

Retrieval-Augmented LLM-Based Empathetic Chatbot for Early Postpartum Depression Screening in Aceh

¹Cut Amalia Saffiera, ²Fiqey Indriati Eka Sari, ³Nur Amalia Hasma, ⁴Rizaki Akbar
^{1,2,3,4}Institut Teknologi Sepuluh Nopember

saffieraamalia@gmail.com, indriatifiqey16@gmail.com, nuramaliahasma05@gmail.com, rizakiakbar004@gmail.com

Submitted : Jan 21, 2026 | **Accepted** : Mar 3, 2026 | **Published** : April 2, 2026

Abstract: Postpartum Depression (PPD) remains a significant maternal mental health concern, particularly in low-resource settings where access to professional psychological services is limited. Although digital mental health tools have emerged to address this gap, most existing chatbot-based systems rely on rule-based interactions, offer limited personalization, and lack integration of structured clinical screening mechanisms. This study addresses the lack of culturally adapted, LLM-based empathetic chatbots for postpartum mental health screening in low-resource Indonesian settings. We design and implement an AI-driven conversational chatbot that integrates a Retrieval-Augmented Generation (RAG) architecture with a Large Language Model (LLM) to enable context-aware, knowledge-grounded response generation. The system incorporates a Patient Health Questionnaire-9 (PHQ-9)-based screening module to support early identification of depressive symptoms and adaptive conversational support. An early-stage usability evaluation was conducted through a seven-day user interaction study involving 30 postpartum mothers in Aceh, with 12 participants completing the System Usability Scale (SUS). The system achieved an average SUS score of 85.63, indicating excellent perceived usability. While the evaluation focuses on usability rather than clinical effectiveness, the findings suggest that the proposed system demonstrates feasibility as a culturally adapted, scalable digital support tool for early postpartum mental health screening. Further studies with larger samples and long-term evaluation are required to assess clinical impact and sustained user engagement.

Keywords: Artificial Intelligence; Chatbot; Conversational Agent; Digital Health; Large Language Models; Postpartum Depression; Retrieval-Augmented Generation

INTRODUCTION

Postpartum Depression (PPD) is a common maternal mental health condition, affecting approximately 13% of mothers worldwide, with a higher prevalence in developing countries reaching up to 19.8% (World Health Organization, 2025). In Indonesia, the reported prevalence of PPD ranges from 20% to 30%, indicating a substantial burden on maternal health systems (Ernawati, 2020). Beyond its psychological dimension, PPD has broader implications, including disruptions to family dynamics, reduced caregiving capacity, and potential impacts on early child development. These challenges are particularly pronounced in regions such as Aceh, where socio-economic constraints and limited access to healthcare infrastructure, including mental health professionals, remain significant issues (Boothby, Veatch, & Pentecost, 2011). Such conditions increase vulnerability to postpartum mental health problems while simultaneously hindering early detection and timely intervention.

Early identification of depressive symptoms is commonly conducted using psychometric instruments, such as self-report questionnaires. However, standalone questionnaire-based screening approaches face several limitations, including response bias, low mental health literacy, and social or cultural pressures that may discourage mothers from openly reporting symptoms (Rickwood & Coleman-Rose, 2023). These limitations are further compounded by the shortage of trained mental health professionals, such as psychologists and psychiatrists, who are required to interpret assessment results and provide follow-up interventions, particularly in underserved regions like Aceh (Mawarpury, Sari, & Safrina, 2017). Consequently, alternative approaches that can provide accessible, supportive, and culturally sensitive early screening are needed.

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Recent advances in Artificial Intelligence (AI) have created new opportunities for supporting mental health services, particularly through the development of empathic conversational agents. AI-based chatbots can provide accessible, scalable, and anonymous conversational support, enabling users to engage without geographical or temporal constraints (Abd-Alrazaq et al., 2021; Tsai et al., 2022). Furthermore, progress in Natural Language Processing (NLP), machine learning, and Large Language Models (LLMs) has significantly enhanced chatbot capabilities, allowing systems to better capture conversational context and generate more natural and empathetic responses (Brännström, Wester, & Nieves, 2024). These technological developments suggest the potential for more adaptive and context-aware digital mental health interventions.

Several studies have demonstrated the feasibility of AI-based chatbots for postpartum mental health support, particularly for symptom self-screening and preliminary psychological assistance (Ahtisham et al., 2023; Suharwardy et al., 2023; Kaur et al., 2024; Mane et al., 2023). Earlier systems frequently relied on predefined conversational flows or rule-based architectures, which limited contextual adaptability (Vinarti et al., 2024), while more recent approaches have begun incorporating machine learning and LLM-based components. However, the integration of advanced language models with structured clinical screening mechanisms and culturally grounded knowledge remains limited, particularly in low-resource Indonesian settings. In addition, the integration of standardized screening tools within dynamic conversational frameworks is not consistently implemented, particularly in low-resource settings.

