Multilingual Heritage Assistant Using A Recurrent Location-Aware Transformer Approach

Ramalingam T
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering
Sriperumbudur, India

Gousiga V S
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering
Sriperumbudur, India

Harish K
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering
Sriperumbudur, India t

Bharat Kumar R J
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering
Sriperumbudur, India

***Abstract*—**In the rapidly evolving landscape of cultural tourism, offering real-time, multilingual, and personalized in- sights plays a pivotal role in enhancing the visitor experience, especially at historically rich and diverse locations such as those found across Tamil Nadu. This project presents an AI-powered voice assistant mobile application specifically designed to make heritage exploration more engaging, educational, and accessible for a wide range of users, including domestic and interna- tional tourists. The system integrates advanced natural language processing and machine learning technologies to interpret user queries in multiple Indian languages, whether spoken or typed, and respond with contextually relevant heritage information. At its core, the application employs a fine-tuned Mistral 7B trans- former model to generate dynamic, human-like responses across various languages, while OpenAI's Whisper model ensures high- accuracy speech-to-text conversion for real-time voice interaction. Complementing this, the Google Text-to-Speech (gTTS) engine synthesizes natural-sounding audio responses, making the inter- face fully voice-driven and inclusive. Personalization is achieved through a GRU-based behavioral model that tracks and analyzes sequential user interactions, learning individual preferences to deliver tailored content and recommendations. The application incorporates real-time geolocation data and applies the Haversine distance formula to determine the user's proximity to various heritage sites, thereby offering hyperlocal suggestions and site- specific content. All user preferences, historical interactions, and metadata are efficiently managed using an SQLite database, ensuring smooth offline support and optimized performance in resource-constrained environments. Drawing from a structured heritage dataset in JSON format, the assistant provides rich narratives, historical facts, travel guidance, and site-specific tips, creating an immersive cultural experience.

***Index Terms*—**Cultural tourism, AI-powered voice assistant, real-time multilingual insights, personalized tours, location-aware algorithms.

I. INTRODUCTION

CULTURAL tourism has become an essential part of the global tourism industry, offering visitors a profound opportunity to engage with history, art, and cultural heritage. Heritage sites, ranging from ancient monuments to modern landmarks, attract millions of tourists every year [1]. However, the rapid increase in the number of tourists visiting these sites has resulted in challenges, including the difficulty of offering personalized, immersive, and accessible experiences that cater to a diverse range of visitor needs [2]. Traditional guided tours often fall short of providing individualized content that caters to specific interests, linguistic preferences, and learning styles, which is where innovative technological solutions can play a vital role [3], [4].

The growing demand for enhanced cultural tourism ex- periences has led to the exploration of integrating artificial intelligence (AI), machine learning (ML), and voice-based interfaces with cultural heritage sites. These technologies have the potential to transform how visitors interact with heritage, offering dynamic and personalized content tailored to the spe- cific context of the visitor's location, interests, and language [5]. In response to this need, the proposed project introduces an AI-powered voice assistant mobile application designed to deliver real-time, multilingual, and personalized insights at heritage sites through machine learning algorithms, natural language processing (NLP), and location-aware services.

The app's intelligence is powered by a fine-tuned Mistral 7B large language model (LLM), which interprets user queries, extracts intent, and generates detailed and relevant responses in a conversational manner. Personalization is enabled us- ing a lightweight Gated Recurrent Unit (GRU)-based model, trained on sentence embeddings derived from the paraphrase-

MiniLM-L6-v2 transformer. This allows the assistant to analyze prior interactions and user preferences over time to offer more accurate and relevant recommendations. The embeddings and chat history are stored in an SQLite database, which supports personalized query matching based on thematic interests (e.g., temples, forts, monuments).

The application also employs location-aware technology, utilizing the Haversine Distance Algorithm [6] to calculate geographic proximity and recommend top-rated nearby heritage sites based on the user's current position. This enables users to receive contextual information dynamically as they explore different regions or landmarks, enhancing their self-guided tour experience [7].

