Applying Large Language Models to Build the Tourmate Smart Travel Support Platform

Tiep Quang Tran[1], Chau Ngo Minh[1], Minh-Anh Vo Ngoc[2], Ngoc Luu Thi Minh[1], Zhang Yuemei[1], Hung Ha Manh[1*]

[1]International School, Vietnam National University, Hanoi, Vietnam

[2]University of Science, Vietnam National University, Ho Chi Minh, Vietnam

[*]Corresponding author: hunghm@vnu.edu.vn

Abstract

The integration of Large Language Models (LLMs) into the tourism sector is reshaping how travelers interact with digital services. This research introduces TourMate, a smart travel support platform that harnesses LLM capabilities to enhance the travel experience for both domestic and international tourists in Vietnam. The system features an AI-driven chatbot, personalized itinerary recommendations, and real-time travel insights, enabling seamless and adaptive assistance. Using natural language processing and machine learning, TourMate can understand user preferences, suggest optimized routes, and provide reliable local service recommendations. This study examines the implementation of LLMs within TourMate, assesses their impact on user engagement, and explores challenges such as data reliability, multilingual functionality, and responsiveness. The findings offer valuable insights into the development of AI-driven tourismapplications, contributing to the advancement of intelligent travel solutions.

Research purpose:

The purpose of this research is to investigate the application of Large Language Models (LLMs) in the tourism sector through the development of TourMate, an intelligent travel support platform. Specifically, the study aims to assess how LLM-driven features—such as personalized itinerary recommendations, adaptive chatbot support, and real-time travel insights—can enhance user engagement and improve the travel experience of domestic and international tourists in Vietnam.

Research motivation:

Tourism in Vietnam is experiencing rapid growth, with increasing demands for personalized, seamless, and digital-first travel services. However, existing solutions often lack adaptability, multilingual support, and contextual awareness. Recent advances in LLMs offer a promising opportunity to overcome these limitations by enabling intelligent, human-like interactions. This research is motivated by the need to bridge the gap between conventional travel platforms and the growing expectations of tech-savvy travelers, while also contributing to the digital transformation of the tourism industry.

Research design, approach, and method:

This study adopts a ***design science research approach****, combining system design, prototyping, and user evaluation. The research process includes: (1)* ***System Development*** *– Designing and implementing the TourMate platform with key modules such as an LLM-based chatbot, itinerary optimizer, and real-time insight generator. (2)* ***Experimental Evaluation*** *– Conducting usability testing and user studies with both domestic and international tourists in Vietnam to assess engagement, satisfaction, and reliability.*

Main findings:

The results indicate that the integration of LLMs significantly enhances user engagement by enabling more natural and context-aware interactions. TourMate was found effective in delivering personalized itineraries, providing accurate local recommendations, and supporting real-time decision-making. Nevertheless, challenges remain in terms of ensuring data reliability, maintaining fast response times, and addressing multilingual complexities.

Practical/managerial implications:

This research offers several implications for tourism stakeholders:

- For service providers: LLM-based platforms can improve customer experience, increase loyalty, and reduce reliance on human support staff.
- For destination managers: Intelligent insights can help optimize visitor flow, reduce congestion, and improve satisfaction.
- For technology developers: The findings highlight the importance of balancing personalization with performance, and of designing scalable, multilingual AI systems.

N. D. Nguyen and P. T. K. Ngoc (eds.), *Proceedings of the International Conference on Emerging Challenges: Business Dynamics in Disruptive Economy (ICECH 2025)*, Advances in Economics, Business and Management Research 377,
https://doi.org/10.2991/978-94-6239-622-7_26

Overall, the study demonstrates that adopting LLM-driven solutions can accelerate the digital transformation of tourism, positioning Vietnam as a leader in smart tourism innovation.

Keywords: Large Language Models (LLMs), Artificial Intelligence (AI), Smart

1. INTRODUCTION

The evolution of travel technologies has increasingly shifted traditional tourism support toward AI-driven, personalized assistance, enhancing traveler experiences in real time. Recent advancements in natural language processing, particularly the development of Retrieval-Augmented Generation (RAG) systems, have demonstrated significant potential in transforming the way information is accessed and delivered. RAG systems integrate large language models (LLMs)—especially those based on transformer architectures such as GPT—with external knowledge retrieval mechanisms. This hybrid architecture enhances the capabilities of LLMs by grounding their responses in up-to-date and contextually relevant information, which is particularly valuable in dynamic and information-sensitive domains such as tourism (Brown et al., 2020).