Despite these developments, an important gap remains in the development of conversational systems that integrate three key components at once, namely LLM-based context understanding, retrieval-augmented knowledge support, and embedded standardized screening in a culturally adapted Indonesian setting. Although recent studies have introduced LLM-based or retrieval-augmented architectures, these elements are often implemented separately rather than within a unified and culturally contextualized framework. This indicates the need for a unified, culturally adapted LLM-RAG framework specifically designed to support postpartum mental health in low-resource regions such as Aceh.

In response to these challenges, this study focuses on the development of *SeudaraBunda*, a digital mental health support application for postpartum mothers, featuring an AI-driven empathic chatbot named *Meurabot*. The research aims to design, develop, and conduct an early-stage evaluation of an AI-based conversational system that integrates a LLM with a RAG architecture to enable context-aware, knowledge-grounded responses. The system incorporates a PHQ-9-based screening module to facilitate early identification of depressive symptoms and guide adaptive conversational support tailored to postpartum mothers in Aceh.

The novelty of this research lies in the integration of an Indonesian-language LLM-based empathic chatbot with a retrieval-augmented mechanism and embedded PHQ-9 screening, explicitly adapted to the local cultural context of Aceh for early PPD detection. The findings are expected to contribute to the development of culturally relevant digital mental health solutions and provide insights into the application of advanced conversational AI for maternal mental health services in low-resource settings.

LITERATURE REVIEW

Postpartum Mental Health and Early Screening

Postpartum mental health has become an important research topic due to its implications for maternal well-being, family functioning, and early child development (World Health Organization, 2022). From a research perspective, Postpartum Depression (PPD) is frequently highlighted as a condition that requires early identification mechanisms, particularly in settings where formal mental health services are not consistently available (O'Hara & McCabe, 2013).

In many developing regions, social stigma, cultural beliefs, and limited mental health literacy discourage mothers from seeking professional mental health services, resulting in delayed identification and treatment of postpartum mental health conditions (Upadhyay et al., 2017; Howard et al., 2014; Jones, 2022). Recent large-scale evidence indicates that perinatal and postpartum depression remain highly prevalent in low- and middle-income countries, reflecting persistent gaps in access to timely and appropriate mental health care (Mitchell et al., 2023).

In recent years, digital mental health systems have increasingly incorporated standardized screening instruments to support early detection and continuous monitoring. Such systems enable structured data collection and data-driven mental health support that can be delivered at scale through mobile and web-based platforms (Torous et al., 2021).

Digital Screening Instruments for Postpartum Depression

Early identification of depressive symptoms is commonly conducted using standardized psychometric instruments. One of the most widely used tools is the Patient Health Questionnaire-9 (PHQ-9) (World Health Organization, 2022). The PHQ-9 consists of nine items that measure the frequency of depressive symptoms over the past two weeks, with total scores reflecting symptom severity levels.

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PHQ-9 is designed as a screening instrument to support early symptom identification and severity assessment, while formal diagnosis remains the responsibility of qualified mental health professionals. In digital mental health systems, PHQ-9 is commonly used as a structured screening component that provides standardized input data to support early detection, symptom monitoring, and adaptive system behavior. Table 1 presents the general interpretation of PHQ-9 scores used in screening contexts.

By embedding PHQ-9 within digital mental health applications, screening results can be used as system input to guide personalized feedback, adaptive conversational strategies, and supportive recommendations. This approach enables scalable early screening and ongoing monitoring of postpartum mental health conditions while reducing reliance on direct clinical assessment, particularly in settings with limited access to mental health professionals.

Table 1. Interpretation of PHQ-9 Scores

Score Range	Depression Level
0 - 4	None/Minimal
5 - 9	Mild
10 - 14	Moderate
15 - 27	Moderately Severe to Severe

AI-Based Chatbots for Postpartum Mental Health Support

Advances in Artificial Intelligence (AI) have enabled the development of conversational agents capable of supporting mental health services through interactive dialogue. AI-based chatbots offer several advantages, including accessibility, scalability, anonymity, and availability without geographical or temporal constraints (Abd-Alrazaq et al., 2021; Tsai et al., 2022). These characteristics make chatbots particularly suitable for postpartum mothers who may face barriers to accessing conventional psychological services.

Recent progress in Natural Language Processing (NLP), machine learning, and Large Language Models (LLMs) has significantly enhanced chatbot capabilities. Modern conversational agents are able to process contextual information, understand user intent, and generate more natural and empathetic responses compared to earlier rule-based systems (Brännström, Wester, & Nieves, 2024).

