AI-Based Anonymous Confession Platforms for Mental Health Support: A Systematic Review

Mrs. Umme Ayeman Saqib Gani

Vaishnavi Mashakhetri

Tanvi Wadhai

Gulshan Nagalwade

Rushabh Ingle.

Department of Artificial Intelligence and Data Science
Priyadarshini College of Engineering, Nagpur

vaishnavimashakhetri1@gmail.com
tanviwadhai@gmail.com
gulshannagalwade02@gmail.com
rushabhingle907@gmail.com

Abstract

Mental health challenges have become a significant global concern, especially among young adults, due to increasing academic pressure, social isolation, and limited access to professional psychological support. Many individuals hesitate to share their emotions openly because of fear of judgment, social stigma, or lack of anonymity. To address this issue, anonymous digital platforms have emerged as a promising solution for emotional expression and mental well-being support. This review paper analyzes existing research on AI-based mental health chatbots, anonymous confession systems, and natural language processing techniques used for emotion detection and supportive response generation. The study systematically reviews recent literature to understand the effectiveness, benefits, and limitations of current solutions in providing emotional support while ensuring user privacy. A comparative analysis of selected studies highlights gaps such as limited personalization, insufficient emotional safety mechanisms, and ethical concerns related to data privacy. Based on these findings, the paper discusses the potential role of an AI-powered anonymous confession platform, referred to as a Confession Box, which enables users to express emotions freely while receiving empathetic and supportive responses. The review concludes by emphasizing the need for secure, emotionally intelligent, and ethically designed confession-based systems to enhance mental health support in digital environments.

Keywords: Mental Health, Anonymous Confession Platforms, AI Chatbots, Emotional Well-being, Natural Language Processing, Privacy and Ethics.

---

1. Introduction

Mental health has emerged as a critical concern in modern society, affecting individuals across all age groups and cultural backgrounds. Rapid technological advancement, academic pressure, workplace stress, and social isolation have significantly contributed to increased levels of anxiety, depression, and emotional distress, particularly among young adults. Despite growing awareness, access to timely and professional mental health support remains limited due to factors such as social stigma, high treatment costs, and a shortage of trained mental health professionals. As a result, many individuals suppress their emotions rather than seeking help, which can lead to severe psychological consequences.

With the widespread adoption of digital technologies, online platforms have become an important medium for emotional expression and mental health assistance. In recent years, artificial intelligence (AI)–based mental health applications and chatbots have gained popularity due to their ability to provide immediate, scalable, and cost-effective support. These systems often utilize natural language processing (NLP) techniques to analyze user input, detect emotional states, and generate appropriate responses. While such platforms have shown potential in offering basic emotional support, users often express concerns regarding privacy, personalization, and emotional authenticity.

Anonymity plays a crucial role in encouraging individuals to express their true feelings without fear of judgment or social repercussions. Anonymous confession platforms allow users to share sensitive thoughts and emotions freely, creating a safer space for self-expression. However, many existing confession-based systems lack intelligent emotional understanding and rely primarily on human moderation or static responses, which limits their effectiveness in providing meaningful mental health support.

This review paper explores existing research related to AI-driven mental health chatbots, anonymous confession systems, and emotion analysis techniques. By systematically examining current literature, the paper identifies key contributions, limitations, and research gaps in existing solutions. Furthermore, the study highlights the importance of integrating anonymity, emotional intelligence, and ethical data handling into digital mental health platforms. Based on the insights gained from the literature, the paper discusses the relevance and potential impact of an AI-powered anonymous Confession Box system designed to support emotional well-being while

maintaining user privacy and safety.

Fig.(a)Workflow of Emotional System

2. Literature Review

Recent advancements in artificial intelligence have significantly influenced the domain of mental health care, offering innovative approaches for assessment, support, and intervention. Several studies have explored the application of AI technologies in mental health settings, highlighting their potential to enhance accessibility, efficiency, and scalability of psychological support systems. Olawade et al. (2025) presented a narrative review focusing on the use of AI in forensic mental health, emphasizing its role in risk assessment, behavioral analysis, and decision support. While the study demonstrated the effectiveness of AI-driven tools, it also raised concerns regarding ethical implications, transparency, and the reliability of automated mental health judgments.

