AI and Tourism: A Chatbot for Small and Medium Enterprises in the Tourism Sector

Stefano Bistarelli[1], Marco Cuccarini[1,2]

[1]University of Perugia, Italy
[2]University of Naples Federico II, Italy

Abstract

The following work proposes the development of an artificial intelligence system, based on multimodal chatbots, intended for the tourism and hotel industry. The main objective is to create a virtual assistant capable of performing the main concierge tasks, such as welcoming guests, handling check-in/check-out, and providing personalized information about the facility and the surrounding area. The system integrates OpenStreetMap geospatial data, structural information derived from architectural floor plans using computer vision techniques, and internal documentation through text mining methods. The extracted information will be organized into a knowledge base and made usable through natural interaction mediated by large-scale language models (LLMs). The approach is modular, inclusive (accessible to people with disabilities) and multilingual, and can be extended to different application contexts beyond the hotel industry, such as museums, shopping malls, and large public buildings. This approach will also be feasible when scaling from large to small tourism contexts, offering open-source solutions with improved computational efficiency.

Keywords

LLM, Tourism, Chatbot, Tourist Assistance

1. Introduction

In Italy, tourism accounts for approximately 5% of the national GDP and is a fundamental component of the economy [1]. One of the main limitations of the sector is its high demand for labor, due to the low degree of automation possible in many operational activities. Artificial Intelligence (AI) can provide effective support by automating certain tasks and making services more accessible and efficient [2].

This paper proposes the development of a chatbot capable of performing the main functions of a hotel concierge, including welcoming guests, handling check-in and check-out procedures, and providing information on schedules, regulations, and available services. Using metadata provided by OpenStreetMap (OSM), the chatbot will be able to offer information on local events and nearby points of interest [3]. It will also be possible to obtain detailed information on any object displayed on the map, thanks to the use of large language models (LLMs), which are highly effective in generating and understanding natural language [4, 5].

Similarly, the chatbot will need to be able to guide users within the facility. However, unlike OSM, there are no available metadata for indoor spaces. The objective will therefore be to design a system capable of receiving a floor plan of the building as input and returning a structured output similar to what is provided by services like Google Maps [6]. Since architectural floor plans are generally standardized, the use of computer vision models is particularly suitable [7, 8].

It will also use Knowledge Graph to integrate knowledge inside LLM and provide an explainable structure of the system that can motivate each decision made. The KG will be taken from external resources, like ontologies already defined, and with the process of Knowledge Graph mining.

The goal of this work is to develop a chatbot that operates autonomously once installed. The system must be capable of handling a wide variety of scenarios and inputs, taking into account exceptions, rules, and the specific characteristics of the facility in which it is deployed (e.g., building floor plan, geographic location, internal regulations).

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
stefano.bistarelli@unipg.it (S. Bistarelli); marco.cuccarini@unina.it (M. Cuccarini)
https://orcid.org/0000-0001-7411-9678 (S. Bistarelli); https://orcid.org/0009-0007-8943-8090 (M. Cuccarini)

The chatbot will be designed to support various types of input: text, audio, images, video, and touchscreen interaction. The ideal hardware will be a device equipped with a webcam, speakers, and a touchscreen (such as an interactive television).

The system will be multilingual and accessible, with support for the most widely spoken languages and integration with translation APIs for less common languages. It will also be compatible with various modes of interaction to facilitate access for people with disabilities.

2. State of the Art

This section reviews key theoretical concepts and recent advancements relevant to the development of intelligent systems in social context. It explores the integration of Large Language Models (LLMs) and Knowledge Graphs (KGs) to enhance knowledge representation and reasoning. Additionally, it covers the application of AI-powered chatbots in hotels, the use of geospatial metadata such as OpenStreetMap, and the automatic analysis of architectural floor plans. Together, these topics provide a comprehensive foundation for building sophisticated, context-aware tourist guidance and reservation systems.

