed a critical gap in digital mental health interventions by examining how patient interaction patterns within E-CBT platforms correlate with depression outcomes and program engagement. Through analysis of participant data from a West Virginia primary care network, we demonstrated that engagement metrics, particularly program logins and completed reviews, show meaningful associations with improvements in depression scores as measured by the PHQ-2.

Our analytical approach provides a foundation for understanding how patient interaction data can inform the development of more responsive and personalized digital mental health interventions. By identifying correlations between program engagement and outcome measures, while accounting for intervention type (Full Coaching vs. Reminders Only), we have established baseline metrics for evaluating E-CBT platform effectiveness.

The findings are particularly valuable given the pressing need for accessible mental health care in underserved regions. Our results demonstrate that digital interventions with varying levels of support can produce meaningful improvements in depression scores, with enhanced outcomes observed in the Full Coaching condition. This suggests that while automated systems can provide benefit, the integration of human support elements may optimize treatment outcomes.

For healthcare providers and researchers, this work offers practical insights into the relationship between digital engagement patterns and mental health outcomes. The identification of specific engagement metrics that correlate with improved outcomes provides actionable targets for intervention refinement and patient monitoring.

Moving forward, this research opens several promising avenues for development. Future work can work to develop predictive models that can identify early warning signs of suicidal ideation based on engagement patterns. The research could also expand to investigate the potential for AI-supported personalization of intervention content based on individual engagement patterns as methods to improve user engagement emerge. These next steps aim to further advance our understanding of how digital mental health platforms can be optimized to provide more effective, personalized care while maintaining accessibility and scalability.

References

Boucher, E.M., Raiker, J.S. Engagement and retention in digital mental health interventions: a narrative review. *BMC Digit Health* 2, 52 (2024). https://doi.org/10.1186/s44247-024-00105-9

Gan, D. Z. Q., McGillivray, L., Han, J., Christensen, H., & Torok, M. (2021). Effect of Engagement With Digital Interventions on Mental Health Outcomes: A Systematic Review and Meta-Analysis. *Frontiers in digital health*, *3*, 764079. https://doi.org/10.3389/fdgth.2021.764079

Hao Liu, Huaming Peng, Xingyu Song, Chenzi Xu, Meng Zhang, Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness, Internet Interventions, Volume 27, 2022,100495, ISSN 2214-7829, https://doi.org/10.1016/j.invent.2022.100495.

Kambeitz-Ilankovic, L., Rzayeva, U., Völkel, L., Wenzel, J., Weiske, J., Jessen, F., Reininghaus, U., Uhlhaas, P. J., Alvarez-Jimenez, M., & Kambeitz, J. (2022). A systematic review of digital and face-to-face cognitive behavioral therapy for depression. NPJ digital medicine, 5(1), 144. https://doi.org/10.1038/s41746-022-00677-8

Levis, M., Leonard Westgate, C., Gui, J., Watts, B. V., & Shiner, B. (2021). Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models. *Psychological medicine*, *51*(8), 1382–1391. https://doi.org/10.1017/S0033291720000173

Peter J. Na, Satyanarayana R. Yaramala, Jihoon A. Kim, Hyelee Kim, Fernando S. Goes, Peter P. Zandi, Jennifer L. Vande Voort, Bruce Sutor, Paul Croarkin, William V. Bobo, The PHQ-9 Item 9 based screening for suicide risk: a validation study of the Patient Health Questionnaire (PHQ)−9 Item 9 with the Columbia Suicide Severity Rating Scale (C-SSRS), Journal of Affective Disorders, Volume 232, 2018, Pages 34-40, ISSN 0165-0327, https://doi.org/10.1016/j.jad.2018.02.045.

Werntz A, Amado S, Jasman M, Ervin A, Rhodes J. Providing Human Support for the Use of Digital Mental Health Interventions: Systematic Meta-review J Med Internet Res 2023;25:e42864 URL: https://www.jmir.org/2023/1/e42864 DOI: 10.2196/42864.