By retrieving pertinent data from curated knowledge bases or the web before generating outputs, RAG systems enable real-time, accurate question answering, itinerary planning, and culturally sensitive guidance (Wang et al., 2024). This approach addresses several key challenges in AI-driven travel support, including maintaining factual accuracy and contextual relevance in localized environments—where understanding cultural nuances and language subtleties is critical (TravelRAG, 2024). Furthermore, RAG frameworks offer enhanced control over information sources, contributing to ethical AI practices by mitigating issues related to data bias and privacy (Xiang, Du, Ma, & Fan, 2017). Nevertheless, the integration of RAG systems with real-time, location-based services introduces technical complexities, such as computational overhead and latency management (Gretzel, Sigala, Xiang, & Koo, 2015).

To tackle these challenges, this paper proposes TourMate, a smart travel support platform powered by RAG technology. TourMate is designed to deliver real-time, personalized, and culturally contextualized travel assistance by combining the natural language understanding strengths of LLMs with accurate, situationally relevant information retrieval. By leveraging RAG in this manner, TourMate aims to provide a seamless, intelligent travel experience that supports both domestic and international travelers, and sets a new standard for AI-enhanced tourism solutions.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

2.1 ARTIFICIAL INTELLIGENCE APPLICATIONS IN TOURISM

In recent years, Artificial Intelligence (AI) has become a trans-formative force in the tourism industry, enabling more personalized, efficient, and real-time travel experiences. AI applications span multiple domains, including demand forecasting, price optimization, customer service automation, and personalized recommendation engines. According to Xiang et al. (2021), AI-powered systems are increasingly used by travel agencies, hospitality services, and smart cities to optimize customer engagement and streamline operations. In particular, natural language processing (NLP) technologies have been integrated into customer service systems to offer multilingual support and contextual query handling, significantly improving the user experience. Image recognition, sentiment analysis, and predictive analytic have also improved traveler profile, behavior prediction, and service customization. However, many traditional AI tools still lack the deep contextual understanding and flexibility offered by the more recent advances in large-scale language modeling.

2.2 Overview of Large Language Models (LLMs) and their Applications

Large language models (LLMs) are an important branch of artificial intelligence in the field of Natural Language Processing (NLP). LLMs are essentially deep learning models trained on extremely large linguistic data sets to predict the next word in a text string, thereby learning the syntactic structure, semantics, context, and even the hidden knowledge in the language.

Modern LLMs are mainly built on the Transformer architecture, introduced by Vaswani et al. (2017). Transformers use self-attention to model the relationships between words in a sentence, allowing them to handle context at a distance more efficiently than previous models such as RNNs or LSTMs. Due to their ability to process in parallel and learn complex relationships in text, Transformers have become the core architecture in models such as BERT , GPT-2/3/4, and PaLM.

The versatility of LLMs lies in their zero-shot and few-shot learning abilities, allowing them to generalize across domains with minimal task-specific training. Recent studies by Bommasani et al. (2021) describe LLMs as “foundation models” with wide applicability across education, healthcare, legal reasoning, and more. In tourism, their use remains relatively limited but is rapidly gaining attention for powering intelligent chatbots and automated itinerary planners.

2.3 Travel Chatbots and Intelligent Recommendation Systems

Chatbots are understood as software systems that simulate conversations between humans and machines through text or voice chat interfaces. Chatbots can be programmed according to available scenarios (rule-based) or use artificial intelligence to learn and respond flexibly according to context (AI-based) . Modern virtual assistants often apply deep learning models such as RNNs, Transformers, or large language models (LLMs), allowing for increased understanding and generation of natural language, making the interaction process more natural and "human". In the travel industry, chatbots and virtual assistants are deployed at various touchpoints of the customer journey, from information search, booking services, to in-trip support and after-sales service. These systems help process requests 24/7, reduce personnel costs, and increase traveler satisfaction and loyalty \cite{b10}. For example, companies such as Booking.com, Expedia, and Traveloka have integrated chatbots to assist users in finding rooms, answering frequently asked questions, or handling booking changes. Some major airlines also apply chatbots to support automatic check-in or update flight schedules.