A core feature of the system is its voice interaction capability, which supports both speech-to-text and text-to-speech. OpenAI's Whisper model [8] is used for accurate and multilingual speech recognition, enabling users to interact with the assistant using natural voice commands. For spoken responses, the system uses gTTS for generating natural and intelligible speech output, thus providing a fully hands-free experience that increases accessibility and engagement.

With the ongoing advancement of AI and speech technologies, the need for adaptable and intelligent solutions in cultural tourism is growing [9]. The proposed mobile application is designed to meet these demands by offering an efficient, personalized, and scalable voice assistant that can be tailored for different cultural regions, languages, and user profiles. Whether deployed at small heritage sites or major cultural landmarks, the system enhances user experience by delivering an immersive, context-aware, and educational journey.

In summary, this project bridges the gap between cultural heritage and intelligent technologies by combining large language models, GRU-based personalization, geolocation services, and speech interfaces into one unified platform. It aims to redefine cultural exploration through a voice-based AI assistant that is multilingual, adaptive, and user-centric, making the discovery of history and heritage more engaging, accessible, and meaningful for every visitor.

A significant innovation of the assistant lies in its multilingual support, which ensures inclusivity for users across diverse linguistic backgrounds. The system performs real- time translation between over 20 Indian languages, enabling users to communicate in their preferred language while still benefiting from the full capabilities of the assistant. This feature not only broadens accessibility for domestic and in- ternational tourists but also supports dialectal variations and transliterated input, such as Tamil typed in English script. By integrating language detection and translation within the NLP pipeline, the app ensures that user queries are accurately interpreted and responses are delivered naturally in the target language. This real-time multilingual functionality reinforces the assistant's mission to make Tamil Nadu's rich cultural heritage comprehensible and engaging to all, regardless of linguistic barriers.

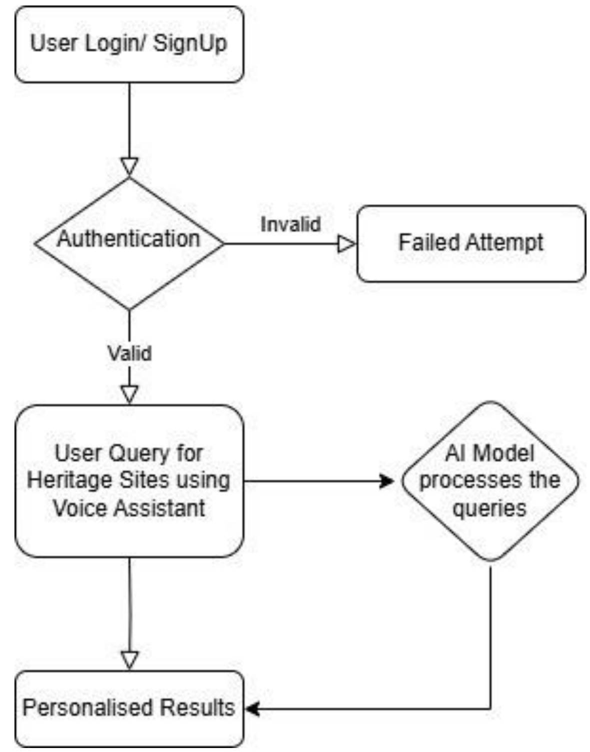

Fig. 1. Flow diagram of the proposed system

II. Literature Survey

The integration of AI and machine learning in cultural tourism has gained significant attention in recent years, with a focus on improving user experiences through personalized content, intelligent recommendations, and interactive interfaces. In the context of AI-powered chatbots, Ahmed et al. [10] explore the use of few-shot learning in transformer- based conversational AI systems, emphasizing the importance of leveraging transformers for developing efficient chatbot systems. This work contributes to the growing understanding of how conversational AI can provide personalized interactions in various domains, including tourism.

Casillo et al. [5] introduced a cultural heritage-aware chatbot designed to enhance tourist experiences by providing contextual information about cultural heritage sites. The CHAT-Bot system not only delivers relevant historical data but also adapts to the visitor's preferences, thus promoting engagement and education during tours. Similarly, Casillo et al. [11] conducted a survey on context-aware recommender systems for cultural heritage, outlining the potential of AI-driven systems to provide tailored experiences based on real-time context and user behavior. This research highlighted the challenges and opportunities in integrating contextual awareness into cultural tourism platforms.

