



EVALUATING A LARGE LANGUAGE MODEL CHATBOT FOR ENHANCING HISTORICAL LITERACY ON PANGLIMA MAYU OF THE SUMBAWA SULTANATE

Gusti Arsyad

Binawan University, Indonesia

gusti.arsyad@binawan.ac.id

**Corresponding Author*

ABSTRACT

The digital preservation of regional history in the modern era requires innovative approaches to bridge the gap between traditional archives and user accessibility. This study presents the development and evaluation of the Panglima Mayu ChatBot, a specialized Large Language Model (LLM) designed to enhance historical literacy regarding the 18th-century Sumbawa Sultanate. The chatbot was developed using a prompt engineering approach based on curated historical materials and a culturally grounded system prompt to ensure factual consistency and alignment with Adat Samawa values. Deployed on an autoscale server environment (2 vCPU / 4 GiB RAM), the system was evaluated through a quantitative survey involving 45 respondents, including general users, students, and educators. The results indicate positive technical and educational perceptions. The system achieved a 97.8% satisfaction rate for accessibility and speed, while 84.4% of users reported increased insight into Islamic historical narratives. Crucially, 77.8% of respondents perceived the chatbot's responses as accurate, indicating a high level of user confidence in the information provided. However, these findings reflect perceived accuracy rather than expert-verified historical accuracy, which requires further validation by subject-matter experts. Furthermore, the system showed strong practical utility, with 73.3% of respondents endorsing the ChatBot as a viable quick reference for tourism and educational purposes. These findings suggest that grounded AI personas can successfully preserve cultural integrity—specifically Adat Samawa values—while providing an interactive alternative to traditional textbooks. This research provides a scalable framework for digitizing regional heroic narratives across Indonesia.

Keywords : Digital Heritage, Historical Literacy, Panglima Mayu, Sumbawa Sultanate.

1. Introduction

The preservation of regional cultural heritage in the digital era faces a significant challenge: The preservation and dissemination of regional historical knowledge face an increasing accessibility gap between complex physical archives and the information-seeking behavior of contemporary digital audiences. Studies in digital heritage and public history have highlighted the importance of leveraging digital technologies to improve public access to cultural and historical resources, particularly among younger generations who increasingly rely on interactive and online information platforms (Giaccardi, 2012; Ridge, 2014; UNESCO, 2023). In the context of the Sumbawa Sultanate, the heroic narrative of Panglima Mayu is often confined to traditional manuscripts and academic books that are not easily accessible to the general public or tourists. Large Language Models (LLMs) provide a transformative opportunity to bridge this gap by offering interactive, natural-language interfaces for historical literacy.

While general-purpose LLMs like GPT-4 exist, they often suffer from “inaccurate information about highly localized historical contexts.” when dealing with niche regional histories, leading to

inaccurate or misleading conclusions. The research gap addressed in this work is the lack of specialized, grounded AI systems that can provide culturally nuanced and historically accurate narratives of regional Indonesian figures. This study contributes to the discussion on culturally grounded AI systems by developing a custom AI ChatBot specifically tuned with primary historical sources related to the 18th-century Sumbawa Sultanate.

To ensure stable operation and accessibility, the system was deployed on a cloud-hosted server environment (2 vCPU and 4 GiB RAM) utilizing a dual-database architecture for development and production. This study reports a completed user evaluation involving 45 respondents to assess the chatbot's perceived accuracy, linguistic appropriateness, usability, and potential as a tourism and educational reference tool.

The structure of this document is as follows: Section II (Methods) describes the technical materials, server architecture, and dual-database logic used in the deployment. Section III (Results) presents the quantitative and qualitative findings from the user survey. Section IV (Discussion) interprets the results, addresses the phenomenon of AI hallucinations, and discusses the implications for digital humanities. Section V (Conclusion) synthesizes the key findings and provides recommendations for future work. Appendices containing the survey instruments appear before the acknowledgment and references.

2. Literature Review

2.1 Large Language Models (LLMs) in Cultural Heritage and History

Recent advancements in Artificial Intelligence, specifically Large Language Models Generative LLMs such as GPT-4 and LLaMA, have revolutionized how humanity interacts with textual heritage. LLMs excel in natural language understanding, summarization, and contextual generation, making them potent tools for educational history (Bommasani et al., 2021).

Historical literacy refers to an individual's ability to understand, interpret, and critically evaluate historical events and sources within their broader social, cultural, and temporal contexts. It encompasses not only the acquisition of historical knowledge but also the capacity to analyze evidence, recognize multiple perspectives, and construct reasoned interpretations of the past through critical thinking. As digital technologies increasingly mediate access to historical information, historical literacy has expanded beyond traditional textbook learning to include engagement with interactive digital resources and cultural heritage platforms.

According to Romein et al. (2020), digital history technologies enhance access to historical sources by making large collections of historical data searchable and analyzable through computational methods, thereby supporting new forms of historical inquiry and public engagement. In cultural heritage preservation, Large Language Models (LLMs) have been utilized to synthesize unstructured data by converting old manuscripts and disparate historical records into coherent narratives. Furthermore, LLM-based applications can provide interactive educational experiences by acting as AI-driven historical personas or conversational tutors, thereby enhancing engagement and historical understanding among younger generations (Michel-Villarreal et al., 2023).