Despite the many benefits, the development of chatbots and virtual assistants in tourism still faces challenges such as: imperfect natural language understanding, difficulty in handling complex or emotional requests, as well as language and cultural barriers. Therefore, current research is aimed at integrating large language models (LLMs), knowledge retrieval, and reinforcement learning to improve the accuracy, adaptability, and self-learning ability of intelligent travel assistant systems

2.4 TourMate's Research Gap and Contribution

Despite advancements in AI-powered travel tools, current solutions often lack deep conversational fluency, multi-modal integration, and adaptability across cultural and linguistic contexts. Many systems still rely on template-based dialogue, keyword extraction, and static decision trees, which limit their responsiveness to nuanced or unpredictable queries.

TourMate aims to address these shortcomings by employing a state-of-the-art LLM as the core engine of a smart travel assistant platform. Its contributions include:

- A multilingual, context-aware dialogue system that can engage in natural, open-ended conversations.
- An intelligent recommendation module that responds to real-time user behavior, preferences, and environmental factors such as location and weather.
- Seamless integration across platforms (web, app, and messaging services) to ensure user convenience.
- A feedback-driven learning loop that allows the system to improve continuously through real-world usage.

By integrating LLMs with tourism-specific knowledge and real-time data streams, TourMate fills a critical gap in the design of adaptive and intelligent travel support systems, especially for culturally diverse and digitally connected travelers.

3. METHODOLOGY

This study follows a design science approach to develop and evaluate TourMate, a smart travel assistant powered by Large Language Models (LLMs). The system integrates a prompt-engineered LLM (GPT-4) with contextual data sources including location, weather, and user preferences, combined through a modular architecture using Retrieval-Augmented Generation (RAG). Tourism-specific datasets, such as MultiWOZ, TripAdvisor UGC, and Vietnamese POI metadata, were collected, cleaned, and used to construct a semantic vector database. To evaluate system performance, we used quantitative metrics focused on real-world travel query handling: Precision, Context Precision, Recall, Relevance, Entities Recall, and average response time. All evaluations were conducted using live user inputs over web and Messenger interfaces.

3.1 Prompt Engineering

To enhance reproducibility and ensure consistent chatbot behavior, we explicitly describe the prompt engineering strategy adopted in TourMate. The LLM (GPT-4) was instructed to extract structured entities from user queries. The main prompt used is as follows: "You are a travel assistant. Extract the following information from the user query: {location}, {category}, {mood}, {price range}. If an entity is missing, return 'N/A'."

Here, **{location}** refers to the geographical area (e.g., District 1, Ho Chi Minh City), **{category}** denotes the type of venue (e.g., café, museum, park), **{mood}** indicates the experiential style desired by the user (e.g., romantic, relaxing, adventurous, traditional), and **{price range}** corresponds to the budget level (low, medium, high). This explicit design ensures consistency across queries and clearly defines the scope of the chatbot's recommendations.

3.2 Dataset

To build a robust and intelligent travel chatbot tailored for users exploring Ho Chi Minh City, a comprehensive dataset of 1,021 tourist destinations was curated. Each record captures a wide range of structured and unstructured information, including:

- Name
- Location
- Opening and Closing Hours
- Detailed Time Schedules
- Type (7 distinct categories)
- Mood (5 categories capturing the ambiance of the location)
- Price
- Image URL
- Descriptive Information

The data was manually collected from reputable and verified sources such as Vietnam's official tourism portals, travel guide platforms (e.g., VinWonders, Traveloka, TripAdvisor), and local tourism publications. Manual curation ensured accuracy and relevance, especially in categorizing types and moods specific to user intent.

4. SYSTEM DESIGN AND ARCHITECTURE

Here we propose the architecture of the chatbot system, which utilizes Retrieval-Augmented Generation (RAG) technology to enable intelligent querying, personalized information delivery, and an enhanced user experience. The system consists of several core components, as illustrated in the architectural diagram in **Figure 1:**