Cossatin et al. [12] presented a framework for incorporating large language models (LLMs) into cultural heritage websites, allowing users to explore detailed information through advanced AI-driven content. The integration of LLMs improves the quality of the information provided and facilitates advanced exploration by offering deeper insights tailored to the visitor's queries. This approach aligns with the objective of enhancing the tourism experience by providing personalized, adaptive, and contextually relevant information at heritage sites.

Fararni et al. [1] proposed a hybrid recommender system for tourism that combines big data and AI to offer personalized suggestions based on user preferences, behavior, and external factors such as location and time. This framework is relevant to cultural tourism as it demonstrates the potential of combining multiple data sources for improved recommendation accuracy. The system's ability to generate dynamic recommendations enhances the cultural tourism experience by adapting to the user's needs in real-time.

Park et al. [9] explored the performance improvement of prompt engineering for generative AI, particularly focusing on large language models. Their findings highlight the importance of prompt optimization in generating high-quality, context-specific responses from AI systems. This research is particularly relevant to the development of AI-powered voice assistants for cultural tourism, where prompt engineering can ensure that the system delivers relevant and accurate information in real-time.

Sperli [3] introduced a deep learning-based chatbot designed specifically for supporting the tourist journey through cultural heritage sites. The framework utilizes AI to guide visitors and provide them with personalized insights, improving the overall tour experience. However, while the chatbot offers dynamic interactions, the integration of advanced location-aware algorithms was not addressed, which could further enhance the personalized experience based on real-time visitor positions.

Tsepapadakis and Gavalas [2] developed an audio augmented reality (AR) conversational guide for cultural heritage sites, combining AI-driven conversations with location-based AR features. This innovative approach allows visitors to interact with virtual guides while exploring heritage sites, significantly improving the experience by providing contextually relevant and immersive content.

Yin et al. [13] introduced a conversational recommendation system that uses multi-granularity latent variable enhancement and user entity focus to deliver personalized recommendations. Their work contributes to the development of more sophisticated recommendation models, which can be adapted to cultural tourism systems for offering dynamic content based on user interests and behavior.

Qu et al. [14] developed CmnRec, a sequential recommendation system that leverages chunk-accelerated memory networks for improving recommendation accuracy and user satisfaction. This approach, based on deep learning, could be utilized in cultural tourism applications to provide more refined and personalized recommendations to visitors as they interact with heritage sites.

Gao et al. [6] proposed an adversarial human trajectory learning system for trip recommendations, highlighting the potential of using human trajectory data to predict tourist behavior and optimize recommendations. By incorporating this approach into cultural tourism systems, recommendations can be dynamically adjusted based on user movements, preferences, and historical behavior.

Chen and Huang [15] explored algorithmic approaches to content recommendation in new media, emphasizing the use of AI to enhance user engagement through effective content delivery. Their research supports the idea that cultural tourism systems can benefit from AI-powered recommendation engines that adapt to user interests, delivering tailored experiences in real-time.

Malodia et al. [4] investigated why people use AI-enabled voice assistants, focusing on the factors that drive adoption. Their study provides valuable insights into user motivations, which can help shape the design of AI-powered voice assistants for cultural tourism by ensuring that the system meets user expectations and promotes engagement.

Lam et al. [8] examined the application of transformer- based models in educational virtual assistants, demonstrating their effectiveness in providing personalized and contextually relevant information. This research underscores the potential of using transformers in AI-driven systems for cultural tourism, where tailored content delivery is essential for enhancing the visitor experience.

Finally, Fontanella et al. [7] discussed the application of pattern recognition and AI techniques for cultural heritage, providing a comprehensive overview of the ways in which these technologies can be leveraged to preserve and enhance cultural heritage sites. Their work contributes to the broader field of AI in cultural tourism by highlighting the importance of integrating AI for both content generation and user interaction.