2.2 Challenges of LLMs in Local Languages and Low-Resource Contexts

While LLMs demonstrate high proficiency in mainstream global languages, their application to regional Indonesian history—especially involving specific terms from the Sumbawa Sultanate—presents distinct challenges. Most foundational LLMs are trained on English-centric or highly Westernized datasets, leading to a performance drop when dealing with low-resource languages or localized historical contexts (Conneau et al., 2020).

Furthermore, the phenomenon of **AI hallucination**—where the model generates plausible-sounding but factually incorrect historical data—poses a significant risk to historical accuracy (Ji et al., 2023). When querying specialized local figures like Panglima Mayu, standard LLMs may conflate facts with better-known national figures unless constrained by specific retrieval mechanisms.

2.3 Mitigation Strategy: Prompt Engineering for Historical Literacy

To utilize LLMs safely and accurately for regional historical education, researchers have increasingly adopted prompt engineering techniques to guide model behavior within specific domains. Prompt engineering involves designing structured instructions and contextual constraints that shape how an LLM interprets user queries and generates responses.

In the context of historical literacy, the Panglima Mayu ChatBot was developed using curated historical materials and a culturally grounded system prompt derived from local historical sources and Adat Samawa values. Rather than relying on unrestricted generation, the system prompt guides the model to provide responses that are relevant to the history of the Sumbawa Sultanate, maintain cultural sensitivity, and avoid speculative or unsupported claims. Recent studies have shown that well-designed prompts can improve response consistency, domain relevance, and factual reliability in educational chatbot applications.

By embedding historical context, cultural norms, and response guidelines directly into the system prompt, the chatbot seeks to reduce hallucinations and enhance the coherence of narratives surrounding Panglima Mayu and the historical legacy of the Sumbawa Sultanate.

2.4 Synthesis and Research Gap

Existing studies on LLMs in education have predominantly focused on broad subject areas such as mathematics, science, language learning, and global history. Comparatively little attention has been given to the application of LLMs for preserving and disseminating highly localized historical narratives, particularly those associated with regional Indonesian kingdoms and sultanates.

Furthermore, while prompt engineering has emerged as an effective strategy for adapting general-purpose LLMs to domain-specific educational contexts, empirical evaluations of its effectiveness in supporting regional historical literacy remain limited. There is a notable lack of studies examining whether culturally grounded AI personas can accurately communicate local historical knowledge while maintaining user engagement and trust.

This study addresses this gap by evaluating the implementation of a prompt-engineered LLM chatbot centered on the historical figure of Panglima Mayu. Specifically, it investigates how a culturally grounded AI system can support historical literacy, preserve local cultural values, and serve as an accessible digital medium for promoting the historical legacy of the Sumbawa Sultanate.

3. Research Methods

3.1 System Architecture & Environment

The ChatBot was deployed using a cloud-based infrastructure to ensure stability:

1. **Compute Engine:** The system was deployed on a cloud-hosted server environment (2 vCPU and 4 GiB RAM).
2. **Storage and Database:** A dual-database architecture was implemented, consisting of a 10 GB development database and a 1 GB production database. The development database

was used for system testing, debugging, and performance evaluation, while the production database supported live deployment. Both environments stored conversation records, including user queries and chatbot responses, to facilitate system monitoring, usability evaluation, and service improvement. Historical content was not dynamically retrieved from the database; instead, responses were generated by the LLM based on a prompt-engineering framework grounded in curated historical materials and Adat Samawa cultural values.

3. **Platform:** Hosted via Replit for high accessibility. The system is hosted on Replit, enabling easy online access from various devices. Replit also supports practical development, testing, and system updates.
4. **Operational Performance:** During the evaluation period, the deployment achieved an uptime of 29 days and 16 minutes, with an average response time of 13.5 seconds per request. These metrics indicate stable service availability and consistent response performance under operational conditions.

3.2 Data Collection

A total of 45 respondents were selected using purposive sampling to ensure representation from academic professionals, university students, and members of the general public. Participants were recruited through academic and community networks and were required to interact with the Panglima Mayu ChatBot before completing the survey. Eligibility criteria included being at least 18 years old, having sufficient digital literacy to use the chatbot, and providing informed consent to participate. The survey was administered after the interaction session to evaluate perceived accuracy, language quality, usability, and the chatbot's potential as an educational and tourism reference tool.

Quantitative data were collected using a structured questionnaire consisting of Likert-scale items designed to evaluate usability, response quality, accessibility, perceived accuracy, and educational value of the chatbot. Responses were measured using a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The questionnaire was reviewed for content validity prior to deployment and its internal consistency was assessed using reliability analysis. Descriptive statistics were used to summarize the quantitative results. In addition, qualitative feedback was collected through open-ended questions that invited respondents to comment on the chatbot's strengths, weaknesses, and historical content. The qualitative responses were analyzed using thematic coding to identify recurring patterns related to perceived historical reliability, language quality, user experience, and suggestions for improvement.