Several studies have explored the use of AI-based chatbots for postpartum mental health support. Prior research has demonstrated the feasibility of chatbots for self-screening of postpartum depressive symptoms with promising performance outcomes (Ahtisham et al., 2023). Controlled evaluations have also shown that chatbot-based interventions may contribute to improved psychological well-being among postpartum mothers (Suharwardy et al., 2023). Other studies have reported the effectiveness of machine learning-based chatbots in mental health detection and interactive health education for new mothers (Kaur, Sharma, & Singh, 2024; Mane et al., 2023).

More recent developments have begun incorporating Large Language Models (LLMs) and retrieval-augmented mechanisms to enhance contextual understanding and knowledge grounding in maternal health applications. For example, Khan et al. (2025) proposed an LLM- and retrieval-based chatbot framework for postpartum depression support, while Utami et al. (2025) explored an LLM-RAG conversational system adapted to Indigenous Australian perinatal mental health contexts.

Despite these encouraging developments, several limitations remain. Many systems are developed primarily in English, and structured screening tools are not consistently embedded within adaptive conversational architectures. Cultural and linguistic adaptation often focuses on translation rather than deeper contextual alignment with local norms, beliefs, and healthcare realities. In the Indonesian context, local-language chatbot initiatives such as KIA-CHAT (Vinarti et al., 2024) have begun to emerge; however, these systems primarily employ rule-based mechanisms and are not specifically designed to provide adaptive, empathetic conversational support for maternal mental health screening.

Table 2. Comparison of Chatbots for Postpartum Mental Health Support

Study	Language	Architecture	Screening Tool	Cultural Adaptation
Ahtisham et al. (2023)	English	AI-based chatbot	Self-screening reported	Not reported
Suharwardy et al. (2023)	English	AI-based chatbot	Not reported	Not reported
Kaur et al. (2024)	English	AI-based chatbot	Screening mentioned (unspecified tool)	Not reported
Mane et al. (2023)	English	AI-based chatbot	Not reported	Not reported

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Vinarti et al. (2024)	Indonesian	Rule-based chatbot	Not reported	Language adaptation only
Khan et. al (2025)	English	LLM + RAG chatbot	Not reported	Not reported
Utami et al. (2025)	English	LLM + RAG chatbot	Not reported	Indigenous Australian context

As summarized in Table 2, most existing postpartum mental health chatbots focus on feasibility, accessibility, or general psychological support. While recent studies have begun integrating LLM-based and retrieval-augmented architectures, the systematic combination of advanced conversational intelligence with embedded standardized screening instruments remains limited. Furthermore, comprehensive cultural adaptation beyond language translation is still underrepresented, particularly within Indonesian and other low-resource settings.

These observations indicate that an integrated system combining LLM-based contextual understanding, retrieval-augmented knowledge grounding, and embedded standardized screening within a culturally adapted Indonesian postpartum mental health framework remains underexplored. Addressing this gap provides the primary motivation for the present study.

METHOD

Research Design

The research is positioned as an early stage system development and feasibility study rather than a hypothesis testing clinical investigation. This study adopts a Design Science Research oriented system development approach within the domain of applied artificial intelligence for digital mental health. In line with Design Science Research principles, the study focuses on the design, implementation, and preliminary evaluation of a functional artifact, namely the SeudaraBunda chatbot system, developed to support postpartum mental health within a specific socio-cultural context.

The research workflow consists of seven main stages, as illustrated in Figure 1, ranging from literature based requirement analysis to usability evaluation. This methodological approach, informed by prior studies on digital mental health chatbots and perinatal support systems (Abd Alrazaq et al., 2021; Mane et al., 2023; Vinarti et al., 2024), emphasizes literature driven requirement analysis, contextual adaptation, and usability focused evaluation. The primary objectives of this study are:

1. To develop a context aware AI chatbot architecture integrating Retrieval Augmented Generation and domain specific knowledge.
2. To validate psychological and medical content through expert review.
3. To evaluate perceived usability among postpartum users.

A total of 30 postpartum mothers participated in the user trial using voluntary sampling. The sample size is consistent with formative usability evaluation practices, where moderate participant numbers are considered sufficient to identify major interaction patterns and usability issues in early stage system assessment.

This study does not include clinical outcome measurement, such as depression symptom reduction or psychiatric diagnosis confirmation. The SeudaraBunda system is designed as a supportive digital companion rather than a medical treatment intervention. Evaluation of clinical effectiveness would require longitudinal monitoring, controlled intervention design, and structured psychiatric assessment under formal clinical research protocols. Clinical effectiveness evaluation is therefore proposed for future research.

