

Developing Conversational AI to Enhance the Tourist Experience: A Methodological Framework Applied to the Case of Porto

Bruno Jardim¹, Ricardo Dona¹, Miguel de Castro Neto¹ and Sandra Barnabé²

¹NOVA Information Management School, Lisboa, Portugal

²Ubiwhere, Aveiro, Portugal

20221359@novaims.unl.pt

bjardim@novaims.unl.pt

mneto@novais.unl.pt

sbarnabe@ubiwhere.com

Abstract: Tourism in Porto has expanded considerably over the last decade, with the number of visitors doubling between 2009 and 2019 and continuing to rise in recent years. This rapid growth creates challenges for the sustainable management of tourist inflows and highlights the need for innovative, data-driven solutions that can both support visitors and relieve pressure on local infrastructures. In response, this study proposes the development of a conversational AI system specifically designed to assist tourists in their daily activities and improve their overall experience of the city. The conversational AI is based on a Retrieval-Augmented Generation (RAG) framework, which combines an information retrieval component with a generative model to deliver accurate, context-aware responses. To ensure reliability, the system draws on more than 7,000 documents from diverse sources, including cultural guides, tourism platforms, and official city websites. Several experiments were conducted to identify the best performing system configuration, testing different retrieval strategies, ranking methods, and model architectures. The final solution demonstrates high retrieval accuracy and generates responses that score strongly on semantic similarity and answer quality metrics. Overall, the study demonstrates the potential of conversational AI systems as valuable tools for urban destinations facing growing tourist demand. Beyond Porto, this work illustrates how data-efficient conversational systems can support sustainable tourism management, improve the visitor experience, and serve as scalable solutions for cities with similar challenges worldwide.

Keywords: Smart Tourism; Tourist Experience; Conversational AI; Chatbot; Large Language Models

1. Introduction

Artificial Intelligence (AI) is reshaping the global tourism sector, driving innovation in visitor experience, management efficiency, and destination competitiveness. Within tourism, AI applications such as predictive analytics, recommendation systems, and conversational systems are increasingly recognized as enablers of smart destinations - urban environments that combine data, technology, and human-centered design to deliver more sustainable and adaptive tourism experiences (Gretzel et al., 2015; Buhalis & Amaranggana, 2015).

Among the most transformative applications of AI in this field is Conversational AI, which allows tourists to access information and services through natural language interactions. Conversational AI encompasses chatbots and virtual assistants capable of understanding, processing, and responding to human language through text or voice (Benaddi et al., 2024). They are increasingly used to answer questions, recommend attractions, manage bookings, and support visitors during their stay. However, many existing solutions remain limited in contextual awareness, often relying on rule-based interactions or pre-defined responses that restrict their capacity to handle complex or nuanced queries (Adamopoulou & Moussiades, 2020; Maryamah et al., 2024).

Recent advances in large language models (LLMs) and information retrieval have enabled more adaptive and context-sensitive conversational systems capable of generating coherent and domain-relevant responses (Benaddi et al., 2024). These developments strengthen the role of Conversational AI in tourism as both a service interface and a strategic tool for destination management, enhancing the quality and inclusivity of visitor experiences while supporting more efficient use of resources.

Building on these advancements, the present study explores how Conversational AI can be applied to support tourists' daily activities in urban destinations. The research focuses on developing a domain-specific chatbot that assists visitors in obtaining reliable, localized, and timely information. Porto was selected as the case study due to its continuous tourism expansion—doubling the number of visitors between 2009 and 2019 (Pavel & Romeiro, 2023) - and its position as one of Europe's leading city destinations. These dynamics intensify the need for digital tools that can improve visitor engagement while contributing to the sustainable management of tourism flows.

Although the city's main tourism platform, Visit Porto¹, offers valuable content, it could benefit from a conversational interface that provides personalized and interactive assistance to tourists.

Beyond its technical implementation, this work contributes to tourism research by illustrating how Retrieval-Augmented Generation systems can support reliable, scalable, and context-aware information provision in urban tourism settings. By grounding conversational responses in destination-specific knowledge, the proposed approach has implications for tourism management, visitor experience, and destination governance within smart tourism ecosystems.

2. Literature Review

2.1 Conversational AI in Tourism

The concept of *smart tourism* has become central to the digital transformation of destinations, combining data infrastructures, intelligent systems, and participatory governance to enhance competitiveness and sustainability (Buhalis & Amaranggana, 2015). More recent research positions Artificial Intelligence (AI) as a key component of this transformation, supporting predictive analytics, visitor flow management, and personalized experience design (Dyduch & Brzozowska, 2025; Kim, So, Shin, & Li, 2024).