2.1. Integration of Large Language Models (LLMs) and Knowledge Graphs (KGs)

The integration between Large Language Models (LLMs) and Knowledge Graphs (KGs) is an emerging research area that combines the generative and contextual understanding capabilities of LLMs with the semantic structure and informational robustness of KGs. This approach improves the accuracy, consistency, and explainability of generated responses, enabling advanced applications in question solving and automated reasoning [9]. By embedding structured knowledge from KGs into LLMs, systems can better handle complex queries, reduce ambiguity, and provide more reliable and context-aware answers. Furthermore, this integration facilitates knowledge updating and reasoning over dynamic information, making it particularly valuable in domains requiring up-to-date and precise knowledge representation.

2.2. Chatbots and Artificial Intelligence in the Hospitality Sector

The adoption of AI-based chatbots in the hotel industry has shown a significant impact on customer trust and satisfaction. A study by Nguyen et al. [10] highlighted that factors such as empathetic responses, anonymity, and personalization positively influence customer interactions and trust in hotel chatbot systems.

The implementation of real-world chatbots for hotel reservation was explored by Li et al. [11], who presented a conversational AI system capable of handling hotel search and reservation tasks by text messaging, demonstrating the effectiveness of such systems on a commercial scale.

2.3. Use of OpenStreetMap (OSM) Metadata and LLMs

The integration of geospatial metadata with large language models (LLMs) was investigated by Wang et al. [12], who employed ChatGPT as a mapping assistant to enrich maps with content derived from street-level photographs. This approach not only improved the accuracy of annotations in OpenStreetMap but also enhanced the semantic understanding of geographic features by leveraging the contextual knowledge embedded within the LLM. Such integration demonstrates the potential of combining AI-driven language models with spatial data to create more detailed and user-friendly mapping tools.

2.4. Recognition and Analysis of Architectural Floor Plans

The automatic detection of objects in architectural floor-plan images is a growing field. Hashmi et al. [13] proposed an approach based on Cascade Mask R-CNN to detect objects such as furniture, doors, and windows in floor plans, improving robustness through data augmentation techniques. Zeng et al.

[14] introduced a multitask neural network with room-boundary guided attention for deep floor plan recognition, achieving superior results compared to previous methods.

Finally, Yuan et al. [15] presented a large-scale dataset for the analysis of multi-unit floor plans, facilitating the development of accurate models for the recognition and labeling of rooms. These studies provide a solid foundation for the development of an advanced hotel chatbot, integrating natural interaction capabilities, the use of geospatial data, and spatial understanding of indoor environments.

3. Proposed work

The system consists of three integrated modules: the first extracts information from regulations using BERT for semantic retrieval and LLaMA for response generation; the second interprets urban maps and indoor floor plans, transforming them into semantic-spatial graphs to support spatial interaction; the third builds a Knowledge Graph specialized in monuments and points of interest, integrating data from various sources to improve the retrieval of tourist information.

3.1. Information Extraction System from Regulations

These modules work together to provide an intelligent and contextual conversational experience.

The first component aims to create a system that, given a regulation as input, can extract information according to the user's requests. This approach will utilize LLM systems capable of answering questions. To reduce risk, a system based on **BERT** [16], which is lighter and has fewer parameters, will be used to extract the relevant portion of the text through semantic matching algorithms. The extracted portion of the text will then be passed as input to the LLM **LLaMA** [17]. LLMs have demonstrated remarkable capabilities across a wide range of fields and are often used in question-answering systems to filter critical or potentially harmful content - a necessary configuration when dealing with the public [18].

BERT has been used in various contexts for question-answering extraction systems, and many studies explore methods to improve performance using zero-shot learning [19], without requiring any training data [20], in order to preserve the generality of the system. This is one of the main points of this work, the training of the model can improve the performance of the provided questions, but can be applied only in one case. Once the model is trained, it loses generalization, this requires technical experience to train it, and this possibility can be an excessive cost for small and small business.