Literature-Based Requirement Analysis

The initial stage involved a focused literature review to inform system requirements and design decisions. The review covered three primary aspects: (1) digital screening instruments for postpartum depression, (2) conversational chatbot technologies for mental health support, and (3) contextual factors influencing postpartum mental health support in Aceh.

Findings from the literature were translated into functional and non-functional system requirements, including the integration of a standardized screening instrument, support for empathic conversational interaction, and adaptation to local linguistic and cultural characteristics.

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



Fig. 1 Research Methodology Workflow

Contextual Requirement Analysis

In addition to literature analysis, contextual requirements were identified to ensure system suitability for postpartum mothers in Aceh. Several key considerations emerged from this analysis. First, the system must use clear and easily understandable language to allow users to express emotions and comprehend guidance without difficulty. Second, cultural and religious sensitivity is essential, particularly in greeting styles, expressions, and communication tone. Third, anonymous interaction and ease of access were identified as critical factors to reduce social stigma and encourage user engagement.

As a supportive system rather than a replacement for professional care, the chatbot is designed to provide emotional validation, initial psychological support, and postpartum-related information in a safe and non-judgmental manner.

System Design

Based on the identified requirements, the *SeudaraBunda* application was designed with three core components. The first component is an AI-based empathic chatbot, referred to as *Meurabot*, which functions as a conversational companion providing supportive dialogue and postpartum mental health information. The second component is a self-assessment feature that adapts the PHQ-9 for early screening of depressive symptoms. The third component is an adaptive recommendation module that provides supportive tips based on screening results.

User interface design prioritizes simplicity, accessibility, and comfort, taking into account the physical and emotional conditions commonly experienced by postpartum mothers.

Expert Validation

To ensure content accuracy and safety, chatbot conversation flows, psychological content, and PHQ-9–based recommendations were reviewed and validated by domain experts, including a psychiatrist, a clinical psychologist, and an obstetrician–gynecologist. This validation process aimed to confirm alignment with medical and psychological standards and to minimize the risk of inappropriate or misleading responses.

System Architecture and Prototype Development

The chatbot system was developed using a RAG architecture to support context-aware and knowledge-grounded response generation, as illustrated in Figure 2. The architecture integrates a Large LLM with a structured postpartum knowledge base and user-specific data.

User inputs are submitted through the chatbot interface and processed by an AI engine implemented using FastAPI, which coordinates communication between system components. The RAG mechanism utilizes two primary information sources. The first is a domain-specific knowledge base covering eight postpartum-related

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

domains, including mental health, breastfeeding, infant care, maternal self-care, family relationships, postpartum symptoms, nutrition and physical activity, and postpartum warning signs. Knowledge documents are indexed using FAISS to enable efficient semantic retrieval.