AI in tourism has progressively evolved from automation toward human-centered design, integrating data-driven systems with user interfaces that facilitate natural communication and inclusivity (Calvaresi et al., 2021). Within this framework, Conversational AI has emerged as a vital interface between visitors and destination knowledge, allowing travelers to access information through natural, context-aware dialogue. This aligns with the broader goals of smart tourism — efficiency, sustainability, and participatory governance (Gretzel et al., 2015; Samancioglu et al., 2024).

Conversational AI refers to systems capable of interpreting and generating natural language to facilitate interaction between humans and computers (Adamopoulou & Moussiades, 2020). In tourism, these systems have been implemented for trip planning, bookings, on-site recommendations, and destination support (Benaddi et al., 2024). Chatbots and virtual assistants have demonstrated potential to improve service efficiency and visitor satisfaction by offering continuous, personalized communication (Li et al., 2021).

Earlier systems (pre-2022) were predominantly rule-based or retrieval-based, relying on pattern recognition and predefined templates (see the example of Amalia & Suprayogi (2019)'s chatbot). While functional for structured tasks, these systems struggled to handle complex or multi-intent questions (Casillo et al., 2020; Alhumoud et al., 2022). Recent advances in large language models (LLMs) have significantly expanded the scope of Conversational AI in tourism, enabling adaptive, contextually grounded dialogue. LLM-based systems increasingly employ hybrid retrieval architectures to access verified destination knowledge while generating coherent, domain-specific responses. These innovations have opened possibilities for destination-wide assistants that can engage with tourists across multiple information domains.

2.2 Destination-Level Conversational AI Applications

In recent years, the number of empirical studies describing destination-scale conversational systems has begun to increase, although still modestly. Several implementations illustrate the gradual evolution from narrow, rule-based chatbots toward domain-grounded assistants capable of managing diverse information needs across a city or region. Anand et al. (2023) presented a convolutional neural-network-based chatbot that integrates recommendation and retrieval modules for multiple Indian cities, while Sperlí (2021) developed a generative model using a sequence-to-sequence recurrent architecture to support cultural-heritage guidance in Italy. Both studies demonstrated improvements in conversational flexibility compared to static interfaces but remained limited by restricted corpora and the absence of continuous knowledge updates.

More recently, large-language-model (LLM) approaches have been applied to tourism settings. Volchek and Ivanov (2024) examined ChatGPT's capabilities as a travel itinerary planner. They found that while ChatGPT could

¹ <https://visitporto.travel/>

quickly generate itineraries that were easy to understand and accessible, these AI-generated plans were less accurate and less specific than those created by human tourism experts. The evaluation of 11 quality criteria showed that ChatGPT's itineraries often included only well-known attractions and missed detailed local insights. The authors concluded that ChatGPT is a useful tool for initial travel inspiration and draft itineraries, but it cannot yet serve as an exclusive tool for trip planning without human vetting.

In another study, Kirtl et al. (2024) focused on fine-tuning a GPT-3.5 model with domain-specific tourism data to create a tailored chatbot for niche travel topics. By training on specialized local content, the fine-tuned chatbot provided more accurate, human-like answers to user queries in that domain, outperforming the base ChatGPT model which often produced misleading or hallucinated responses on those specific topics. The fine-tuned model's responses were observed to be significantly superior to the untrained ChatGPT-3.5/4, especially in handling local heritage and culturally specific questions. This result suggests that customizing large language models with up-to-date, domain-focused data can substantially improve a tourism chatbot's relevance and reliability for particular destinations or themes.

Similarly, Cruz et al. (2025) introduced LISA, a city-wide conversational assistant for Lisbon that combines hybrid retrieval (dense + sparse search) with a generative model (ChatGPT) to provide information and recommendations from thousands of curated tourism documents. The system underwent both automated and expert evaluations. This highlights the effectiveness of LLM-driven chatbots in delivering accurate answers and engaging experiences for tourists.

Overall, research on conversational AI in tourism remains fragmented and lacks methodological consistency. Few peer-reviewed studies detail comprehensive destination-level systems, and available reports often omit crucial information about data, retrieval, and evaluation. This study introduces a RAG-based Conversational AI for a major urban destination, using a transparent, reproducible approach to support sustainable digital transformation and improve visitor experiences.

3. Methodology

This study develops a Conversational AI system designed to support tourists visiting Porto by providing accurate, context-based recommendations and information². The methodological approach is based on Jeong's (2023) six-step framework for the development of Retrieval-Augmented Generation (RAG) applications, adapted here to the tourism context.

3.1 Data Collection and Preparation

The dataset was constructed following Buhalis's (2000) *6 A's Framework*, which defines six fundamental components of a tourism destination: Attractions, Accessibility, Amenities, Available Packages, Activities, and Ancillary Services. This framework ensured that the conversational AI system could address a broad range of tourist inquiries, from accommodations and dining to events and local transportation.