This approach, typical of **Retrieval-Augmented Generation (RAG)** systems [21], reduces the computation time, provides more accurate information, and mitigates the phenomenon of hallucinations.

3.2. Interpretation of City Maps and Indoor Structures

To support intelligent and contextual interaction between the user and the urban or architectural environment, the system must be equipped with the ability to understand and represent the spatial and semantic relationships between contextual elements. This section describes the approach adopted to interpret city maps and indoor structures, transforming this information into structured representations useful for conversational interaction [22].

The system must be capable of recognizing and interpreting **spatial** and **semantic relationships** between different elements on the city map, to support a more natural and effective interaction with the user. Specifically, it will be necessary to identify spatial relationships, for example, "in front of," "next to," "connected by a road" [23], as well as semantic relationships, such as "located on the same street," "within the same area" [24].

For the interpretation of indoor structures, a specialized **object detection model** will be developed to analyze architectural floor plans. This model will be responsible for recognizing and classifying **room types** (e.g., bedrooms, bathrooms, kitchens, halls, etc.) [25], identifying **relevant objects** within each room, and extracting and organizing this information into consistent **meta-data** [26].

Metadata extracted from urban maps and indoor floor plans will be used to construct **structured graphs**, where **nodes** represent spaces, rooms or objects, and **edges** describe spatial relationships (for example, proximity, connection) or functional ones (for example, use, belonging to a specific area) [27].

These graphs will serve as the foundation for the power of a **large language model (LLM)**, which will be able to generate dynamic conversational responses, guide the user inside buildings or among points of interest in the city, and suggest events or potentially relevant locations based on the user's position or preferences [28].

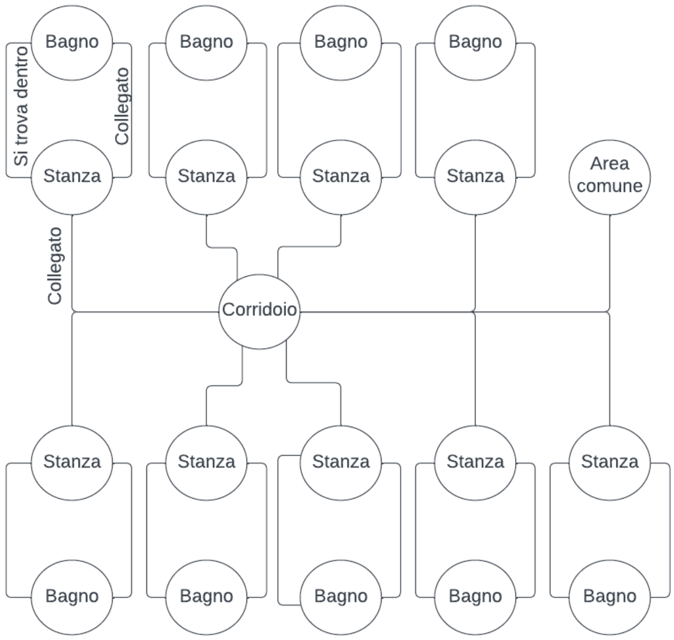

Figure 1: An Italian knowledge graph representing the inside of a structure

3.3. Creation of a Knowledge Graph for Monuments and Points of Interest

The objective of this section is to describe the construction of a Knowledge Graph (KG) designed to represent and organize relevant information about tourist points of interest in a structured and interconnected manner. This structured representation aims to enhance the effectiveness of information retrieval within an intelligent response system, providing more accurate and context-aware guidance to users. The KG will be developed by integrating data from multiple sources, including Wikipedia, public datasets, and other relevant information repositories, alongside already structured data from existing information systems. To extract pertinent information from unstructured texts, a comprehensive text mining process will be employed. This process involves identifying key entities, relationships, and metadata, enabling the transformation of raw textual data into structured knowledge. The adoption of a knowledge graph offers two significant advantages. First, it allows the incorporation of new knowledge contributed by domain experts, enriching the system with up-to-date and specialized information. Second, it leverages existing knowledge embedded within large language models (LLMs) by organizing it into a formal structure. This organization facilitates better inference capabilities, enabling the system to deduce new concepts and relationships more effectively. Overall, the integration of diverse data sources and the application of advanced text mining techniques will result in a robust and dynamic KG that supports intelligent and contextually relevant tourist guidance.