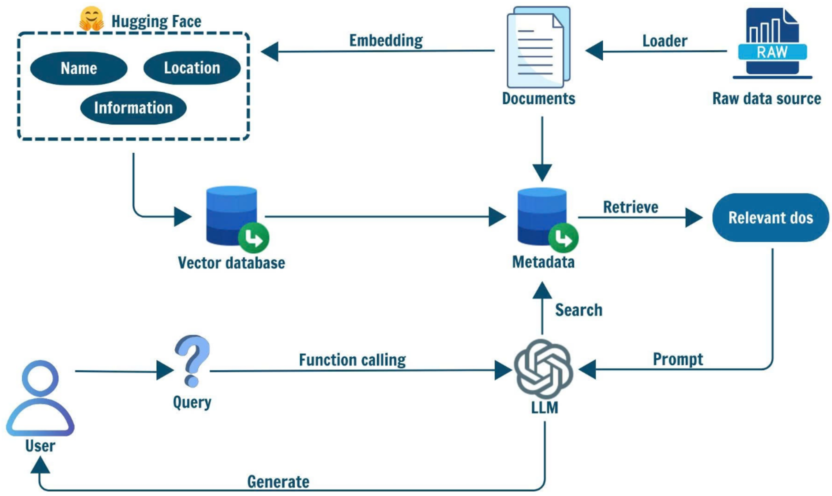

FIG 1. THE ARCHITECTURE OF THE CHATBOT SYSTEM

(Source: Own elaboration)

In this paper, to build a robust and intelligent travel TourMate chatbot for users, we choose HoChiMinh City for the test point. The data was manually collected from reputable and verified sources such as Vietnam's official tourism portals, travel guide platforms (e.g., VinWonders, Traveloka, TripAdvisor), and local tourism publications. Manual curation ensured accuracy and relevance, especially in categorizing types and moods specific to user intent. Then we need preprocessing data with two main parts: Time presentation and Price structure.

Upon receiving a user query through the chatbot interface (e.g.,``Địa điểm cà phê chụp hình tại quận 1"), the system initiates intent recognition using **Function Calling** within a **Large Language Model (LLM)**.

The model parses the query to extract the user's intent and relevant entities such as{location}, {category},{mood}, or {price range}. Based on this extracted information, the system first queries {MongoDB}, which stores structured data including place names, pricing, opening hours, and contextual tags.

MongoDB performs a filtering process to retrieve entries that match these basic attributes and returns a list of associated document IDs for further semantic ranking.

These filtered results are then passed to Milvus, a high-performance vector database, where fields such as descriptions, reviews, and semantic tags are pre-embedded and indexed. Milvus performs similarity-based vector searches using techniques like IVF (Inverted File Index), retrieving semantically relevant entries. The outputs from both MongoDB and Milvus are merged and ranked using a combined scoring mechanism based on semantic similarity and metadata relevance. The final ranked list is returned to the LLM, which generates a fluent, user-friendly natural language response and delivers it back through the chat interface.

For deployment, the system leverages Docker Compose to orchestrate services including **MongoDB, Milvus, Etcd, and MinIO.**

MongoDB is exposed on port 27017 and stores structured data in a JSON-like schema, including fields such as name, location, mood, type of location, and price. Each document may include a milvus id to map with its vectorized counterpart in Milvus. Milvus operates in standalone mode and connects to Etcd and MinIO for metadata and object storage respectively, exposing port 19530 for vector queries.

All containers communicate over a custom Docker network (services milvus-network), with credentials and ports managed securely via environment variables. This architecture allows TourMate to combine symbolic and semantic reasoning for context-aware query handling, ensuring efficient retrieval, multilingual response generation, and a seamless user experience.

While MongoDB is well-suited for storing structured or semi-structured data such as prices, open hours, and categorical tags, it is not optimized for semantic search over unstructured text fields like detailed descriptions or user reviews.

To overcome this limitation, TourMate integrates Milvus—a vector database purpose-built for high-performance similarity search. Milvus enables fast and scalable semantic retrieval using approximate nearest neighbor (ANN) algorithms such as IVF and HNSW, allowing the system to handle nuanced queries like:

“Phong cách cổ kính mang bản sắc dân tộc”

Milvus stores vectorized embeddings of text fields (e.g., name, location, information), each indexed by a shared milvus id that links back to structured entries in MongoDB. This dual-database design allows TourMate to combine symbolic and semantic reasoning in response generation. The system is scalable and robust, supporting real- time search over millions of vectors—which is essential for dynamic travel and lifestyle recommendations. When using both MongoDB and Milvus, the system can be optimized in terms of storage and performance.

4.1 ***Information Retrieval from MongoDB***

User queries are first processed through Function Calling, which extracts relevant fields according to a predefined schema. The system then uses these fields to filter products from MongoDB based on structured attributes.