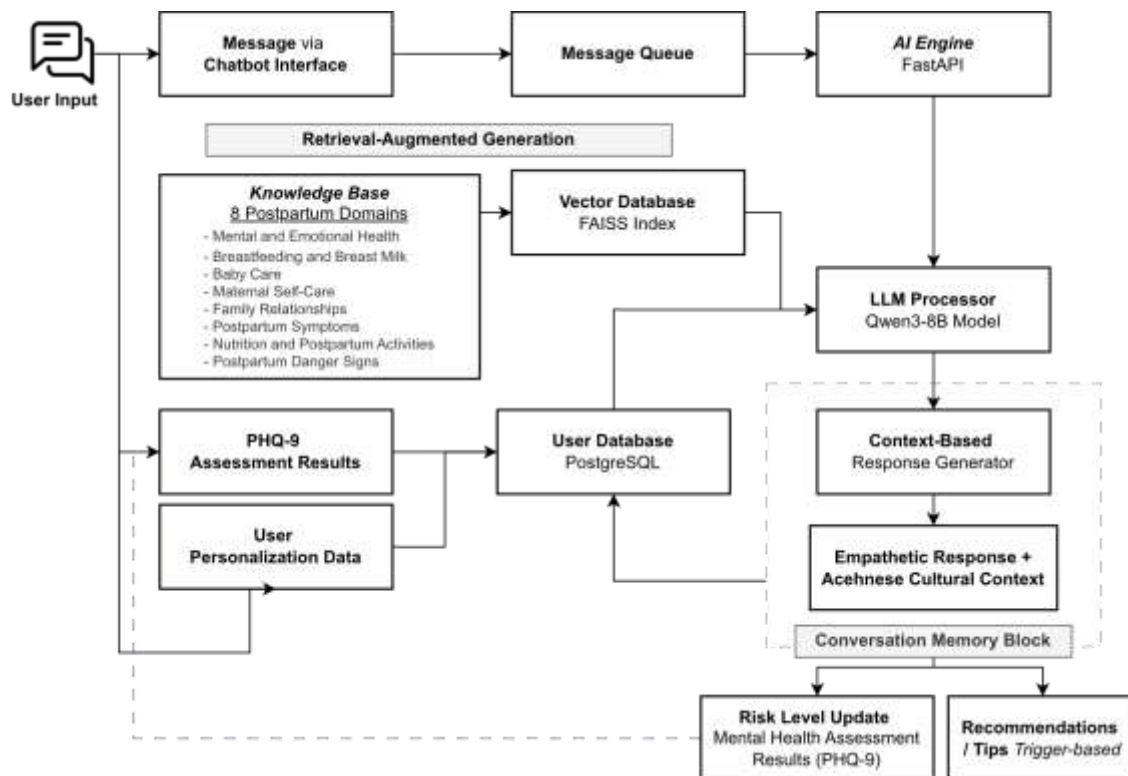


Fig. 2 Chatbot System Architecture

The second information source is a relational database implemented using PostgreSQL, which stores PHQ-9 screening results and user personalization data. Retrieved knowledge, screening outcomes, and conversation context are processed by the LLM core (Qwen3-8B) to generate responses. A response generation layer further refines outputs by incorporating empathic tone and cultural adaptation to align with local communication norms. Conversation history is maintained as contextual memory to support dialogue continuity, dynamic risk level updates based on PHQ-9 scores, and trigger-based supportive recommendations.

Implementation and User Trial

The developed prototype was deployed for a user trial involving 30 postpartum mothers in Aceh who voluntarily agreed to participate. Participants were recruited using voluntary sampling, with eligibility criteria based on postpartum status within a specified timeframe. During the trial period, participants interacted with the system and completed the screening and conversational features.

Usability Evaluation

System evaluation focused on usability assessment using the System Usability Scale (SUS) (Brooke, 1996). SUS was selected due to its reliability and widespread use in evaluating user perceptions of system usability. Participants completed the SUS questionnaire after the trial period, and the resulting scores were used to assess overall system usability.

RESULT

Expert Validation

The *SeudaraBunda* prototype was evaluated by a panel of experts, including a psychiatrist, a psychologist, and an obstetrician-gynecologist (Ob/Gyn), to ensure the content and workflow align with established clinical and psychological standards. The validation focused on three main aspects: screening instruments and application workflow, psychological content and conversational language, and medical accuracy of responses.

*name of corresponding author



Table 2. Expert Validation of Chatbot Content

Aspect Validated	Validator	Outcome
Screening instruments and workflow	Psychiatrist	Validated for clinical appropriateness and alignment with PHQ-9 guidelines
Psychological content	Psychologist	Validated for empathy, clarity, and accuracy
Medical content	Ob/Gyn	Validated for accuracy and appropriateness

The results indicate that the system content is reliable, clinically sound, and communicates empathetically, forming a solid foundation for user interaction.

Prototype

Overview

The *SeudaraBunda* application was implemented using LLM integrated with RAG, a vector database, and lightweight fine-tuning (LoRA). The chatbot, *Meurabot*, is designed to provide empathetic conversation in Indonesian, with consideration of local Acehnese cultural context. Figure 3 illustrates the main interface of the *SeudaraBunda* application, showing the home page that serves as the primary access point to the chatbot conversation feature, emotional state assessment (PHQ-9), and personalized recommendations.



Fig. 3 *SeudaraBunda* Home Interface

Figure 3 illustrates the main interface of the *SeudaraBunda* application. The layout emphasizes minimal cognitive load, with clear navigation toward three primary features: chatbot interaction, PHQ-9 screening, and recommendation access. The simplified interface design may contribute to the high usability score by reducing navigation complexity and improving clarity of user pathways.

Example Chatbot Interaction

To demonstrate the chatbot’s empathetic response capability, an example interaction is presented below (see Fig. 4).

User: “I am confused about how to fulfill my role as a mother.”

Meurabot: “I understand, Bun. Feeling confused about your role as a mother is normal, especially in the beginning. Never feel alone, as many other mothers are also learning and adapting. Start with small steps, like taking care of your baby’s basic needs, and gradually follow your daily routine. You’re doing great, Bun, don’t hesitate to ask for help or discuss with people close to you. Every process takes time, and you are on your own path.”

This interaction illustrates that *Meurabot* responds in a supportive, empathetic, and culturally appropriate manner, providing reassurance and actionable guidance.

*name of corresponding author





Fig. 4. Example Meurabot interaction showing empathetic response

Example Personalized Tips

Based on PHQ-9 assessment scores, the system provides personalized recommendations to support emotional well-being. For example, users with a mild score (0-4) receive light supportive advice, including encouragement to maintain daily routines, engage in brief self-care activities, and reflect on positive moments (see Fig. 5). This functionality enables adaptive support tailored to the user’s current mood, promoting early intervention while maintaining autonomy.