² https://github.com/novacidade-org/RAGChatbot_Porto

Data were gathered from seven complementary sources - VisitPorto³, TimeOut⁴, Civitalis⁵, Booking⁶, Tripadvisor⁷, Wikipedia⁸, and Visitar-Porto⁹ - encompassing both experiential and operational content. The final corpus contained 7,382 documents, with Booking and Tripadvisor representing the largest share. In addition, manually curated summaries were developed to capture generic tourist intents (e.g., “*What are the main events in Porto?*”). After collection, all documents were normalized and cleaned to remove duplicates, empty entries, and formatting inconsistencies. Finally, Text chunking was implemented using LangChain’s RecursiveCharacterTextSplitter, testing chunk sizes of 3,000 and 7,000 characters without overlap. The larger chunk configuration preserved semantic coherence in longer narratives such as itineraries.

3.2 Model Architecture

The conversational AI system integrates two core modules — retrieval and generation — connected through the LangChain framework (LangChain, 2024). This architecture transforms a user query into an enriched prompt, combining retrieved factual knowledge with natural language generation to produce grounded and contextually relevant responses (Lewis et al., 2020). The overall workflow includes embedding generation, vector storage, retrieval, and answer synthesis (Figure 1).

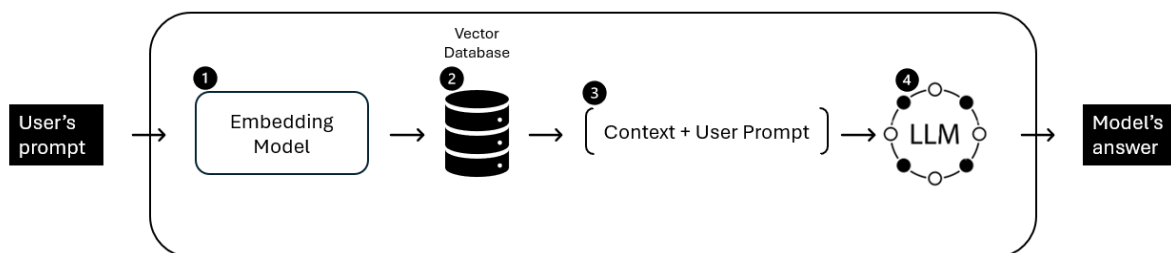


Figure 1: LLM enhanced through the Retrieval Augmented Generation process.

The embedding model selected was gte-large-en-v1.5 (Zhang et al., 2024), chosen for its high semantic retrieval performance on the Massive Text Embedding Benchmark (MTEB) (Muennighoff et al., 2022) and extended input capacity of 8,192 tokens. This configuration allows the model to encode long tourism-related document chunks efficiently while preserving contextual information.

All embeddings were stored in ChromaDB (Chroma, 2024), an open-source vector database optimized for dense retrieval and similarity search. The cosine similarity metric was applied to rank documents by semantic closeness. Four separate databases were generated to test the impact of chunk-size and text normalization (lowercasing and punctuation removal) on retrieval accuracy.

The retrieval stage determines which contextual passages are provided to the generator. Three configurations were tested: Dense Retrieval, relying solely on semantic embeddings; Hybrid Retrieval, combining semantic and lexical searches through the BM25 algorithm (Ma et al., 2021; Karpukhin et al., 2020); Hybrid Retrieval with Re-

³ www.visitporto.travel/

⁴ www.timeout.com/porto

⁵ www.civitalis.com/en/

⁶ www.booking.com

⁷ www.tripadvisor.com

⁸ <https://www.wikipedia.org/>

⁹ <https://www.visitar-porto.com/en/>

ranking, which employed the ms-marco-MiniLM-L-6-v2 cross-encoder to reorder retrieved results by semantic relevance (Akkiraju et al., 2024; Erak et al., 2024).

In the hybrid configurations, Reciprocal Rank Fusion (RRF) (Cormack et al., 2009) was applied to merge ranked results from both retrievers. Text preprocessing ensured alignment between lexical and semantic search — all text was converted to lowercase and punctuation selectively removed to avoid splitting named entities such as “Porto.” and “Porto”.

For the generation module, two transformer-based models were evaluated: Llama 3.1 – 8B (Dubey et al., 2024) and Llama 3.2 – 3B (Meta, 2024). Both are multilingual decoders capable of processing up to 128,000 tokens. The smaller model, obtained through pruning and distillation of Llama 3.1, was compared to the larger version to assess trade-offs between computational cost and quality.