The research outlined in these works highlights the grow- ing role of AI, machine learning, and voice assistants in transforming the cultural tourism industry. However, many of these systems lack full integration of multilingual capabilities, real-time location-based features, and seamless personalized interactions, which are crucial for creating truly immersive and adaptive tourism experiences. The proposed system aims to address these gaps by incorporating advanced AI models, such as mT5 for multilingual content generation, GRU for sequential behavior prediction, and location-aware algorithms for context-based recommendations.

Motivated by these existing works, this paper aims to design and develop an AI-powered voice assistant mobile application for cultural tourism, which integrates multilin- gual, personalized, and location-aware features. The following sections of the paper are organized as follows: Section III discusses the proposed system model; Section IV outlines the design requirements; Section V details the implementation and results; Section VI presents the conclusion; and Section VII highlights the future scope of the work.

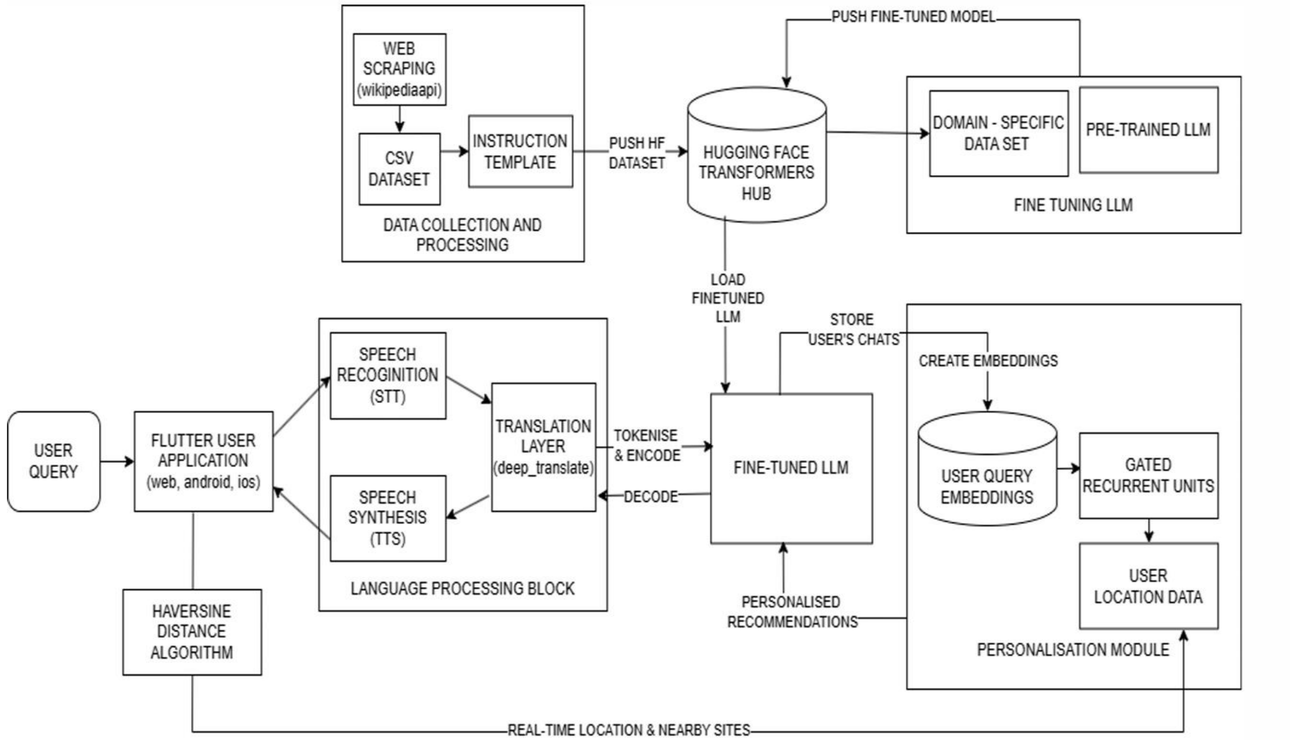

Fig. 2. Block diagram of our proposed system

III. System Model

The proposed system focuses on a cultural tourism assis- tant powered by AI-driven voice interaction and personalized recommendations. The system leverages advanced models such as the fine-tuned Mistral 7B large language model for content generation and a GRU-based model for user preference learning. Additionally, it incorporates the Haversine Distance Algorithm for proximity-based site suggestions. The assistant is designed to deliver real-time insights to tourists at heritage locations by understanding both user history and geolocation, thereby enriching the overall cultural tourism experience. The proposed system block diagram is shown in Fig. 2.