Ethical challenges associated with conversational AI systems in mental health care have been widely discussed in recent literature. Meadi et al. (2025) conducted a scoping review examining ethical risks such as data privacy, consent, emotional dependency, and bias in AI-driven mental health platforms. The study highlighted that many conversational systems operate in environments where user conversations are not fully anonymous, potentially discouraging users from expressing deeply personal emotions. These findings underline the importance of anonymity and responsible AI design in emotionally sensitive applications.

Systematic reviews have further analyzed the effectiveness of AI-based mental health interventions. Dehbozorgi et al. (2025) reviewed various AI applications including chatbots, predictive analytics, and emotion recognition systems. Their findings indicated that AI-based conversational agents can provide timely and accessible emotional support, particularly in regions with limited mental health resources. However, the study noted limitations such as insufficient personalization and the inability of AI systems to fully understand complex human emotions.

Similarly, Ni and Jia (2025) conducted a scoping review mapping AI-driven digital mental health interventions across screening, monitoring, prevention, and clinical support. The review emphasized the role of natural language processing techniques, such as recurrent neural networks, in analyzing anonymous user data. The authors observed that anonymity improves user engagement, but many systems lack integrated emotional safety mechanisms to detect distress or crisis situations effectively.

Earlier studies have also contributed valuable insights into AI-enabled mental health chatbots. Omarov et al. (2023) systematically reviewed AI-based chatbots used in mental health care and identified perceived anonymity, immediate response, and personalization as major advantages. Despite these benefits, the review pointed out challenges related to data security, emotional accuracy, and long-term user trust, which remain unresolved in many existing platforms.

Privacy and emotional intimacy in AI-driven mental health systems have been examined by Kneese et al. (2025), who explored how users interact with conversational agents in sensitive contexts. Their study suggested that users often treat chatbots as emotionally intimate entities, sharing personal confessions and vulnerabilities. This behavior raises critical concerns regarding privacy protection, ethical responsibility, and the emotional boundaries of AI systems.

Security and risk mitigation have also gained attention in AI-driven mental health research. Banerjee et al. (2024) analyzed security protocols in workplace mental health chatbots and emphasized the need for robust data protection measures to safeguard sensitive user information. The study stressed that unintended data exposure could undermine user trust and compromise mental well-being.

Legal and ethical aspects of data disclosure in AI-based mental health systems were explored by Petročnik et al., who examined suicide prediction algorithms and personal data handling. The study highlighted the tension between user confidentiality and intervention responsibility, reinforcing the need for ethically designed systems that respect user anonymity while ensuring safety.

Furthermore, Khawaja (2025) emphasized responsible design principles for therapeutic AI systems, particularly voice-based conversational agents. The research argued that AI should complement, rather than replace, human mental health services, and must be designed with transparency, accountability, and user well-being as core priorities. Hamdoun et al. (2023) similarly discussed the balance between the growing demand for digital mental health applications and the associated risks, noting that anonymous AI-based platforms can reduce stigma but also introduce ethical and security challenges.

Overall, the reviewed literature demonstrates that AI-driven mental health systems and anonymous platforms offer significant potential for emotional support. However, limitations such as inadequate emotional intelligence, privacy concerns, ethical risks, and lack of personalization persist. These gaps highlight the need for an AI-powered anonymous Confession Box system that prioritizes emotional safety, user anonymity, and ethical data handling while providing empathetic and supportive interactions.

3. Research Gaps

The review of existing literature reveals several gaps in current AI-driven mental health and anonymous communication systems. Although numerous studies have explored the use of artificial intelligence for mental health assessment and support, most existing platforms focus primarily on clinical assistance, risk prediction, or general conversational support rather than providing a dedicated space for anonymous emotional expression. The integration of anonymity with emotionally intelligent AI responses remains limited.

Many AI-based mental health chatbots lack advanced emotional understanding and contextual awareness, which restricts their ability to respond empathetically to complex or deeply personal confessions. While natural language processing techniques are widely used for emotion detection, current systems often fail to accurately interpret nuanced emotional states, such as mixed emotions or suppressed distress. Additionally, personalization in responses is minimal, leading to generic interactions that may not adequately support users over time.