The temperature parameter was fixed at 0 to ensure factual and deterministic outputs (Maryamah et al., 2024). Prompt engineering followed the role-prompting approach (Chen et al., 2023), explicitly instructing the model to respond only to questions within the Porto tourism domain. The system message defined was the following:

“This is a Chatbot that only answers to questions related to Porto (Portugal) Tourism. More specifically, to topics related to attractions, accessibility, amenities, activities, available packages, and Ancillary Services. When not specified by the user assume the question is related to Porto. If the question is not about Porto Tourism just write: "I am sorry, but my knowledge only allows me to help you with Porto Tourism topics. Can I help you with something related to Porto Tourism?" Answer to the user's question objectively, using correct syntax and based on context written below:

{context}\n

User: {input}”

This design ensured factual consistency, thematic focus, and reduction of hallucinations during inference, aligning with best practices in grounded conversational AI (Chen et al., 2023). Additionally, the system incorporates chat history to support follow-up questions, implemented using the “create_history_aware_retriever” method from LangChain. This ensures that users can have coherent, multi-turn conversations with the chatbot.

3.3 Evaluation and Experimental Setup

To assess the system’s retrieval and generation performance, a benchmark dataset was developed comprising 71 question–answer pairs and 118 context passages, reflecting the diversity of user intents within the tourism domain. The dataset was divided into three categories — generic, specific, and non-related queries — to evaluate not only factual accuracy but also domain adherence and fallback behaviour. “Generic” relates to topics such as “What are the best places to visit in Porto?” and “What to do in Porto in one day?”. “Specific” questions refer to questions about a particular fact, like “When is the São João festival?” or “What is the Check in time at Vincci Bonjardim?”. Lastly, non-related aspects refer to questions like “Tell me about Algarve history?” or “What are the best attractions in Lisbon?”, since the goal in these situations is to specifically answer “I am sorry, but my knowledge only allows me to help you with Porto Tourism topics. Can I help you with something related to Porto Tourism?”

Three complementary metrics were employed to evaluate different aspects of system performance based on previous work: Hit Rate (HR) was used to assess retrieval effectiveness by measuring the proportion of correctly retrieved documents (Gao et al., 2023; Salemi & Zamani, 2024); Semantic Answer Similarity (SAS) quantified the semantic overlap between the generated and reference responses, using a cross-encoder-based scoring approach (Risch et al., 2021); BertScore (Zhang et al., 2019) was applied to measure contextual similarity between the predicted and reference answers using token-level embeddings from BERT.

This combination of retrieval and generation-oriented metrics provided a robust evaluation of the conversational AI’s grounding and factual reliability, consistent with recent approaches in retrieval-augmented systems (Es et al., 2024; Doan et al., 2024).

In this phase, two main stages of experimentation were conducted.

1. Retriever Optimization: A grid search tested 56 combinations of parameters, varying chunk-size (3,000 vs. 7,000 characters), retriever type (dense, hybrid, hybrid with re-ranking), and Top-K

retrieval size (3 to 9). Each configuration was evaluated according to Hit Rate, identifying the setup that maximized recall without introducing excessive noise.

2. Model Comparison: Four configurations were tested — baseline Llama 3.1–8B, baseline Llama 3.2–3B, and both models integrated with the best-performing retriever (see Table 3 in section 4). Performance was compared using SAS and BertScore to determine the most effective generator–retriever pairing.

The optimal setup combined a Hybrid Retriever with Re-ranking and Llama 3.1–8B as the generator. This configuration achieved the highest retrieval precision and semantic alignment, establishing it as the final architecture for subsequent testing and discussion.

4. Results and Discussion

This section presents the quantitative and qualitative results of the evaluation, followed by a discussion of their implications for conversational AI in tourism. The analysis focuses on retrieval efficiency, generative accuracy, and the contribution of each model configuration to the overall system performance.

4.1 Retrieval Performance

The retrieval phase tested three configurations — Dense Retrieval, Hybrid Retrieval, and Hybrid Retrieval with Re-ranking — across two chunk sizes (3,000 and 7,000 characters) and several Top-K values. Figure 2 displays the main outcomes.

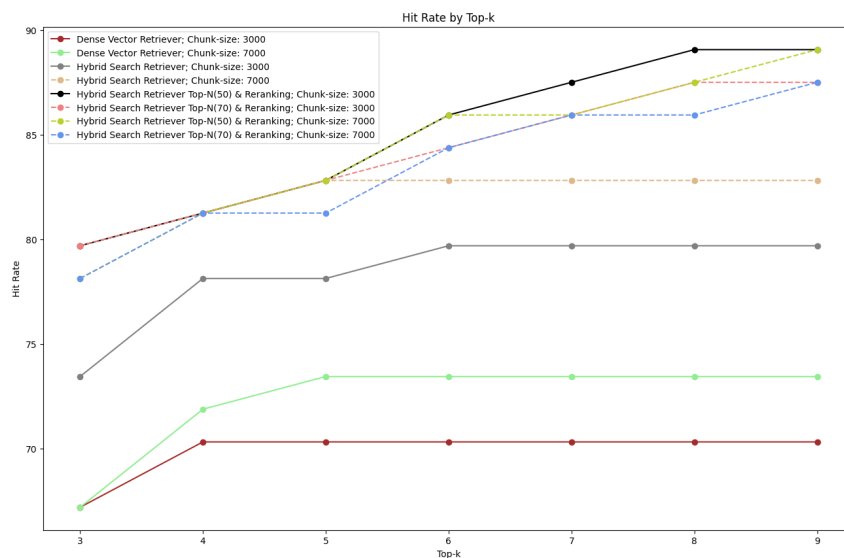


Figure 2: Hit Rate evolution by Retriever configuration.