The fields extracted and used for querying MongoDB include:

- “price”:
- If the value is in the format <XXXXXX, MongoDB searches for products with prices less than or equal to the specified value.
- If the value is in the format >XXXXXX, MongoDB searches for products with prices greater than or equal to the specified value.
- If the value is in the format XXXXXX - XXXXXX, MongoDB filters products within the defined price range.
- If the query includes expressions such as “cheapest possible” or “as expensive as possible”, the system sets the price field to "min" or "max" and retrieves the Top-K cheapest or most expensive products accordingly.
- “mood”: The query filters results that match the mood attribute.
- “type”: Filter results are based on the type of destination or product.
- “time”: Filter based on available time windows or operational hours.

After this initial filtering process in MongoDB, the system extracts the milvus_id values of all matching entries and uses them to query Milvus for semantic ranking and optimization.

In cases where the user requests product bundles or combinations (e.g., combos), the system applies a Dynamic Programming algorithm to optimize the final list of recommended products based on user-defined constraints (e.g., budget, quantity, compatibility).

4.2 ***Information Retrieval from Milvus***

Once MongoDB returns the list of relevant milvus_id entries, the system queries Milvus to further rank and refine the results based on content similarity.

Milvus is used to perform semantic vector-based search using vector embeddings of various product fields. The system performs sa imilarity search across the following fields:

- information: Detailed description of the product or location.
- name: The name of the destination or product.
- location: The address or geographic information.

To rank products based on comprehensive relevance, the system calculates the average similarity score across these fields.

- Field-level similarity: Computed using cosine similarity between the vector embedding of the user query and the vector embedding of each corresponding field in the product data.

Aggregate Similarity Score ($similarity_{avg}$):

$$similarity_{avg} = \frac{similarity_{inormation} + sim_{name} + sim_{locaation}}{3}$$

After obtaining a ranked list from Milvus based on semantic relevance, the system reconnects to MongoDB to retrieve full product details, including all structured information fields, which are then compiled into the final response delivered to the user.

4.3 **Integration of LLMs for Rewriting**

The information retrieved from the selected products is passed to a Large Language Model (LLM) to be rewritten and formulated into natural language responses that address the user's informational needs. At the same time, the system returns detailed information for the Top 10 locations that best match the user's query.

5. EXPERIMENTAL EVALUATION

5.1. Chatbot Deployment on Messenger and Web Interface

To ensure wide accessibility and ease of use, our travel chatbot was deployed on both Facebook Messenger and a custom web interface. This dual-platform deployment allows users to interact with the system in real time, either through social media or directly via browser.

Web Interface: A responsive React-based web client was developed for browser-based access. This interface interacts with the backend API via HTTPS and supports text-based queries, displays results with images and detailed location descriptions, and enables real-time feedback collection, as shown in **Figure 2**.

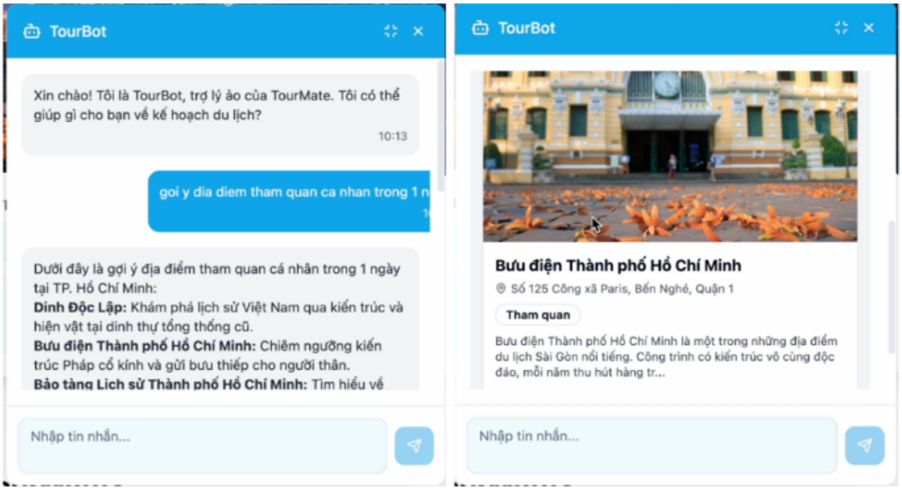

FIG 2. INTERFACE OF WEBSITE

(Source: Own elaboration)