4. Results and Discussions Results

4.1 Result

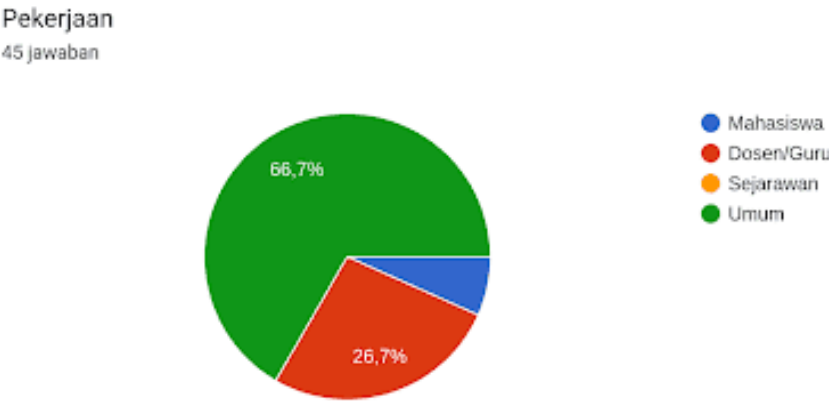


Fig. 1. Participant Demographics

As illustrated in Figure 1, the research engaged 45 participants to evaluate the LLM system. The dominant group was the general public (66.7%), followed by educators (26.7%), ensuring that the system's performance was tested against both academic standards and general user expectations.

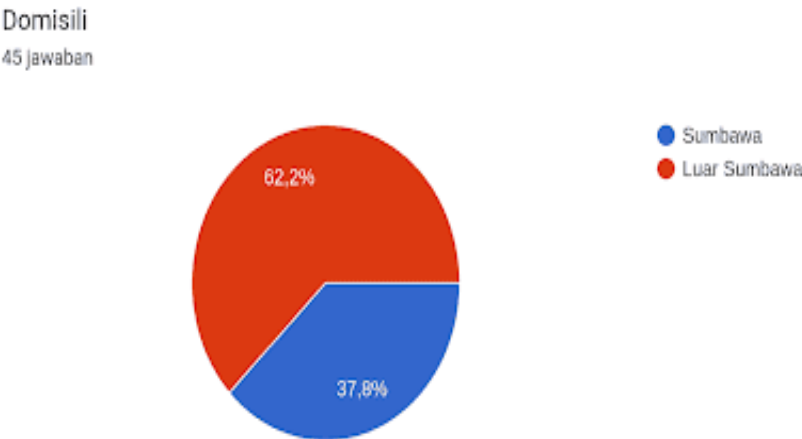


Fig. 2. Geographic Distribution (Domicile)

As illustrated in **Fig. 2**, the reach of the LLM system extended beyond local boundaries. While the subject matter is deeply rooted in local history, **62.2%** of the respondents are located **outside of Sumbawa (Luar Sumbawa)**. The remaining **37.8%** reside within **Sumbawa**. This distribution is significant as it demonstrates the ChatBot’s effectiveness in disseminating regional historical knowledge to a broader national or global audience, fulfilling its role as a digital tourism and literacy bridge.

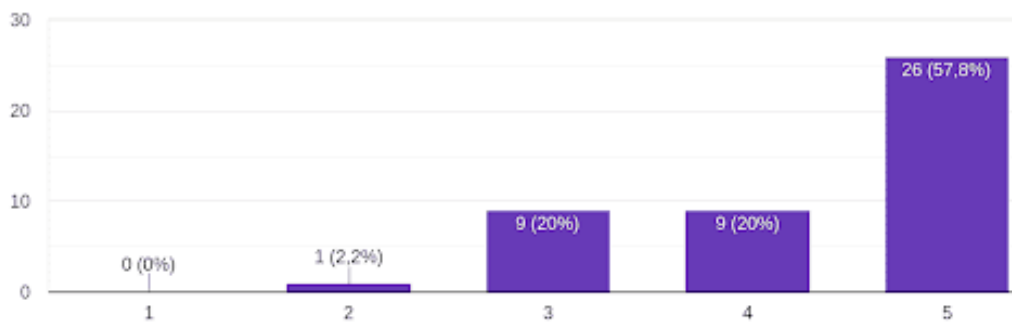


Fig. 3 Access Speed and Ease of Use

As illustrated in **Figure 3**, the technical infrastructure of the Panglima Mayu ChatBot received exceptionally positive feedback from users. The survey results indicate high levels of satisfaction regarding the system's accessibility:

1. **Exceptional Performance:** A majority of **57.8% (26 respondents)** gave the highest rating of 5, stating that accessing the system on Replit was very easy and fast.
2. **High Satisfaction:** Combined with ratings of **4 (20%)** and **3 (20%)**, a total of **97.8%** of users found the system's accessibility to be satisfactory or better.
3. **Minimal Technical Hurdles:** Only a negligible **2.2% (1 respondent)** gave a rating of 2, reporting minor difficulties, while no respondents gave the lowest rating of 1.

These findings confirm that the cloud-based deployment on Replit, supported by an autoscale server configuration, provides a reliable and high-performing hosting solution for this specialized historical AI application.