3.4. LLM and KG Integration

Various methods will be explored to integrate the information collected in Sections 3.2 and 3.3 into the language model. The extracted data will be incorporated using different techniques, including graph embedding [29], direct injection of triples into the prompt [30], semantic retrieval based on symbolic knowledge [31], and hybrid neuro-symbolic models [32]. The goal is to expand the contextual knowledge accessible to the LLM during response generation, without excessively increasing computational overhead.

This approach will enable the model to access precise, structured, and up-to-date information, helping to improve the precision, consistency, and semantic relevance of its responses [9]. At the same time, integrating a Knowledge Graph can provide a framework for explainability, allowing traceability of the system's decision making, a crucial factor in contexts where transparency and reliability are required [33]. This mechanism not only improves the models' capabilities, but also fosters greater user trust, as users can better understand the basis on which responses are generated. The user will be able to explore the sources and underlying semantic relationships, actively participating in the system's decision-making process.

In the future, these techniques may be further extended with modules for continuous updates and source verification, making the system more resilient to the dynamics of information content and potentially more compliant with regulatory requirements such as algorithmic transparency and AI trustworthiness mandated by European legislation [34].

4. Conclusions

This approach is specifically designed to minimize computational requirements, making it accessible and practical to small and medium enterprises (SMEs) in the tourism sector. Using lightweight models and efficient algorithms, the system ensures that essential functionalities such as natural language understanding, semantic search, and interactive responses can be executed on affordable hardware with limited resources. This reduces both operational costs and technical barriers, allowing smaller businesses, often with constrained budgets and infrastructure, to implement advanced AI-driven chatbots without the need for expensive servers or cloud computing services. Ultimately, this approach democratizes access to intelligent customer assistance, fostering greater innovation and competitiveness across the tourism industry regardless of the size of the company.

The proposed project aims to make a significant contribution to the tourism sector, with potential extensions to other application domains such as shopping malls, banking institutions, museums, trade fairs, and large public facilities. The system is designed to be modular and flexible, allowing for partial and scalable implementations based on specific needs and available resources.

The integration of advanced artificial intelligence technologies, natural language processing (NLP), and computer vision represents a turning point for the intelligent automation of hospitality and navigation services. Thanks to the semantic-spatial interpretation of environments and personalized conversational interaction, the system can enhance the user experience, making it more engaging, efficient, and accessible.

In the long term, this approach can foster a profound transformation in the way users interact with complex physical spaces, promoting more informed, sustainable, and inclusive use. Moreover, the system's adaptive nature allows for continuous evolution in response to technological advancements and new social demands, positioning it as a strategic platform for innovation in intelligent guidance and assistance services.

5. Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and DeepL to perform grammar and spelling checks, translation, and to paraphrase or reword parts of the text. After using these tools, the authors thoroughly reviewed and edited the content as needed, and take full responsibility for the accuracy and integrity of the publication's content.

References

[1] Statista, Tourism total contribution to gdp in italy 2014-2028, 2023. URL: https://www.statista.com/statistics/628849/tourism-total-contribution-to-gdp-italy-share/, accessed: 2025-06-15.

[2] Deloitte, Ai in the travel and tourism industry, https://www.deloitte.com/uk/en/Industries/consumer/blogs/embracing-the-future.html, 2021. Disponibile su: https://www.deloitte.com/uk/en/Industries/consumer/blogs/embracing-the-future.html.