Facebook Messenger Integration: The chatbot was integrated using the BotBanHang platform, which enables message handling, user authentication, and webhook configuration. Users can ask travel-related questions naturally and receive intelligent, context-aware answers, as illustrated in **Figure 3:**

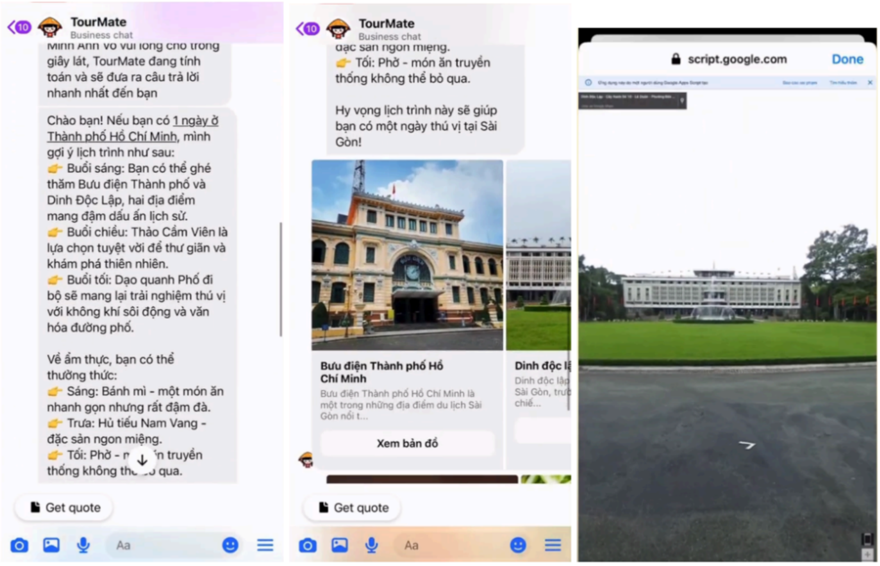

FIG 3. INTERFACE OF MOBILE APP

5.2 Evaluation

To evaluate the performance of the proposed travel chatbot system, both quantitative and qualitative indicators were employed. Our evaluation focuses on the system's capacity to retrieve relevant information, correctly interpret user intent, extract structured entities, and generate timely responses. The experiments were conducted using real-world travel queries through both the web and Messenger interfaces. In particular, five evaluation metrics were adopted: **Precision** (the proportion of correct answers among the system's responses), **Recall** (the proportion of relevant answers successfully retrieved), **Relevance** (the extent to which responses align with the user's intent, as judged by human evaluators), **Entities Recall** (the accuracy in extracting details such as location, category, mood, or price), and **Latency** (the average response time of the system). Together, these measures capture both the technical accuracy and the practical effectiveness of the chatbot in supporting real travel scenarios. The overall results across these metrics are presented in Table 1:

Precision	ContextPrecision	Recall	Relevance	Entities Recall	Time (s)
0.95	0.89	1	0.91	0.95	8.14

***Table 1** Summarizes the evaluation results across key metrics*

The Hybrid approach, which combines MongoDB for structured filtering and Milvus for semantic vector search, demonstrates exceptional performance in handling complex, multi-faceted queries. As shown in Table 1, the method achieves strong evaluation outcomes: Precision reaches 0.93, Context Precision is 0.90, Recall remains perfect at 1.00, Entities Recall stands at 0.95, and Relevance scores 0.91. These metrics highlight the system's ability to deliver accurate, context-aware, and semantically rich results. In addition, the system processes each query in an average of 8.14 seconds, reflecting both high computational efficiency and fast user response times.

These findings confirm that the hybrid design not only ensures technical robustness but also aligns with the practical needs of real-world travel assistance. By effectively combining structured filtering with semantic retrieval, the system can provide travelers with recommendations that are both precise and contextually relevant. Such performance underscores the feasibility of integrating advanced AI-driven search mechanisms into tourism platforms, thereby supporting more personalized services, improving user satisfaction, and contributing to the broader goals of digital transformation and innovation in the tourism industry.