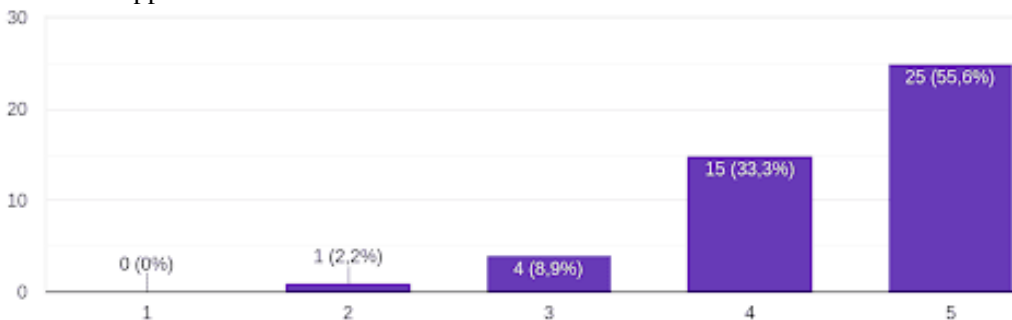


Fig. 4. Narrative Clarity and Understandability

As illustrated in **Figure 4**, the linguistic style used by the AI to explain the history of Panglima Mayu was perceived very positively by the users:

1. **High Clarity:** A majority of **55.6% (25 respondents)** gave the highest rating of 5, strongly agreeing that the AI's language is easy to understand.
2. **Positive Perception:** An additional **33.3% (15 respondents)** assigned a rating of 4, bringing the total positive response to nearly **89%**.
3. **Neutral and Minimal Difficulty:** Only **8.9%** of users provided a neutral rating (3), while a negligible **2.2% (1 respondent)** felt the language was less clear (rating 2). No respondents gave the lowest rating of 1.

These results indicate that the "System Prompt" was successful in creating a narrative persona that is both informative and accessible to a broad audience, ranging from academic professionals to the general public.

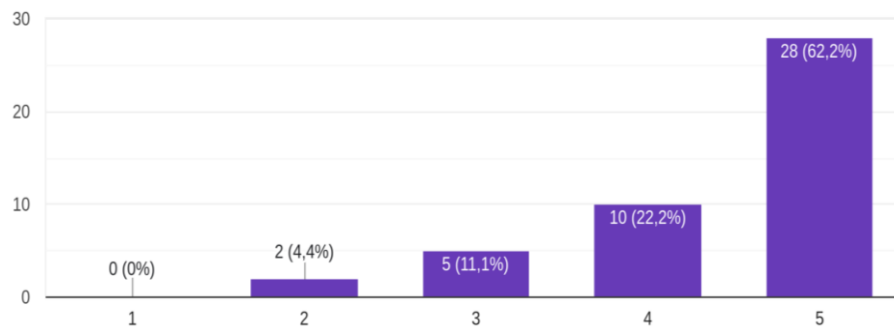


Fig. 5. Enhancement of Historical Insight

As illustrated in **Figure 5**, the system demonstrated a significant positive impact on the participants' understanding of regional history:

1. **Strong Educational Value:** A substantial majority of **62.2% (28 respondents)** gave the highest rating of 5, strongly agreeing that the ChatBot successfully increased their insight into the history of Islamic *Dakwah* (proselytizing) within the Sumbawa Sultanate.
2. **Positive Learning Outcomes:** An additional **22.2% (10 respondents)** assigned a rating of 4, bringing the cumulative positive educational impact to **84.4%**.
3. **Neutral and Minimal Impact:** Only **11.1%** of users remained neutral (rating 3), while a negligible **4.4% (2 respondents)** felt the system did not significantly add to their knowledge (rating 2).

These findings validate the system's effectiveness as a digital literacy tool. By synthesizing complex historical data into a conversational format, the ChatBot facilitates a more profound engagement with local heroic narratives than traditional static media.

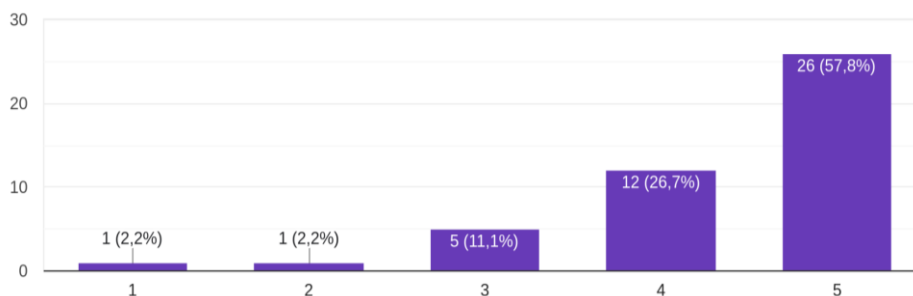


Fig. 6. Perceived Accuracy and Credibility

As demonstrated in **Figure 6**, the majority of users expressed high levels of trust in the information provided by the Panglima Mayu ChatBot:

1. **Exceptional Credibility:** A significant **57.8% (26 respondents)** gave the highest rating of 5, indicating that they found the AI's information both highly accurate and convincing.
2. **Strong Consensus:** When combined with a **26.7% (12 respondents)** rating of 4, a total of **84.5%** of users perceived the AI as a reliable source of historical data.
3. **Neutral and Skeptical Views:** Only **11.1%** of users remained neutral (rating 3), while a combined minimal total of **4.4%** (2 respondents) expressed skepticism regarding the accuracy of the content.