In the final evaluation phase of our system, two distinct retriever configurations achieved a Hit Rate of 89%, marking the highest level of retrieval accuracy observed throughout the experiments. While this result demonstrates that multiple approaches can reach strong performance, a closer analysis reveals that one configuration stands out for its efficiency and practical value. Specifically, the Hybrid Search Retriever with re-ranking not only matched the top Hit Rate but did so by retrieving fewer context passages, thereby reducing the amount of irrelevant or noisy information presented to the generative model.

The optimal configuration was built using Alibaba’s gte-large-en-v1.5 embedding model, which is known for its robust semantic retrieval capabilities. The data was segmented into chunks of up to 3,000 characters, a choice that balanced the need for semantic coherence with computational efficiency. Retrieval parameters were set with a Top-N of 50 and a Top-K of 8, ensuring that the system considered a broad yet manageable set of candidate passages for each query.

Figure 1 further illustrates the superiority of the Hybrid Search Retriever with re-ranking, showing that it consistently outperformed other configurations as the Top-K parameter increased beyond 5. This trend is consistent with the findings of Erak et al. (2024), who also observed that hybrid approaches with re-ranking mechanisms tend to yield better retrieval outcomes in complex information domains. The present study’s results

are in line with the conclusions drawn by Ma et al. (2021) and Akkiraju et al. (2024), both of whom highlighted the advantages of hybrid retrieval strategies over purely dense vector-based methods.

In summary, while several retriever setups could achieve high retrieval accuracy, the Hybrid Search Retriever with re-ranking distinguished itself by delivering this performance more efficiently and consistently. These findings reinforce the growing consensus in the literature that hybrid retrieval architectures, particularly those enhanced with re-ranking, represent a best practice for retrieval-augmented generation systems in knowledge-intensive applications.

4.2 Generative Performance

The second experimental stage compared four model configurations. Table 1 presents the comparative results using Semantic Answer Similarity (SAS) and BertScore metrics.

Table 1: Results comparing RAG models with baseline models with Llama3.2-3B and Llama3.1-8B.

Generator Model	BertScore	SAS
Llama3.1-8B - baseline	0,72	0,52
Llama3.1-8B w/ RAG	0,82	0,64
Llama3.2-3B - baseline	0,72	0,54
Llama3.2-3B w/ RAG	0,80	0,61

The Llama 3.1–8B + Hybrid Retriever with Re-ranking configuration achieved the highest performance across both metrics: SAS = 0.64 and BertScore = 0.82, outperforming all other setups. In contrast, the baseline Llama 3.1–8B without retrieval scored SAS = 0.52 and BertScore = 0.72, demonstrating the substantial impact of retrieval grounding on factual completeness and semantic precision.

The smaller Llama 3.2–3B model achieved a SAS of 0.54 and BertScore of 0.72 when paired with the hybrid retriever. This result is particularly relevant for practical applications, as it suggests that lighter architectures can deliver competitive accuracy with lower computational requirements, making them suitable for deployment in tourism organizations with limited infrastructure.

This improvement is consistent with findings by Lewis et al. (2020) and Akkiraju et al. (2024), who emphasized that retrieval-augmented architectures improve factual reliability in knowledge-intensive tasks.

4.3 Discussion

These findings contribute to addressing a gap in the literature concerning the limited adoption of advanced conversational AI systems in the tourism sector. By adapting approaches previously explored in other economic domains—such as technology (Akkiraju et al., 2024), telecommunications (Erak et al., 2024), medicine (Singh et al., 2024), and education (Alsafari et al., 2024; Maryamah et al., 2024)—this work introduces to tourism a set of techniques related to Retrieval-Augmented Generation (RAG), hybrid retrieval, re-ranking strategies, and evaluation methods. The integration of these techniques enabled the development of a solution tailored to the specific informational needs and contextual constraints of urban tourism in Porto.

At the same time, the application of conversational AI in tourism raises several practical and governance-related challenges that must be considered. The effectiveness of such systems is highly dependent on the quality, coverage, and timeliness of the underlying data sources, which may limit reliability if content becomes outdated or incomplete. In addition, the interpretation of user queries and the generation of responses introduce potential risks related to bias, oversimplification of complex information, and misalignment with official policies or regulations. From a destination management perspective, deploying conversational systems also entails operational constraints, including system maintenance, performance monitoring, and the need to ensure transparency and accountability in automated information provision.