As shown in **Table 2**, the hybrid approach leverages the combined strengths of MongoDB and Milvus to provide a robust and versatile search framework, supporting filtering, semantic retrieval, and vector-based search simultaneously. This integration enables the system to balance high-speed performance with real-time responsiveness while also extending support for multilingual interaction. In contrast, using MongoDB or Milvus independently provides only partial functionality and lacks the flexibility required to process complex or nuanced user queries.

The findings highlight the advantages of the hybrid method in managing intricate queries and demonstrate its potential for seamless integration with advanced techniques such as Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG). Such capabilities are particularly valuable in application domains where accuracy, responsiveness, and multilingual support are critical. For instance, the tourism and hospitality sectors can greatly benefit from search systems that deliver fast, precise, and context-aware recommendations, thereby enhancing user experience and supporting sustainable service innovation.

Factors	MongoDB	Milvus	Hybrid Method (MongoDB + Milvus)
Filter-based Search	✓	x	✓
Vector-based Search	x	✓	✓
Semantic Search	x	✓	✓
Search Speed	✓	✓	✓
Feedback Tuning Capability	x	✓	✓
Handling Complex Data	✓	✓	✓
Handling Complex Queries	✓	✓	✓
Real-time Data Processing Capability	✓	✓	✓
Multilingual Support	x	✓	✓

Table 2 *Comparison of MongoDB, Milvus, and Hybrid (MongoDB + Milvus) Methods across Various Search Factors.*

5.3 Qualitative Analysis

To complement the quantitative evaluation, a qualitative analysis was performed to better understand system behavior.

For example, the query *"Find romantic cafes in District 1"* was correctly parsed with entities {location: District 1, category: cafe, mood: romantic}, and the chatbot returned three highly relevant venues, including rooftop cafés with ambient lighting that matched the intended user mood.

In contrast, the query *"Suggest adventurous activities near Saigon"* produced less precise matches, with some generic suggestions such as indoor games, rather than outdoor experiences that better fit the "adventurous" mood. This indicates that while the system performs well for common categories and moods, it remains limited in handling niche experiential styles. Future improvements may include expanding the taxonomy of moods and enriching training data with diverse activity types.

6. DISCUSSION AND FUTURE WORKS

6.1 DISCUSSION

The development of TourMate represents a significant effort to harness the potential of artificial intelligence, geospatial technologies, and multimedia integration in addressing the evolving needs of modern travelers in Vietnam. This study has demonstrated that intelligent platforms, when designed with user-centric principles and supported by robust data infrastructures, can greatly enhance the travel experience by offering context-aware information, real-time support, and culturally meaningful interactions.

From **Table 2**, it can be observed that the Milvus-only configuration achieves nearly the same semantic performance as the hybrid approach. This indicates that semantic vector search is already highly effective in capturing contextual nuances. However, MongoDB adds unique value by enabling filter-based retrieval, such as filtering by price range, opening hours, or exact location tags, which Milvus alone cannot efficiently support. Therefore, the hybrid design combines semantic richness with symbolic filtering, providing a more practical and flexible solution for real-world travel assistance.

One key point of discussion lies in the way TourMate facilitates not only trip planning and navigation but also engagement with authentic local experiences. The inclusion of 3D mapping, personalized itineraries, and natural language interaction promotes greater accessibility, particularly for international tourists who may face language or cultural barriers. Furthermore, the system's ability to dynamically adapt its recommendations based on user preferences and real-time conditions contributes to its practicality and relevance in diverse travel contexts.

However, the integration of advanced technologies into the tourism sector also raises several complex challenges, especially concerning data governance, user trust, and digital inclusivity. While TourMate aims to mitigate these risks through ethical design and adherence to privacy regulations, continuous monitoring and stakeholder involvement will be essential to ensure its responsible deployment.

Moreover, from a broader socio-economic perspective, TourMate presents an opportunity to support local businesses and communities by directing tourists toward verified, culturally rich, and sustainable services. This aligns with current national strategies for digital transformation and sustainable tourism development in Vietnam. The platform can also act as a digital bridge, connecting tourists with underrepresented destinations, thereby contributing to more balanced tourism flows.

Despite the promising outcomes, this study acknowledges several limitations that may influence the interpretation and generalizability of the results.

First, the prototype implementation and user testing were conducted within a limited geographic and demographic scope. While the platform was designed to support a wide range of users, including international tourists, the actual testing population consisted primarily of domestic users with access to digital devices and internet connectivity. This may introduce bias in the evaluation of usability and system effectiveness.