Privacy, security, and ethical concerns are frequently highlighted in the literature, yet practical implementations addressing these challenges are insufficient. Existing studies emphasize risks related to data storage, confidentiality, and user trust, but few propose concrete frameworks that ensure complete anonymity while maintaining emotional safety and ethical responsibility. Furthermore, mechanisms for identifying emotional crises and providing appropriate support without violating user privacy are underexplored.

Another significant gap lies in the lack of user-centric design approaches tailored to anonymous confession platforms. Most reviewed systems are designed for structured mental health interventions rather than open-ended emotional expression. As a result, there is limited research on platforms that allow users to freely share confessions while receiving supportive, non-judgmental, and ethically guided AI responses.

These gaps indicate the need for an AI-powered anonymous Confession Box system that combines emotional intelligence, user anonymity, ethical data handling, and supportive interaction design to enhance digital mental health support.

4. Discussion

The reviewed literature highlights the growing role of artificial intelligence in supporting mental health care through digital platforms, particularly conversational agents and decision-support systems. AI-driven chatbots have demonstrated potential in improving accessibility, reducing stigma, and providing immediate emotional assistance. Features such as anonymity and availability have been identified as key factors encouraging users to share personal thoughts and emotions more openly. However, despite these advantages, existing systems often fall short in delivering emotionally nuanced and context-aware support.

A recurring theme across studies is the challenge of balancing user anonymity with ethical responsibility and emotional safety. While anonymity enables honest emotional expression, it also complicates crisis detection and intervention strategies. Many reviewed platforms either compromise privacy to ensure safety or fail to implement effective mechanisms for identifying high-risk emotional states. This highlights a critical design trade-off that remains insufficiently addressed in current research.

Ethical concerns related to data privacy, emotional dependency, and transparency are also widely discussed. Although several studies emphasize these issues, there is a lack of standardized frameworks guiding the ethical design and deployment of AI-based mental health systems. Moreover, personalization and long-term emotional engagement are limited in most existing solutions, leading to generic interactions that may reduce user trust and effectiveness over time.

In this context, an AI-powered anonymous Confession Box system represents a promising direction by combining free emotional expression, empathetic AI responses, and ethical data handling. Such a system can complement existing mental health services by offering a safe, non-judgmental digital space for emotional release while addressing the limitations identified in current platforms.

5. Conclusion

This review paper examined existing research on AI-driven mental health systems, anonymous confession platforms, and emotion-aware conversational agents. The analysis reveals that while artificial intelligence has significantly enhanced the accessibility and scalability of mental health support, current systems face limitations related to emotional intelligence, personalization, privacy, and ethical design. Anonymity has emerged as a crucial factor in encouraging emotional disclosure, yet it remains inadequately integrated with intelligent and responsible AI mechanisms.

The findings indicate a clear research gap in the development of anonymous, emotionally intelligent, and ethically designed digital platforms that support mental well-being. An AI-powered Confession Box system can address these challenges by providing users with a secure and supportive environment to express emotions freely while receiving empathetic responses. Future research should focus on improving emotion detection accuracy, establishing ethical guidelines, and integrating crisis-support mechanisms without compromising user anonymity. Overall, such systems have the potential to play a complementary role in modern mental health care by bridging gaps between emotional expression and accessible support.

6.Acknowledgements

7.Referances

[1] D. B. Olawade, F. I. Ayoola, T. O. Ebo, A. J. Asaolu, *et al.*, "Artificial intelligence in forensic mental health: A review of applications and implications," *Journal of Forensic Mental Health*, Elsevier, 2025.

[2] M. R. Meadi, T. Sillekens, S. Metselaar, *et al.*, "Exploring the ethical challenges of conversational AI in mental health care: A scoping review," *JMIR Mental Health*, 2025.

[3] R. Dehbozorgi, S. Zangeneh, E. Khooshab, D. H. Nia, *et al.*, "The application of artificial intelligence in the field of mental health: A systematic review," *BMC Psychiatry*, Springer, 2025.

[4] Y. Ni and F. Jia, "A scoping review of AI-driven digital interventions in mental health care," *Healthcare*, MDPI, 2025.

[5] T. Kneese, B. Vecchione, and A. Marwick, "A chatbot for the soul: Mental health care, privacy, and intimacy in AI-based conversational agents," *Communication and Change*, Springer, 2025.