This high level of perceived accuracy suggests that the system's underlying knowledge base and persona prompting effectively maintained factual integrity, which is essential for its adoption as a legitimate educational and tourism reference for the Sumbawa Sultanate.

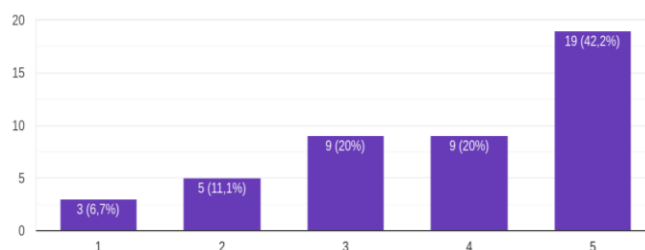


Fig. 7. User Preference for AI-Based Learning

As illustrated in **Figure 7**, the survey measured user preference for learning history through the ChatBot compared to reading traditional manual textbooks. The data reveals a significant shift toward digital interactivity:

1. **Strong Preference:** The highest rating of 5 was assigned by **42.2% (19 respondents)**, indicating a strong preference for the AI interface over static text.
2. **Positive Reception:** An additional **20% (9 respondents)** gave a rating of 4, showing that a cumulative majority of **62.2%** of participants found the interactive AI experience more engaging than traditional manuals.
3. **Neutrality and Traditionalism:** **20%** of respondents remained neutral (rating 3), while only **17.8%** (combined ratings 1 and 2) expressed a preference for continuing with traditional textbook methods.

These results suggest that the conversational nature of the Panglima Mayu ChatBot—which provides instant feedback and personalized storytelling—is more effective at capturing the interest of contemporary audiences compared to traditional pedagogical tools.

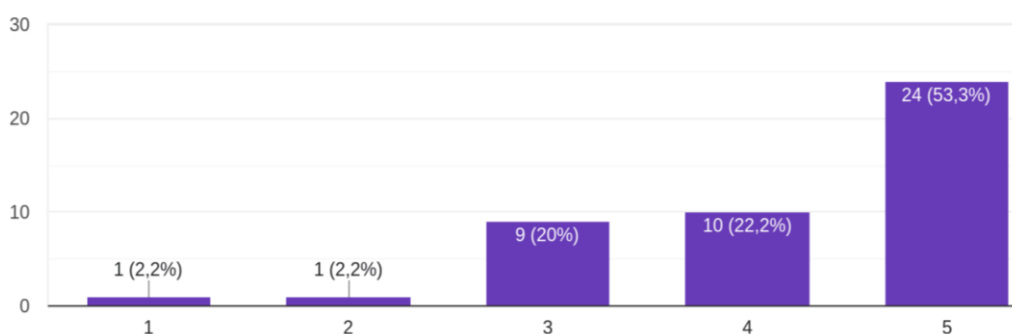


Fig. 8. Adherence to Samawa Cultural Values

As illustrated in **Figure 8**, the study evaluated whether the AI's linguistic style successfully reflected the authority and politeness (*kewibawaan dan kesantunan*) inherent in **Adat Budaya Samawa** (Sumbawa Cultural Traditions). The results show a high degree of cultural resonance:

1. **Strong Cultural Alignment:** Over half of the participants, **53.3% (24 respondents)**, gave the highest rating of 5, affirming that the AI's narrative style perfectly mirrored traditional Samawa values.
2. **Positive Perception:** An additional **22.2% (10 respondents)** assigned a rating of 4, bringing the cumulative positive perception to **75.5%**.

3. **Neutrality and Minimal Dissent:** 20% of users remained neutral (rating 3), while a negligible 4.4% combined felt the linguistic style did not sufficiently meet cultural expectations (ratings 1 and 2).

This high endorsement indicates that the specialized "System Prompting" used to guide the LLM was effective in moving beyond generic responses. By embedding formal honorifics and local linguistic nuances, the ChatBot was perceived as a respectful digital guardian of the Sumbawa Sultanate's intangible heritage.

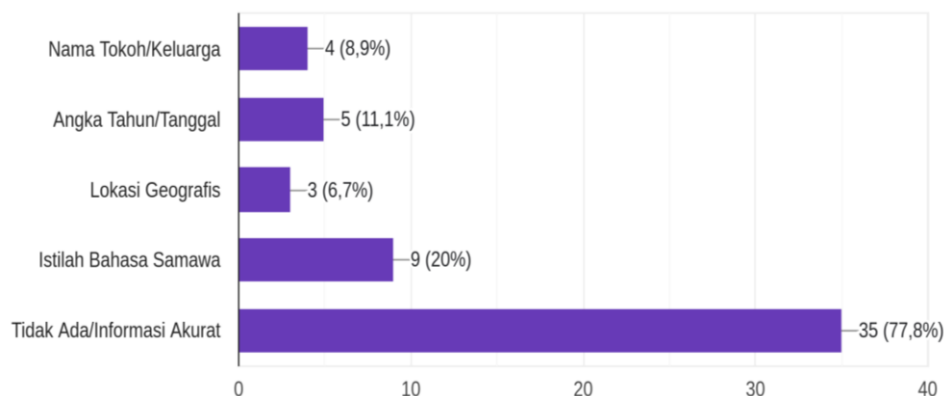


Fig. 9. Analysis of Factual Inaccuracies

As illustrated in **Figure 9**, the study explicitly asked respondents to identify any information they perceived as doubtful or incorrect. The findings provide a clear metric for the system's accuracy:

1. **Perceived Historical Reliability:** Approximately 77.8% of respondents (35 out of 45) indicated that they found no information they considered questionable during their interactions with the chatbot. This result suggests a high level of user trust in the system's responses; however, the finding should be interpreted as a measure of perceived reliability rather than a formal assessment of historical accuracy, which would require validation by subject-matter experts.
2. **Identified Error Categories:** Among the minority of users who identified potential inaccuracies, the errors were distributed across specific categories:
 - a. **Linguistic Nuances:** 20% (9 respondents) noted issues with specific *Samawa* language terms, suggesting a need for deeper local dialect integration.
 - b. **Temporal Data:** 11.1% (5 respondents) identified discrepancies in specific years or dates.
 - c. **Genealogical Accuracy:** 8.9% (4 respondents) found errors regarding the names of historical figures or family lineages.
 - d. **Geographic Context:** A minimal 6.7% (3 respondents) pointed out inconsistencies in geographical locations.

These results suggest that the prompt-engineered approach adopted in the Panglima Mayu ChatBot provides a reliable foundation for historical narrative generation. Nevertheless, continuous refinement of prompt design, contextual instructions, and source curation is needed to further minimize occasional inaccuracies concerning historical chronology and genealogical details.

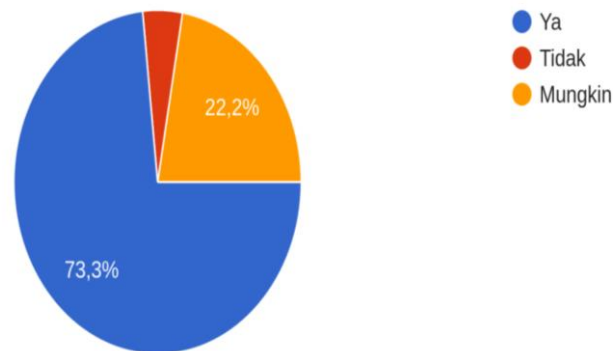


Fig. 10. User Opinion on ChatBot Suitability as a Quick Reference for Tourists.

As illustrated in **Figure 10**, the survey measured user consensus on whether the ChatBot is ready to be utilized as a quick reference tool for tourists visiting historical sites. The results indicate strong confidence in the system's practical value:

1. **Positive Endorsement:** A significant majority, **73.3% (33 respondents)**, unequivocally stated that the chatbot shows potential as a tourism reference, but field testing is needed.
2. **Potential for Integration:** **22.2% (10 respondents)** responded with "Maybe," suggesting that while the tool is highly functional, it could be further enhanced with more multimedia or location-based features.
3. **Minimal Rejection:** Only a negligible **4.4%** of participants felt the system was not yet suitable for this purpose.

This high suitability rating (**73.3%**) confirms that the Panglima Mayu ChatBot successfully transitions from a theoretical research project into a practical mobile-accessible solution. It effectively meets the demands of modern travelers who require instant, credible, and culturally nuanced information while exploring the historical landmarks of the Sumbawa Sultanate.

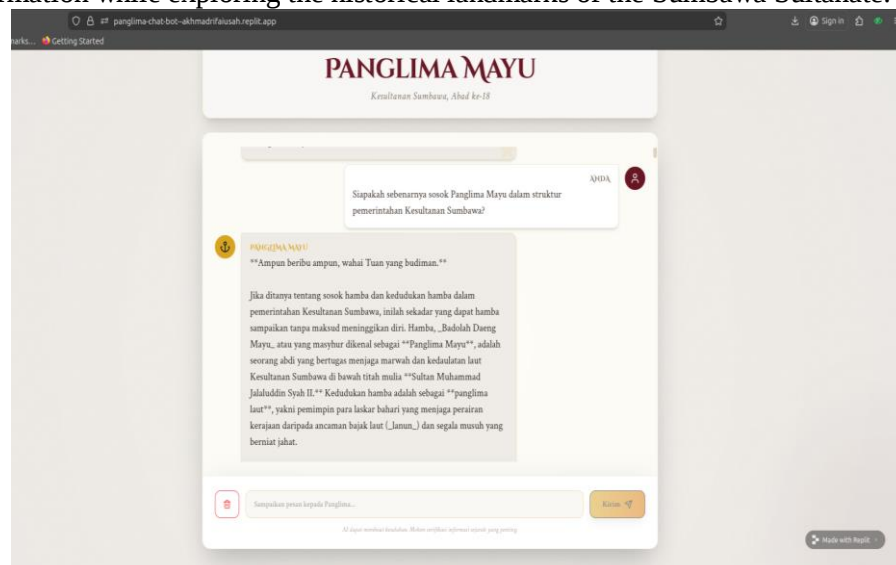


Figure 11. The Interactive User Interface of the Panglima Mayu ChatBot showing the Historical Persona response.