Despite these challenges, the proposed system demonstrates potential to support tourism management by facilitating engagement between Porto and its visitors and by enabling more data-informed decision-making. Destination management organizations may leverage aggregated interaction data to better understand visitor interests and information needs, as suggested by Liberato et al. (2018), while tools such as LangSmith (LangChain, 2024) allow for monitoring system behaviour and usage patterns. To support experimentation and future user-

based evaluations, a front-end interface was developed using Streamlit¹⁰, enabling controlled interaction with the conversational agent (Figure 3).

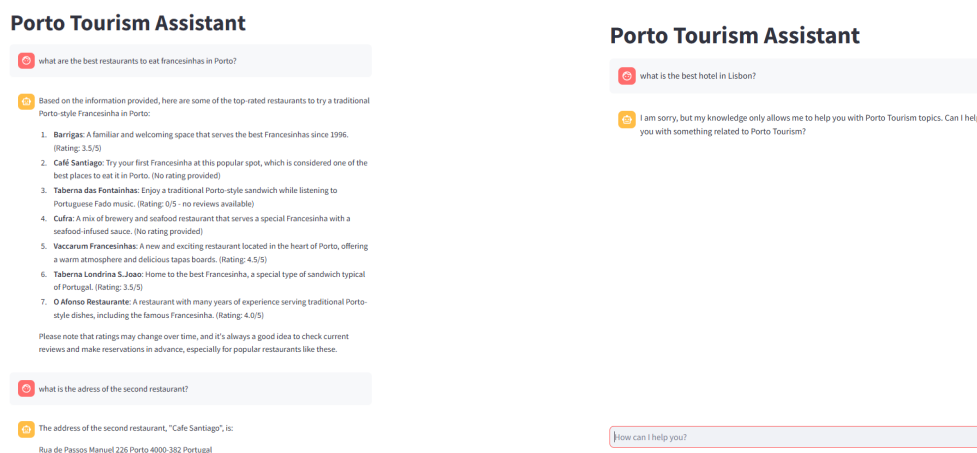


Figure 3: Chatbot conversation with one relevant question and another unrelated to Porto Tourism.

5. Conclusions, Limitations and Future Work

This study designed and assessed a conversational AI system for delivering tourism information in Porto, using a Retrieval-Augmented Generation (RAG) framework enhanced by hybrid retrieval and re-ranking techniques. The integration of these methods with the Llama 3.1–8B model yielded the highest performance, while the more compact Llama 3.2–3B also showed strong results, highlighting the potential of lightweight solutions for practical tourism deployment. The system improved information accuracy and contextual relevance, demonstrating the value of conversational AI for improving visitor experience and supporting destination management.

This study is subject to several limitations. It focuses on a single destination (Porto), relies on static and English-only data sources, and introduces some latency due to the re-ranking process. In addition, the evaluation is limited to automated metrics, as no user-based assessment was conducted. Future work should address these limitations by incorporating real-time and multilingual data, improving retrieval efficiency, enabling adaptive learning mechanisms, and conducting user-centered evaluations to assess usability and real-world impact.

Overall, the findings offer a robust framework for scalable, data-driven conversational AI to support digital innovation in tourism, while laying the groundwork for further research that includes user-based evaluation and system improvements.

Acknowledgements

This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the project - UID/04152/2025 - Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS - <https://doi.org/10.54499/UID/04152/2025> (2025-01-01/2028-12-31) and UID/PRR/04152/2025 <https://doi.org/10.54499/UID/PRR/04152/2025> (2025-01-01/ 2026-06-30). This work is also funded by PRR – Plano de Recuperação e Resiliência and by the NextGenerationEU funds, through the scope of the Agenda for Business Innovation “ATT – Agenda Mobilizadora Acelerar e Transformar o Turismo” (Project no. 47 with the application C645192610-00000060).

AI Declaration

This paper was partially developed with the assistance of the AI tool ChatGPT (OpenAI, GPT-5 model). The tool was used to support the drafting and refinement of text, including grammar checking, rephrasing for clarity, and

¹⁰ <https://streamlit.io>

generating structure suggestions. All ideas, analyses, and final interpretations presented in the paper are the author's own, and all outputs generated by the AI were reviewed, edited, and validated by the author before inclusion.

Ethics Declaration

This research did not involve human participants, personal data, or any activity requiring formal ethical approval according to institutional or national guidelines. Therefore, ethical clearance was not required for the development of this study.