The main components of the system are:

a) Multilingual Transformer Model (Mistral 7B): The Mistral 7B language model dynamically generates informative and context-aware responses. It supports over 20 Indian languages, adapting responses based on user queries and language preferences to ensure accessibility across diverse audiences.

b) GRU-Based User Preference Modeling: The assistant applies a lightweight GRU-based neural model to capture sequential user behavior using sentence embeddings. This model enables continuous learning from prior interactions, thereby allowing for dynamic, preference-aware heritage site recommendations and adaptive conversation flow.

c) GPS and Haversine-Based Retrieval: Using real-time GPS coordinates, the system computes the nearest heritage sites through the Haversine formula. This enables accurate, radius-based retrieval of location-relevant information, enhancing the contextual delivery of cultural insights as the user explores physical spaces.

d) Voice Input/Output Interface: The assistant integrates speech-to-text and text-to-speech modules to support natural language voice interaction. This hands-free interface allows users to query heritage-related information and receive spoken, language-specific responses, enriching the immersive experience.

e) Context-Aware Dialogue Management: To maintain coherent and engaging conversations, the system incorporates a context-aware dialogue manager that retains session history and interprets follow-up questions intelligently. This module leverages attention mechanisms to maintain continuity across user interactions, enabling the assistant to provide more relevant and context-sensitive responses throughout the user's cultural journey.

The operating principle of the proposed system is shown in Fig. 1. It starts with voice input processed by the Whisper model, followed by query understanding via Mistral 7B, and retrieval of relevant heritage data based on location and preferences. The GRU module continuously learns from user input to refine future suggestions. The system outputs responses via TTS in the user's language, offering a real-time, personalized cultural tour experience.

IV. Design Requirements

The experimental setup of the proposed system is illustrated in Fig. 3. The system architecture comprises the following hardware and software components:

A. Hardware Components

- **Smartphone / Tablet** – The mobile application is developed using Flutter and runs on standard smartphones or tablets. The device's built-in microphone and speaker are used for voice input and output, eliminating the need for external peripherals.
- **GPS Module** – The device's integrated GPS enables real-time location tracking, allowing the system to offer context-aware heritage site recommendations based on the user's position.
- **Internet Connectivity (Wi-Fi / Mobile Data)** – A stable internet connection is required to access cloud-based AI models, retrieve multilingual content, and update dynamic recommendations.

B. Software Components

- **Flutter Framework** – The application is built using Flutter, providing cross-platform support for both Android and iOS devices.
- **Whisper Model** – An advanced speech recognition model that transcribes user voice input into text, enabling real-time interaction.
- **Multilingual AI Model (Mistral 7B)** – A large lan- guage model responsible for generating multilingual text responses, ensuring accessibility for users from diverse linguistic backgrounds.
- **Text-to-Speech (TTS) Engine** – A TTS system, such as Google TTS, converts textual responses into natural-sounding speech, facilitating hands-free interaction.
- **GPS-based Context Engine** – This module processes geolocation data to provide dynamic, location-aware insights tailored to the user's surroundings.
- **Cloud Backend (Firebase)** – A cloud-based backend infrastructure is used for storing user preferences, managing updates, and maintaining system scalability.
- **Personalization Module** – A lightweight recurrent model tracks user interactions and refines recommendations over time, enhancing the personalization of responses.

The integration of these hardware and software components forms the foundation of a robust and scalable AI-powered heritage assistant. By combining mobile accessibility, geolocation awareness, real-time speech processing, and personalized content generation, the system is well-equipped to deliver an interactive and immersive cultural tourism experience. The modular design ensures ease of maintenance and future extensibility, allowing for the seamless incorporation of additional features and enhancements as user needs evolve.