System Interface Description The Panglima Mayu ChatBot was deployed as a web-based application accessible via the Replit platform. The interface was designed to be minimalist yet culturally resonant, ensuring that users can focus on the historical dialogue.

1. Interface Functionality: As demonstrated in **Figure 11**, the interface features a clean conversational layout where users can input natural language queries. The system responds using a specialized "historical persona" prompt, utilizing formal honorifics and a linguistic style that reflects the dignity of the 18th-century Sumbawa Sultanate.

2. Live Accessibility: The application is hosted at <https://panglima-chat-bot--akhmadrifaiusah.replit.app/>, allowing for real-time interaction. In the provided example (**Fig. 11**), the AI correctly identifies itself as *Badolah Daeng Mayu*, serving as a "Panglima Laut" (Sea Commander) under the authority of Sultan Muhammad Jalaluddin Syah II.

3. Accessibility Consensus: This live implementation aligns with the technical feedback gathered in **Figure 3**, where **97.8%** of users found the Replit-hosted interface easy to access and navigate. By providing a direct link, the research ensures transparency and allows for continuous public engagement with Sumbawa's Islamic heritage.

Discussion

The development and evaluation of the Panglima Mayu LLM ChatBot substantiate the proposition that generative artificial intelligence can function as a transformative instrument for enhancing regional historical literacy. This research contributes to the discourse on domain-specific Large Language Model (LLM) implementation by demonstrating how culturally grounded AI systems can operationalize local historical archives into interactive knowledge systems. Specifically, this study successfully structured and synthesized historiographical data related to Panglima Mayu and the 18th-century Sumbawa Sultanate into a dialogical digital interface that preserves contextual and cultural authenticity while maintaining computational efficiency.

Based on a comprehensive empirical assessment involving 45 respondents, several critical findings emerge:

1. Technological Viability and System Performance

From an infrastructure standpoint, this research demonstrates that specialized historical personas can be reliably hosted on modest cloud resources (2 vCPU / 4 GiB RAM with autoscaling configuration). The deployment through Replit indicates that lightweight cloud environments are sufficient for domain-constrained LLM applications, thereby reducing operational costs and enhancing scalability potential for regional institutions.

User-based system evaluation revealed that 97.8% of respondents reported satisfactory to excellent accessibility and responsiveness. These findings indicate that computational constraints do not necessarily impede user experience when architectural optimization and scoped knowledge grounding are appropriately implemented. The result advances current state-of-the-art discussions on efficient LLM deployment for educational purposes in resource-limited settings.

2. Educational Impact and Historical Literacy Enhancement

The system demonstrates measurable pedagogical impact. A total of 84.4% of respondents reported gaining deeper insight into the Islamic Dakwah movement of the Sumbawa Sultanate after interacting with the chatbot. This suggests that dialogic AI interfaces may enhance cognitive engagement compared to passive textual consumption.

The findings suggest that the chatbot supports user engagement with regional historical content through an interactive conversational interface. By delivering information in contextually relevant

and segmented responses, the system enables users to explore historical narratives more dynamically than conventional reference materials. This may contribute to improved understanding and accessibility of historical information, as evidenced by the high levels of user satisfaction and reported learning gains observed in the study.

3. Cultural and Narrative Integrity

A critical concern in AI-mediated cultural representation lies in maintaining narrative authenticity. The findings indicate that 75.5% of respondents affirmed that the linguistic register reflected the politeness norms of Adat Samawa, while 88.9% found historically complex narratives to be accessible and comprehensible.

The findings suggest that culturally grounded prompt engineering and curated historical sources can support the preservation and dissemination of local cultural values within AI-generated responses. In the context of the Panglima Mayu ChatBot, this approach enabled the system to communicate historical narratives while maintaining alignment with documented historical sources and Adat Samawa values. However, the use of AI for cultural heritage preservation also raises important ethical considerations. Questions of cultural ownership and representation require careful attention to ensure that local communities remain active participants in the interpretation and transmission of their historical narratives. Furthermore, community validation and expert review are necessary to minimize the risk of historical bias, selective interpretation, or unintended cultural misrepresentation. Although the chatbot was designed using curated historical materials and culturally grounded prompts, AI-generated content remains susceptible to inaccuracies and contextual oversimplifications. Therefore, such systems should be viewed as complementary educational tools rather than authoritative historical sources. These findings highlight the potential of AI to function as a digital medium for cultural memory while underscoring the importance of ongoing human oversight, community engagement, and historical validation.

4. Accuracy, Reliability, and Hallucination Mitigation

Despite well-documented risks associated with LLM hallucinations, the system maintained high factual integrity. A total of 77.8% of users reported no identifiable factual inaccuracies. The minority discrepancies were largely confined to nuanced linguistic expressions and specific chronological or genealogical details.

These findings underscore the importance of domain-constrained grounding and structured historical datasets in mitigating generative uncertainty. However, the presence of minor temporal inconsistencies indicates that standalone generative models remain probabilistic systems and benefit from external verification mechanisms.