[6] Z. Khawaja, "Fostering responsible therapeutic AI support: Guiding ethical design and usage of conversational agents in digital mental healthcare," 2025.

[7] V. K. Barbudhe and V. M. Rakhade, "AI chatbots for mental health therapist – Review paper," *International Journal of Advanced Research in Computer Science*, 2025.

[8] E. M. Provost, S. H. Sperry, J. Tavernor, *et al.*, "Emotion recognition in the real world: Passively collecting and estimating emotions from natural speech data of individuals with bipolar disorder," *IEEE Transactions on Affective Computing*, vol. 16, no. 1, pp. 28–40, 2024.

[9] S. Banerjee, A. Agarwal, and A. K. Bar, "Securing well-being: Exploring security protocols and mitigating risks in AI-driven mental health chatbots," *American Journal of Computer Science*, 2024.

[10] B. Omarov and S. Narynov, "Artificial intelligence-enabled chatbots in mental health: A systematic review," *Computers, Materials & Continua*, 2023.

[11] S. Hamdoun, R. Monteleone, *et al.*, "AI-based and digital mental health apps: Balancing need and risk," *IEEE Technology and Society Magazine*, 2023.

[12] S. Toblad, "Users' attitudes on changing anonymity on social media: Retention of pseudonymity and identity," *DIVA Portal: Social Computing*, 2023.

[13] M. D. R. Haque, *et al.*, "An overview of chatbot-based mobile mental-health apps," *IEEE Access*, 2023.

[14] P. Kumar, *et al.*, "A BERT-based dual-channel explainable text emotion recognition model," *Journal of Artificial Intelligence Research*, 2022.

[15] X. Song, "A qualitative research on usage intention and platform features in anonymous social media applications," *JMIR / Frontiers in Psychology*, 2022.

[16] E. M. Boucher, *et al.*, "Artificially intelligent chatbots in digital mental health: A scoping review," *Digital Mental Health Review*, 2021.

[17] M. Zhang and T. Fung, "Examining the role of anonymity in online self-disclosure: A study on confession-based apps," *Computers in Human Behavior*, vol. 122, p. 106850, 2021.

[18] J. Deng and F. Ren, "A survey of textual emotion recognition and its challenges," *IEEE Transactions on Affective Computing*, 2021.

[19] V. Scotti, R. Tedesco, and L. Sbattella, "A modular data-driven architecture for empathetic conversational agents," in *Proc. IEEE BigComp*, Jeju, Korea, 2021, pp. 365–368.

[20] A. Sharma and R. Gupta, "Anonymous social media applications: Opportunities and challenges," *International Journal of Computer Applications*, vol. 182, no. 35, pp. 15–20, 2020.

[21] J. Xu, Z. Hu, J. Zou, and A. Bi, "Intelligent emotion detection method based on deep learning in medical and health data," *IEEE Access*, vol. 8, pp. 3802–3811, 2020.

[22] S. Kaur, P. Singh, and J. K. Chhabra, "Exploring online anonymity: A case study of confession platforms," *International Journal of Recent Technology and Engineering*, vol. 8, no. 4, pp. 234–238, 2019.

[23] J. S. Park and H. S. Lee, "Psychological effects of anonymous communication in mobile confession applications," *Journal of Information Technology Services*, vol. 18, no. 2, pp. 55–66, 2019.

[24] C. V. Clark-Gordon, N. D. Bowman, A. K. Goodboy, and A. Wright, "Anonymity and online self-disclosure: A meta-analysis," *Communication Reports*, vol. 32, no. 2, pp. 1–14, 2019.

[25] S. Wang, Y. Zhao, and X. Li, "Understanding user engagement in anonymous mobile applications," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, CSCW, pp. 1–23, 2019.

[26] J. Han, Z. Zhang, Z. Ren, and B. Schuller, "EmoBed: Strengthening monomodal emotion recognition via training with crossmodal emotion embeddings," *IEEE Transactions on Affective Computing*, 2019.

[27] J. X. Chen, *et al.*, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44317–44328, 2019.

[28] S. Misoch, "Online self-disclosure and the role of visual anonymity," *Computers in Human Behavior*, vol. 51, pp. 246–256, 2015.