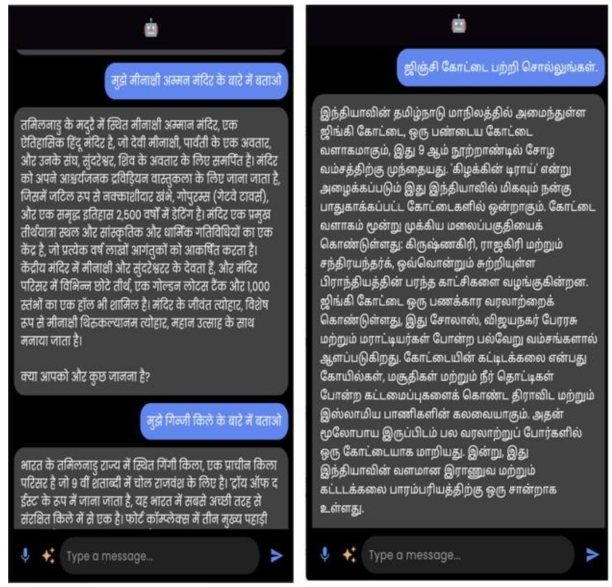

Fig. 3. Experimental setup of the proposed system

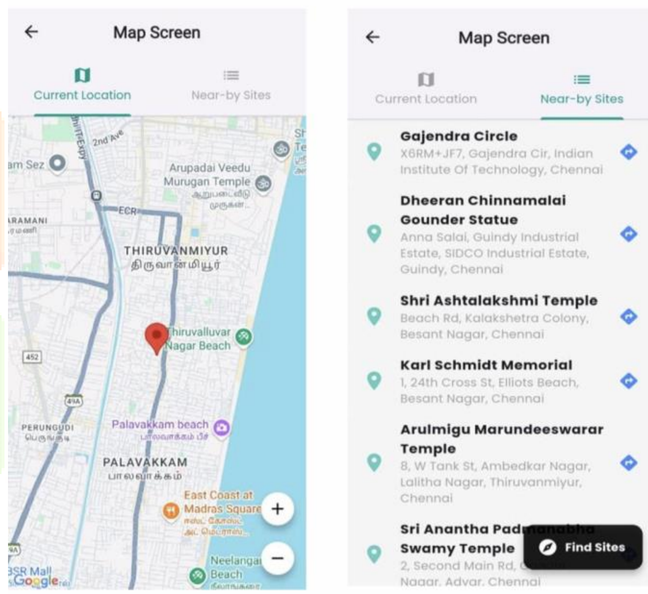

Fig. 4. Functional prototype of the cultural tourism assistant

V. Implementation and Results

The proposed AI-based cultural tourism assistant system is deployed in a real-world heritage site environment. The system continuously processes real-time GPS data and user voice input to provide relevant multilingual insights. The edge computing device successfully integrates speech recognition, AI-based content generation, and text-to-speech synthesis to facilitate seamless interaction with tourists.

A prototype implementation of the proposed system is shown in Fig. 4. During model fine-tuning, key training metrics such as loss, gradient norm, and learning rate were continuously monitored. The training loss decreased steadily from an initial value of 2.5 to around 0.5, indicating strong convergence. Gradient norms stabilized around 0.5 after an initial spike, suggesting healthy learning dynamics without vanishing or exploding gradients. The learning rate followed a decaying schedule, aiding in stable convergence during later epochs. These trends confirm the effective fine-tuning of the Mistral 7B model, optimizing it for generating multilingual, context-aware responses in the proposed cultural tourism assistant system.