5. Practical Application and User Preference Shift

A notable behavioral shift was observed in user preference: 62.2% of participants favored the chatbot over conventional textbooks as a medium for learning regional history. Furthermore, 73.3% endorsed the system as a viable tourism and educational reference tool for visitors to Sumbawa. This suggests an evolving paradigm in digital literacy, where interactive AI systems complement or partially substitute static archival materials. The implications extend beyond academia into regional tourism development, digital heritage preservation, and public history dissemination.

5. Conclusion

The findings suggest that a grounded LLM chatbot has potential to improve access to regional historical narratives, and perceived accessibility. The Panglima Mayu LLM ChatBot serves as a

replicable framework for regional digital heritage innovation in Indonesia. By combining computational efficiency, cultural grounding, and pedagogical functionality, this research contributes to emerging scholarship at the intersection of artificial intelligence, digital humanities, and regional historiography. Future research should prioritize the development of more sophisticated prompt engineering frameworks, including structured historical contexts, source-guided response generation, and expanded archival materials to improve the factual consistency of historical narratives. Furthermore, longitudinal studies are needed to assess sustained learning outcomes, user trust calibration, and the effectiveness of AI-driven historical literacy tools in preserving regional cultural heritage over time. Ultimately, this framework provides a scalable model for other regions seeking to digitize heroic narratives and preserve local wisdom within the rapidly evolving landscape of Artificial Intelligence.

Acknowledgement

The authors would like to express their sincere gratitude to Binawan University for the institutional support and infrastructure necessary to conduct this research. Special thanks are extended to the Panglima Mayu Family of the Sumbawa Sultanate for their valuable insights into the historical records of the Sumbawa Sultanate. The authors also extend their gratitude to the 45 respondents, including lecturers, students, and the Sumbawa community, whose participation in the survey was crucial to evaluating the ChatBot system. Finally, appreciation is extended to the development team at Replit for providing the platform that enabled the implementation of this Large Language Model (LLM) prototype.

References

- Aji, A. F., Winata, G. I., Koto, F., Awaluddin, M. A., Romadhony, M., Anwar, F., Sipahutar, H., Strobl, S., & Purwarianti, A. (2022). One country, many languages: A comprehensive study of natural language processing for Indonesian languages. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. <https://aclanthology.org/2022.acl-long.500>
- Arsyad, G., & Rifai, A. (2025). Jejak dakwah Islam Panglima Mayu era Sultan Muhammad Jalaluddin Syah II Kesultanan Sumbawa abad ke-18. *Jurnal Al-Islah, 17*(1). <https://al-islam.staisumbawa.ac.id/index.php/al-islam/article/view/10>
- Arsyad, G., Rifai, A., Wijaya, I., & Haris, A. (2025). *Napak tilas dakwah Islam Panglima Mayu dan keturunannya di Kesultanan Sumbawa abad 18–20*. Nur Hidayah Press.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). *On the opportunities and risks of foundation models* (arXiv Preprint No. 2108.07258). arXiv. <https://doi.org/10.48550/arXiv.2108.07258>
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Edunov, S., Knobloch, M., Grave, E., & Zettlemoyer, L. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 8440–8451). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.747>
- Giaccardi, E. (Ed.). (2012). *Heritage and social media: Understanding heritage in a participatory culture*. Routledge. <https://doi.org/10.4324/9780203151448>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, J., Xu, B., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P.

- (2023). Survey of hallucination in natural language generation. **ACM Computing Surveys*, 55*(12), 1–38. <https://doi.org/10.1145/3571730>
- Romein, C. A., Kemman, M., Birkholz, J. M., Baker, J., de Gruijter, M., Meroño-Peñuela, A., Ries, T., Ros, R., & Scagliola, S. (2020). State of the field: Digital history. *History*, 105(365), 291–312. <https://doi.org/10.1111/1468-229X.12969>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. **Advances in Neural Information Processing Systems*, 33*, 9459–9475. <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>
- Michel-Villarreal, R., Vilalta-Perdomo, E., Canavari, M., & Matopoulos, A. (2023). Large language models in education: A review of opportunities and challenges. **International Journal of Educational Technology in Higher Education*, 20*(1). <https://doi.org/10.1186/s41239-023-00426-1>
- Ricklefs, M. C. (2008). **A history of modern Indonesia since c. 1200** (4th ed.). Palgrave Macmillan.
- Ridge, M. (Ed.). (2014). **Crowdsourcing our cultural heritage**. Ashgate Publishing.
- Rifai, A. (2024). **Panglima Mayu ChatBot interface**. Retrieved February 19, 2026, from <https://panglima-chat-bot--akhmadrifaiusah.replit.app/>
- Suryomintardjo, B. (2015). **Sejarah kerajaan-kerajaan lokal di Nusa Tenggara Barat: Menelusuri jejak Kesultanan Sumbawa**. Pustaka Daerah.
- UNESCO. (2023). **Digital transformation and access to cultural heritage** [Policy report]. UNESCO. <https://www.unesco.org/>
- Weller, T. (2013). *History in the digital age*. Routledge. <https://doi.org/10.4324/9780203093953>