Second, the dynamic nature of travel-related data—such as changes in local conditions, transportation schedules, or service availability—poses an ongoing challenge for maintaining data accuracy and relevance. TourMate currently relies on third-party APIs and manual data curation, which may limit its scalability and real-time responsiveness in certain scenarios.

Third, while the platform includes basic multilingual support and chatbot interaction, its natural language processing capabilities are still in early development stages. This may affect the depth and accuracy of conversations, particularly for users with complex or nuanced queries.

Finally, the ethical and legal considerations discussed, although addressed in principle, require long-term operational mechanisms and institutional partnerships to be fully realized in practice. Without consistent oversight, even well-intentioned systems can encounter unforeseen consequences, such as algorithmic bias or privacy breaches.

6.2. Future Work

Future research directions should focus on expanding the technical capabilities, contextual adaptability, and ethical safeguards of the TourMate platform.

From a technological perspective, advancements in natural language processing—particularly in multilingual and low-resource language contexts—will be critical to improving the platform's ability to serve a diverse user base. Integration with IoT-based environmental sensing and real-time traffic data could further enrich the personalization and situational awareness features of the platform.

Moreover, future studies should explore the deployment of TourMate in various geographic locations, including rural and less-visited destinations, to assess its adaptability and socio-economic impact. This includes evaluating how the platform influences tourist behavior, supports local service providers, and contributes to cultural preservation.

Ethically, additional research is needed to develop frameworks for participatory governance of smart tourism systems, where users and local communities are not merely data subjects but active stakeholders in shaping the digital tourism landscape. The long-term sustainability of such systems will depend not only on technical robustness but also on community trust, institutional support, and cross-sector collaboration.

Finally, given the increasing relevance of climate-conscious tourism, future versions of TourMate may incorporate features that help users reduce their environmental footprint—such as promoting low-emission transport options or eco-certified accommodations—thereby aligning with global efforts toward sustainable and responsible travel.

7. ACKNOWLEDGMENTS

“The authors would like to express heartfelt appreciation to the International School-Vietnam National University, Hanoi (VNU-IS) for its financial support. Additionally, we are indebted to the anonymous reviewers whose insightful feedback significantly enriched the quality of this work.”

REFERENCES

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Wang, H., Li, J., Chen, Z., Zhang, Y., & Liu, X. (2024). Unims-RAG: A unified multi-source retrieval-augmented generation for personalized dialogue systems. *arXiv preprint* arXiv:2401.13256.

TravelRAG. (2024). A tourist attraction retrieval framework based on multi. *ISPRS International Journal of Geo-Information, 13*(11), Article 2024.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51–65.

Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: Foundations and developments. *Electronic Markets, 25*(3), 179–188.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint* arXiv:1810.04805.

Ha, M. H., & Chen, Q. T. C. (2021). Deep neural networks using residual fast-slow refined highway and global atomic spatial attention for action recognition and detection. *IEEE Access, 9*, 164887–164902. https://doi.org/10.1109/ACCESS.2021.3136494

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Pesonen, J., & Neidhardt, J. (Eds.). (2019). *Information and Communication Technologies in Tourism 2019.* Cham: Springer. https://doi.org/10.1007/978-3-030-05940-8

Derham, R., Di Gangi, P., & Dolen, W. (2020). Intelligent agents in tourism: A systematic review. *Tourism Management Perspectives, 35*, 100713. https://doi.org/10.1016/j.tmp.2020.100713

Bekele, H. and Raj, S. (2024), "Digitalization and digital transformation in the tourism industry: a bibliometric review and research agenda", Tourism Review. https://doi.org/10.1108/TR-07-2023-0509

Izacard, G. & Grave, E. (2020). *Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering*. arXiv:2007.01282.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., … & Riedel, S. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. Advances in Neural Information Processing Systems (NeurIPS), 33, 9459–9474.

Chowdhery, A. et al. (2022). *PaLM: Scaling Language Modeling with Pathways*. arXiv:2204.02311

Bender, E. et al. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*. FAccT Conference.

Følstad, A., & Skjuve, M. (2019). *Chatbots for customer service: User experience and motivation*. *Proceedings of the 1st International Conference on Conversational User Interfaces*.

Zamora, J. (2017). *Rise of the Chatbots: Finding A Place for Artificial Intelligence in Service Encounters*. *In Digital Marketing*.