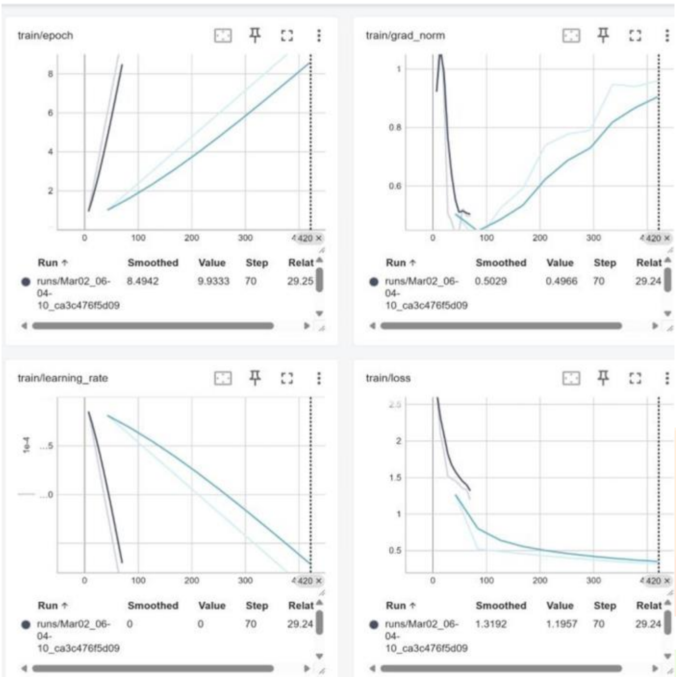

Fig. 5. Fine-tuned model metrics

In addition to quantitative training metrics, qualitative evaluation was conducted through real-time user interaction trials at selected heritage locations. Users reported high satisfaction with the accuracy, fluency, and relevance of the AI-generated responses. The voice interface was particularly appreciated for its ease of use, especially in noisy outdoor environments where hands-free operation is advantageous. Multilingual support was validated across several Indian languages, demonstrating the system's inclusivity and adaptability.

The personalized recommendations generated by the GRU-based module showed improved contextual alignment over repeated user sessions, validating the system's ability to adapt based on historical user interactions.

Overall, the implementation results demonstrate both the technical feasibility and practical effectiveness of the proposed system, positioning it as a promising tool for enhancing smart tourism experiences through intelligent, voice-driven, and culturally aware interactions.

VI. Conclusion

In the era of smart tourism, the proposed system introduces an AI-driven, multilingual voice assistant for heritage exploration. By integrating NLP models, speech recognition, and GPS-based recommendations, the system delivers real-time, location-aware insights, enhancing the visitor experience. The hands-free interaction reduces reliance on traditional guides and printed materials, making cultural tourism more accessible and immersive. With its ability to personalize responses based on user interactions, the system effectively bridges technology with historical storytelling, demonstrating practical feasibility for real-world deployment.

The system's adaptability allows it to cater to different user preferences, making it an inclusive tool for tourists from diverse linguistic and cultural backgrounds. Its lightweight implementation using Flutter ensures cross-platform compatibility, making it easy to deploy on mobile devices. Furthermore, by leveraging AI-driven responses, the system provides an engaging and dynamic experience, keeping users informed and entertained throughout their exploration.

By combining AI and cultural heritage, the system not only improves visitor engagement but also contributes to the digital preservation of historical knowledge. The interactive nature of the assistant makes learning about heritage sites more engaging, fostering a deeper appreciation for history and culture.

Moreover, the modular design of the system enables easy integration with future enhancements, such as augmented reality features, sentiment-aware dialogue, or extended datasets for broader regional coverage. As cultural tourism evolves, such AI-driven tools have the potential to become indispensable companions for modern travelers, empowering users to discover, learn, and connect with heritage in a meaningful and technologically enriched manner.

VII. Future Scope

Future enhancements can extend coverage to more heritage sites with improved personalization features. The integration of augmented reality (AR) can provide interactive historical reconstructions, enriching the user experience. Implementing solar-powered modules can improve sustainability, enabling operation in remote locations. AI-driven storytelling and real-time multilingual translation can further enhance engagement by adapting content dynamically. Additionally, blockchain technology can be explored for secure data storage and user authentication, ensuring privacy while maintaining a personalized experience.

Further research can focus on optimizing speech recognition for noisy outdoor environments, ensuring seamless interactions in crowded heritage sites. The inclusion of machine learning-based sentiment analysis could enable the system to adapt responses based on user emotions, providing a more intuitive and engaging experience.

REFERENCES

[1] K. A. Fararni, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi, and A. Sabri, "Hybrid recommender system for tourism based on big data and AI: A conceptual framework," *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 47–55, 3 2021, [Online; accessed 2025-02-02].