Figure 5. Personalized tips based on PHQ-9 score

User Testing

A field trial was conducted with 30 postpartum mothers in Aceh over one week (24 August - 31 September 2025). Participants interacted with the chatbot independently according to their needs. Twelve participants completed the System Usability Scale (SUS) assessment.

Although the response rate was limited relative to the total trial sample, recent usability research indicates that a sample size of approximately 10-15 participants is generally sufficient to obtain stable SUS estimates in

*name of corresponding author



formative system evaluations (Lorincz et al., 2026). Therefore, the obtained SUS sample ($n = 12$) falls within the recommended range for reliable usability assessment.

The system achieved a mean SUS score of 85.63 ($SD = 12.39$), with scores ranging from 65 to 100 (Median = 90). According to established SUS interpretation guidelines, scores above 80 are categorized as "excellent" usability. Eight of the twelve participants (66.7%) rated the system above 80, indicating strong perceived usability. Only one participant scored below the commonly accepted usability threshold of 68.

The observed standard deviation ($SD = 12.39$) indicates moderate variability in user perception, suggesting that while overall usability was rated positively, individual experiences were not entirely uniform. This dispersion may reflect differences in digital literacy, prior experience with chatbot systems, or variation in individual postpartum conditions.

The relatively high usability ratings may be attributed to three key design factors. First, the integration of a RAG mechanism enhances contextual relevance and reduces ambiguous responses. Second, the incorporation of culturally adapted empathetic language likely improves emotional acceptance and comfort during interaction. Third, the simplified interface structure minimizes cognitive load and navigation complexity, which is particularly important for postpartum users who may experience fatigue or emotional vulnerability.

Nevertheless, the results should be considered preliminary and exploratory. Further evaluation with a larger and more diverse sample is required to confirm usability stability and assess long-term engagement patterns.

DISCUSSIONS

The development and evaluation of the SeudaraBunda application demonstrate the potential of an AI-driven empathetic chatbot to support postpartum mental health in Indonesia, particularly in culturally specific contexts such as Aceh. The results from expert validation, prototype demonstration, and user testing highlight several key findings regarding usability, clinical appropriateness, and cultural relevance.

Usability and User Perception

The System Usability Scale (SUS) assessment yielded an average score of 85.63, placing the system in the "excellent" usability category. This finding indicates that postpartum mothers perceived the SeudaraBunda application as easy to use, supportive, and engaging. The high usability likely stems from multiple factors: the empathetic conversational responses provided by Meurabot (Fig. 4), personalized recommendations based on PHQ-9 assessments (Fig. 5), and a culturally and linguistically adapted interface tailored to Acehnese norms. These results align with prior studies suggesting that AI-based chatbots can enhance psychological well-being among postpartum mothers when designed to be user-friendly and contextually relevant (Ahtisham et al., 2023; Suharwardy et al., 2023). However, while high usability indicates positive user acceptance, it does not necessarily equate to clinical effectiveness. Future longitudinal studies are required to determine whether sustained interaction leads to measurable reductions in depressive symptoms.

Clinical and Psychological Validity

Expert validation by a psychiatrist, psychologist, and obstetrician-gynecologist confirmed that the system content is clinically sound, empathetic, and accurate. The integration of the PHQ-9 screening tool enables early identification of depressive symptoms while preserving user autonomy. Unlike traditional questionnaire-based screening, SeudaraBunda combines a retrieval-augmented generation (RAG) approach with large language models (LLMs) to generate dynamic, context-sensitive responses, supporting a more interactive and personalized experience. This approach addresses known limitations in early PPD detection, including barriers due to social stigma, limited mental health literacy, and reduced access to trained professionals in low-resource settings (Upadhyay et al., 2017; Howard et al., 2014; Jones, 2022).

Nevertheless, it is important to critically acknowledge that chatbot-generated responses, even when clinically informed, do not constitute medical diagnosis or psychotherapy. The system provides early screening support and emotional assistance but does not replace professional clinical evaluation. This system is not intended to replace professional diagnosis or therapy, but rather to support early screening and emotional assistance. Clear referral pathways and crisis escalation mechanisms remain essential components for safe deployment.

Cultural and Linguistic Adaptation

A distinctive contribution of SeudaraBunda lies in its cultural and linguistic localization. The incorporation of Acehnese-friendly salutations, culturally appropriate expressions, and sensitivity to religious values enhances perceived trust and reliability. Cultural congruence is particularly significant in postpartum mental health, where emotional experiences are often framed within family expectations, spiritual beliefs, and community norms.

Compared to prior Indonesian chatbot initiatives that focus primarily on postnatal education through rule-based flows (Vinarti et al., 2024), SeudaraBunda emphasizes adaptive emotional dialogue and early mental health

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

screening. This highlights the importance of moving beyond informational chatbots toward emotionally intelligent systems in maternal health contexts.