References

- Akkiraju, R., Xu, A., Bora, D., Yu, T., An, L., Seth, V., Shukla, A., Gundecha, P., Mehta, H., Jha, A., Raj, P., Balasubramanian, A., Maram, M., Muthusamy, G., Annepally, S. R., Knowles, S., Du, M., Burnett, N., Javiya, S., ... Boitano, J. (2024). FACTS About Building Retrieval Augmented Generation-based Chatbots. <https://arxiv.org/html/2407.07858v1>
- S. ALHumoud et al., "Rahhal: A Tourist Arabic Chatbot," *2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, Riyadh, Saudi Arabia, 2022, pp. 66-73, doi: 10.1109/SMARTTECH54121.2022.00028.
- Alsafari, B., Atwell, E., Walker, A., & Callaghan, M. (2024). Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants. *Natural Language Processing Journal*, 8, 100101. <https://doi.org/10.1016/j.nlp.2024.100101>
- Amalia A, Suprayogi MS (2019) Engaging millennials on using chatbot messenger for eco-tourism. *Proc Third Int Conf Sustainable Innov 2019 – Humanity Educ Social Sci (IcoSIHESS 2019)* 484–487. <https://doi.org/10.2991/icosihess-19.2019.84>
- Anand S, Abhishek Sai AM, Karthikeya M (2023) Chatbot Enabled Smart Tourism Service for Indian Cities: an AI Approach. *11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks, IEMECON 2023*. <https://doi.org/10.1109/IEMECON56962.2023.10092286>
- Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2. <https://doi.org/10.1016/j.mlwa.2020.100006>
- Benaddi, L., Ouaddi, C., Jakimi, A., & Ouchao, B. (2024). A Systematic Review of Chatbots: Classification, Development, and Their Impact on Tourism. *IEEE Access*, 12, 78799–78810. <https://doi.org/10.1109/ACCESS.2024.3408108>
- Buhalis, D. (2000). Marketing the competitive destination of the future. *Tourism Management*, 21(1), 97–116.
- Buhalis, D. and Amaranggana, A. (2015) 'Smart tourism destinations enhancing tourism experience through personalisation of services', in Xiang, Z. and Tussyadiah, I. (eds.) *Information and Communication Technologies in Tourism 2015*. Cham: Springer, pp. 377–389.
- Calvaresi, D., Ibrahim, A., Calbimonte, J. P., Schegg, R., Fragniere, E., & Schumacher, M. (2021). The evolution of chatbots in tourism: A systematic literature review. In Wörndl, W., Koo, C., & Stienmetz, J. L. (Eds.), *Information and Communication Technologies in Tourism 2021* (pp. 3–16). Springer. https://doi.org/10.1007/978-3-030-65785-7_1
- Casillo, M., Clarizia, F., D'Aniello, G., De Santo, M., Lombardi, M., & Santaniello, D. (2020). CHAT-Bot: A cultural heritage aware teller-bot for supporting touristic experiences. *Pattern Recognition Letters*, 131, 234-243.
- Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2023). Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review. <https://arxiv.org/abs/2310.14735>
- Chroma. (2024). ChromaDB: The AI-native open-source embedding database. Chroma. <https://www.trychroma.com>
- Cormack, G. V., Clarke, C. L. A., & Buettcher, S. (2009). Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 758–759.
- Cruz, M., Jardim, B. & Neto, M.d.C. Lisa: a touristic chatbot for Lisbon. *Inf Technol Tourism* (2025). <https://doi.org/10.1007/s40558-025-00339-x>
- Doan, N. N., Aki Härmä, Remzi Celebi, & Valeria Gottardo. (2024). A Hybrid Retrieval Approach for Advancing Retrieval-Augmented Generation Systems. *Association for Computational Linguistics*, 397–409. <https://aclanthology.org/2024.icnlp-1.41.pdf>
- Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A., Korenev, A., Hinsvark, A., Rao, A., Zhang, A., ... Zhao, Z. (2024). The Llama 3 Herd of Models. <https://arxiv.org/abs/2407.21783>
- Dyduch, W., & Brzozowska, A. (2025). Artificial intelligence in tourism management: Theoretical underpinnings, empirical tests and the SmartTourAI framework. *Central European Management Journal*. <https://doi.org/10.1108/CEMJ-07-2024-0227>
- Erak, O., Alabbasi, N., Alhusein, O., Lotfi, I., Hussein, A., Muhaidat, S., & Debbah, M. (2024). Leveraging Fine-Tuned Retrieval-Augmented Generation with Long-Context Support: For 3GPP Standards. <https://arxiv.org/pdf/2408.11775>
- Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2024). RAGAS: Automated Evaluation of Retrieval Augmented Generation. *Association for Computational Linguistics*, 150–158. <https://aclanthology.org/2024.eacl-demo.16>
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv:2312.10997*