[2] M. Tsepapadakis and D. Gavalas, "Are you talking to me? An Audio Augmented Reality conversational guide for cultural heritage," *Pervasive and Mobile Computing*, vol. 92, no. 1, p. 101797, 5 2023, [Online; accessed 2025-02-02].

[3] G. Sperl´ı, "A cultural heritage framework using a Deep Learning based Chatbot for supporting tourist journey," *Expert Systems with Applications*, vol. 183, no. 1, p. 115277, 11 2021, [Online; accessed 2025-02-02].

[4] S. Malodia, N. Islam, P. Kaur, and A. Dhir, "Why Do People Use Artificial Intelligence (AI)-Enabled Voice Assistants?" *IEEE Transactions on Engineering Management*, vol. 71, no. 1, pp. 491–505, 2024, [Online; accessed 2025-02-09].

[5] M. Casillo, F. Clarizia, G. D'Aniello, M. De Santo, M. Lombardi, and D. Santaniello, "Chat-Bot: A cultural heritage aware teller-bot for supporting touristic experiences," *Pattern Recognition Letters*, vol. 131, no. 1, pp. 234–243, 3 2020, [Online; accessed 2025-02-02].

[6] Q. Gao, F. Zhou, K. Zhang, F. Zhang, and G. Trajcevski, "Adversarial Human Trajectory Learning for Trip Recommendation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 1764–1776, 4 2023, [Online; accessed 2025-02-09].

[7] F. Fontanella, F. Colace, M. Molinara, A. Scotto Di Freca, and F. Stanco, "Pattern recognition and artificial intelligence techniques for cultural heritage," *Pattern Recognition Letters*, vol. 138, no. 1, pp. 23–29, 10 2020, [Online; accessed 2025-02-09].

[8] K. N. Lam, L. H. Nguy, V. L. Le, and J. Kalita, "A Transformer-Based Educational Virtual Assistant Using Diacriticized Latin Script," *IEEE Access*, vol. 11, no. 1, pp. 90 094–90 104, 2023, [Online; accessed 2025-02-09].

[9] D. Park, G.-t. An, C. Kamyod, and C. G. Kim, "A Study on Performance Improvement of Prompt Engineering for Generative AI with a Large Language Model," *Journal of Web Engineering*, vol. 1, no. 1, pp. 1187–1206, feb 22 2024, [Online; accessed 2025-02-02].

[10] M. Ahmed, H. U. Khan, and E. U. Munir, "Conversational AI: An Explication of Few-Shot Learning Problem in Transformers-Based Chatbot Systems," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 1888–1906, 4 2024, [Online; accessed 2025-02-02].

[11] M. Casillo, F. Colace, D. Conte, M. Lombardi, D. Santaniello, and C. Valentino, "Context-aware recommender systems and cultural heritage: a survey," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 3109–3127, aug 19 2021, [Online; accessed 2025-02-02].

[12] A. G. Cossatin, N. Mauro, F. Ferrero, and L. Ardissono, "Tell me more: integrating LLMs in a cultural heritage website for advanced information exploration support," *Information Technology Tourism*, vol. 1, no. 1, pp. 1–32, jan 28 2025, [Online; accessed 2025-02-02].

[13] Y. Yin, Y. Pan, X. Bao, and F. Huang, "Conversational Recommendations With User Entity Focus and Multi-Granularity Latent Variable Enhancement," *IEEE Transactions on Knowledge and Data Engineering*, vol. 37, no. 3, pp. 1126–1139, 3 2025, [Online; accessed 2025-02-09].

[14] S. Qu, F. Yuan, G. Guo, L. Zhang, and W. Wei, "Cmnrec: Sequential Recommendations With Chunk-Accelerated Memory Network," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3540–3550, apr 1 2023, [Online; accessed 2025-02-09].

[15] Y. Chen and J. Huang, "Effective Content Recommendation in New Media: Leveraging Algorithmic Approaches," *IEEE Access*, vol. 12, no. 1, pp. 90 561–90 570, 2024, [Online; accessed 2025-02-09].