However, cultural adaptation also presents trade-offs. Over-localization may limit scalability across heterogeneous Indonesian populations, and linguistic nuances may affect model generalization across dialects. Further research is necessary to evaluate cross-regional adaptation and ensure ethical alignment across diverse sociocultural settings.

Ethical and Safety Considerations

The use of LLM-based conversational systems in mental health contexts raises ethical considerations related to privacy, misinformation risk, emotional dependency, and crisis management. Although the RAG architecture reduces hallucination risk by grounding responses in curated knowledge sources, the possibility of unintended outputs cannot be fully eliminated.

There is also a risk that users may attribute excessive authority or therapeutic legitimacy to the chatbot. To mitigate this, the system should always make clear that it offers support, not a diagnosis, and advise users to seek professional help if moderate or severe symptoms appear. Future iterations should incorporate automated risk escalation protocols for suicidal ideation and integrate direct referral pathways to licensed healthcare providers.

Limitations

Several limitations should be acknowledged. First, the field trial involved a relatively small sample size ($n = 30$) and a short observation period (1 week), limiting generalizability and preventing evaluation of long-term mental health outcomes. Second, reliance on self-reported PHQ-9 responses introduces potential response bias and social desirability effects. Third, cultural adaptation was limited to the Aceh context; further validation is required in other Indonesian regions with different linguistic and sociocultural backgrounds. Finally, system performance indicators such as response latency, engagement frequency, and dropout rates were not quantitatively analyzed.

CONCLUSION

The development and evaluation of the SeudaraBunda application demonstrate that an AI-driven empathic chatbot can effectively support postpartum mental health in Indonesia, particularly in culturally specific contexts such as Aceh. Expert validation confirmed that the system content is clinically sound, empathetic, and accurate, while prototype testing and user trials highlighted its high usability, acceptability, and potential for engagement among postpartum mothers. SeudaraBunda's integration of PHQ-9-based emotional state assessment with RAG and a LLM enables dynamic, context-sensitive interactions that support early detection of depressive symptoms while maintaining user autonomy. Moreover, the system's linguistic and cultural adaptations, including localized salutations, culturally appropriate idioms, and attention to religious norms, enhance trust, engagement, and user comfort, distinguishing it from prior rule-based digital interventions. These findings suggest that AI-driven chatbots like SeudaraBunda can serve as scalable, personalized, and culturally relevant solutions for digital mental health support.

Future work will include longitudinal evaluation to assess sustained user engagement and impact on postpartum mental health outcomes over time. Additionally, comparisons with non-AI interventions, such as traditional counseling or rule-based chatbots, will be conducted to better understand the added value of AI-driven conversational support.

ACKNOWLEDGMENT

The author(s) wish to express gratitude for funding support from the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia through the 2025 fiscal year program. We also extend our appreciation to the LPPM of Universitas Islam Kebangsaan Indonesia for their ongoing support and facilitation throughout this research.