- Gretzel, U., Sigala, M., Xiang, Z. and Koo, C. (2015) 'Smart tourism: Foundations and developments', *Electronic Markets*, 25(3), pp. 179–188.
- Jeong, C. (2023). A Study on the Implementation of Generative AI Services Using an Enterprise Data-Based LLM Application Architecture. *Advances in Artificial Intelligence and Machine Learning*, 03(04), 1588–1618. <https://doi.org/10.54364/AAIML.2023.1191>
- Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense Passage Retrieval for Open-Domain Question Answering. <https://arxiv.org/abs/2004.04906>
- Kim, H., So, K. K. F., Shin, S., & Li, J. (2024). Artificial Intelligence in Hospitality and Tourism: Insights From Industry Practices, Research Literature, and Expert Opinions. *Journal of Hospitality & Tourism Research*, 49(2), 366-385. <https://doi.org/10.1177/10963480241229235> (Original work published 2025)
- Kirtıl, İ. G., Çizel, B., Uzut, İ., & Uzun, S. (2024). Bridging the Gap: Fine-Tuning Artificial Intelligence (AI) Chatbots for Tourism. In *Proceedings of the Conference on Managing Tourism Across Continents (MTCO 2024)* (pp. 127–130). Istanbul, Turkey: Association of Turkish Tourism Academics.
- LangChain. (2024). LangSmith: Platform for evaluating and monitoring LLM applications. LangChain. <https://docs.langchain.com/langsmith/home>
- 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)*.
- Liberato, P., Alen, E., & Liberato, D. (2018). Smart tourism destination triggers consumer experience: the case of Porto. *European Journal of Management and Business Economics*, 27(1), 6–25. <https://doi.org/10.1108/EJMBE-11-2017-0051>
- Li, M., Yin, D., Qiu, H., & Bai, B. (2021). A systematic review of AI technology-based service encounters: Implications for hospitality and tourism operations. *International Journal of Hospitality Management*, 95. <https://doi.org/10.1016/j.ijhm.2021.102930>
- Ma, X., Sun, K., Pradeep, R., & Lin, J. (2021). A replication study of dense passage retriever (arXiv preprint arXiv:2104.05740). <https://arxiv.org/abs/2104.05740>
- Maryamah, M., Irfani, M. M., Tri Raharjo, E. B., Rahmi, N. A., Ghani, M., & Raharjana, I. K. (2024). Chatbots in Academia: A Retrieval-Augmented Generation Approach for Improved Efficient Information Access. 2024 16th International Conference on Knowledge and Smart Technology (KST), 259–264. <https://doi.org/10.1109/KST61284.2024.10499652>
- Meta. (2024, September 25). Llama 3.2: Revolutionizing edge AI and vision with open, customizable models. <https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edgemobile-devices/>.
- Muennighoff, N., Tazi, N., Magne, L., & Reimers, N. (2023). MTEB: Massive Text Embedding Benchmark. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 2014–2037). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.eacl-main.148>
- Pavel, F., & Romeiro, P. (2023). Challenges of Public Policies for Tourism in Medium-Sized Cities in Europe. The Cases of Lisbon and Porto (Portugal). *GOT - Journal of Geography and Spatial Planning*, 25, 129–150. <https://doi.org/10.17127/got/2023.25.007>
- Richardson, L. (2007). Beautiful Soup Documentation. Retrieved from <https://www.crummy.com/software/BeautifulSoup/>
- Risch, J., Möller, T., Gutsch, J., & Pietsch, M. (2021). Semantic Answer Similarity for Evaluating Question Answering Models. <https://arxiv.org/abs/2108.06130>
- Samancioglu, E., Kumlu, S., & Ozkul, E. (2024). Smart tourism destinations and sustainability: Evidence from the tourism industry. *Worldwide Hospitality and Tourism Themes*, 16(6), 680–693. <https://doi.org/10.1108/WHATT-07-2024-0167>
- Salemi, A., & Zamani, H. (2024). Evaluating Retrieval Quality in Retrieval-Augmented Generation. <https://arxiv.org/abs/2404.13781>
- Singh, A., Ehtesham, A., Mahmud, S., & Kim, J. H. (2024). Revolutionizing Mental Health Care through LangChain: A Journey with a Large Language Model. 2024 IEEE 14th Annual Computing and Communication Workshop and Conference, CCWC 2024. <https://doi.org/10.1109/CCWC60891.2024.10427865>
- Sperli G (2021) A cultural heritage framework using a deep learning based chatbot for supporting tourist journey. *Expert Syst Appl* 183. <https://doi.org/10.1016/j.eswa.2021.115277>
- Volchek, K., & Ivanov, S. H. (2024). ChatGPT as a travel itinerary planner. In K. Berezina, et al. (Eds.), *Information and Communication Technologies in Tourism 2024* (pp. 365–370). Springer, Cham. https://doi.org/10.1007/978-3-031-58839-6_38
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). BERTScore: Evaluating Text Generation with BERT. <https://arxiv.org/abs/1904.09675>
- Zhang, X., Zhang, Y., Long, D., Xie, W., Dai, Z., Tang, J., Lin, H., Yang, B., Xie, P., Huang, F., et al. (2024). mGTE: Generalized long-context text representation and reranking models for multilingual text retrieval. arXiv preprint arXiv:2407.19669. <https://arxiv.org/abs/2407.19669>