REFERENCES

- Abd-Alrazaq, A. A., Alajlani, M., Ali, N., Denecke, K., Bewick, B. M., & Househ, M. (2021). Perceptions and opinions of patients about mental health chatbots: Scoping review. *Journal of Medical Internet Research*, 23(1), e17828. <https://doi.org/10.2196/17828>
- Ahtisham, B., Tausif, S., Bhatti, Z. H., Masood, A., & Shahid, S. (2023). An AI chat-based solution aimed to screen postpartum depression. In *IFIP Conference on Human-Computer Interaction* (pp. 319–328). Cham: Springer Nature Switzerland.
- Boothby, N., Veatch, M., & Pentes, M. (2011). Evaluating treatment of Axis I mental health disorders in Aceh, Indonesia. *The Psychiatrist*, 35(7), 248–255. <https://doi.org/10.1192/pb.bp.110.034777>
- Brännström, A., Wester, J., & Nieves, J. C. (2024). A formal understanding of computational empathy in interactive agents. *Cognitive Systems Research*, 85, 101203. <https://doi.org/10.1016/j.cogsys.2023.101203>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor & Francis.
- Ernawati, E. (2020). Faktor yang berhubungan dengan kejadian post partum blues pada ibu nifas di Ruang Nuri Rumah Sakit Bhayangkara Makassar. *Media Keperawatan*, 11(1), 25–30.
- Howard, L. M., Molyneaux, E., Dennis, C. L., Rochat, T., Stein, A., & Milgrom, J. (2014). Non-psychotic mental disorders in the perinatal period. *The Lancet*, 384(9956), 1775–1788. [https://doi.org/10.1016/S0140-6736\(14\)61276-9](https://doi.org/10.1016/S0140-6736(14)61276-9)
- Jones, A. (2022). Postpartum help-seeking: The role of stigma and mental health literacy. *Maternal and Child Health Journal*, 26(5), 1030–1037. <https://doi.org/10.1007/s10995-022-03399-1>
- Kaur, P., Sharma, R., & Singh, M. (2024). Perinatal mental detection with chatbots using machine learning. In *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 1498–1503). IEEE. <https://doi.org/10.1109/ICTACS57996.2024.00031>
- Khan, Z., Yang, J., Jawad, R., Martinez, I., Hoque, M., Zhao, X., & Samuel, J. (2025, May). MomCare with AI: A dual embedding-based RAG-LLM chatbot for postpartum depression. In *International Conference on Information and Communication Technology for Intelligent Systems* (pp. 459–477). Singapore: Springer Nature Singapore.
- Lorincz, J., Barišić, K., & Vlahović, V. (2026). Usability testing and the System Usability Scale effectiveness assessment on different sensing devices of prototype and live web system counterpart. *Sensors*, 26(2), 679. <https://doi.org/10.3390/s26020679>
- Mane, H. Y., Doig, A. C., Gutierrez, F. X., Jasczynski, M., Yue, X., Srikanth, N. P., Mane, S., Sun, A., Moats, R. A., Patel, P., & He, X. (2023). Practical guidance for the development of Rosie, a health education question-and-answer chatbot for new mothers. *Journal of Public Health Management and Practice*, 29(5), 663–670. <https://doi.org/10.1097/PHH.0000000000001800>
- Mawarpury, M., Sari, K., & Safrina, L. (2017). Layanan kesehatan mental di puskesmas: Apakah dibutuhkan? *Insight: Jurnal Pemikiran dan Penelitian Psikologi*, 13(1), 1–10.
- Mitchell, A. R., Gordon, H., Lindquist, A., Walker, S. P., Homer, C. S., Middleton, A., et al. (2023). Prevalence of perinatal depression in low- and middle-income countries: A systematic review and meta-analysis. *JAMA Psychiatry*, 80(5), 425–431. <https://doi.org/10.1001/jamapsychiatry.2023.0205>
- O’Hara, M. W., & McCabe, J. E. (2013). Postpartum depression: Current status and future directions. *Annual Review of Clinical Psychology*, 9(1), 379–407. <https://doi.org/10.1146/annurev-clinpsy-050212-185612>
- Rickwood, D. J., & Coleman-Rose, C. L. (2023). The effect of survey administration mode on youth mental health measures: Social desirability bias and sensitive questions. *Heliyon*, 9(9), e14902. <https://doi.org/10.1016/j.heliyon.2023.e14902>
- Suharwardy, S., Ramachandran, M., Leonard, S. A., Gunaseelan, A., Lyell, D. J., Darcy, A., et al. (2023). Feasibility and impact of a mental health chatbot on postpartum mental health: A randomized controlled trial. *AJOG Global Reports*, 3(3), 100165. <https://doi.org/10.1016/j.xagr.2023.100165>
- Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., et al. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, 20(3), 318–335. <https://doi.org/10.1002/wps.20883>
- Tsai, Z., Kiss, A., Nadeem, S., Sidhom, K., Owais, S., Faltyn, M., & Van Lieshout, R. J. (2022). Evaluating the effectiveness and quality of mobile applications for perinatal depression and anxiety: A systematic review and meta-analysis. *Journal of Affective Disorders*, 296, 443–453. <https://doi.org/10.1016/j.jad.2021.09.020>
- Upadhyay, R. P., Chowdhury, R., Salehi, A., Sarkar, K., Singh, S. K., Sinha, B., et al., & Kumar, A. (2017). Postpartum depression in India: A systematic review and meta-analysis. *Bulletin of the World Health Organization*, 95(10), 706–717. <https://doi.org/10.2471/BLT.17.200139>
- Utami, M. S. M., Kwok, W. H., Kotz, J., Walker, R., Wang, G., & Marriott, R. (2025, July). Facilitating Aboriginal perinatal mental health information access with a retrieval-augmented LLM-based chatbot. In *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1–7). IEEE.
- Vinarti, R. A., Sani, N. A., Amalia, R., Sulistyani, E., Noveyani, A. E., Suryaputra, E., & Arwan, M. A. (2024). KIA-CHAT: A QnA chatbot for postnatal and newborn care. *Jurnal Promkes: The Indonesian Journal of Health Promotion and Health Education*, 12(1 SP), 97–101. <https://doi.org/10.20473/jpk.V12.I1SP.2024.97-101>
- World Health Organization. (2025). Perinatal mental health. Geneva: World Health Organization. Retrieved April 10, 2025, from <https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/maternal-mental-health>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.