[3] M. Haklay, How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets, Environment and Planning B: Planning and Design 37 (2010) 682–703. URL: https://journals.sagepub.com/doi/10.1068/b35097. doi:10.1068/b35097.

[4] T. B. Brown, B. Mann, N. Ryder, et al., Language models are few-shot learners, in: Advances in Neural Information Processing Systems (NeurIPS), 2020. URL: https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

[5] R. Bommasani, et al., On the opportunities and risks of foundation models, arXiv preprint (2021). URL: https://arxiv.org/abs/2108.07258. arXiv:arXiv:2108.07258.

[6] A. R. Zamora, et al., A survey of indoor navigation systems for smartphones, IEEE Access 9 (2021) 99146–99171. URL: https://www.mdpi.com/1424-8220/24/21/6876. doi:10.1109/ACCESS.2021.3096560.

[7] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (2018) 296–312. URL: https://arxiv.org/abs/1703.06870. doi:10.1109/TPAMI.2017.2787607.

[8] J. Chen, Y. Li, J. Wu, Plan2vec: Unsupervised representation learning by latent plans, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. URL: https://arxiv.org/abs/2005.03648.

[9] Z. Ji, Y. Cheng, X. Liu, et al., A survey on knowledge-enhanced text generation, ACM Computing Surveys (CSUR) 55 (2022) 1–38.

[10] V. T. Nguyen, L. T. Phong, N. T. K. Chi, The impact of ai chatbots on customer trust: an empirical investigation in the hotel industry, Consumer Behavior in Tourism and Hospitality 18 (2023) 293–305.

[11] B. Li, N. Jiang, J. Sham, H. Shi, H. Fazal, Real-world conversational ai for hotel bookings, in: 2019 Second International Conference on Artificial Intelligence for Industries (AI4I), IEEE, 2019, pp. 58–62.

[12] S. Wang, T. Hu, H. Xiao, Y. Li, C. Zhang, H. Ning, R. Zhu, Z. Li, X. Ye, Gpt, large language models (llms) and generative artificial intelligence (gai) models in geospatial science: a systematic review, International Journal of Digital Earth 17 (2024) 2353122.

[13] M. Hashmi, S. Srivastava, V. Kumar, Object detection in architectural floor plans using cascade mask r-cnn, Applied Sciences 11 (2021) 11174.

[14] Z. Zeng, C.-W. Fu, P.-A. Heng, Deep floor plan recognition using multi-task learning with room-boundary-guided attention, arXiv preprint arXiv:1908.11025 (2019).

[15] Z. Yuan, X. Chen, H. Zheng, et al., A large-scale dataset for multi-unit architectural floor plan analysis, Automation in Construction 155 (2023) 105032.

[16] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[17] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., Llama: Open and efficient foundation language models, arXiv preprint arXiv:2302.13971 (2023).

[18] L. Draetta, C. Ferrando, M. Cuccarini, L. James, V. Patti, Reclaim project: Exploring italian

slurs reappropriation with large language models, in: CLiC-it, volume 3878 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024.
[19] S. Bistarelli, M. Cuccarini, Bert-based questions answering on close domains: Preliminary report, in: CILC, volume 3733 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024.
[20] S. Bistarelli, M. Cuccarini, Feature selection on contextual embedding pushing the sparseness, in: AI*IA, volume 15450 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 37–49.
[21] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Kulkarni, M. O. Ju, V. Stoyanov, S. Riedel, Retrieval-augmented generation for knowledge-intensive nlp tasks, in: Advances in Neural Information Processing Systems, volume 33, 2020, pp. 9459–9474.
[22] J. Bateman, S. Farrar, Spatial Representation and Reasoning for Geographic Information Systems, Springer, 2010.
[23] B. Kuipers, Modeling spatial knowledge, in: Proceedings of the 1978 Workshop on Theoretical Issues in Natural Language Processing, 1978, pp. 1–10.
[24] Y. Gao, C. Lv, X. Wang, Y. Liu, Constructing a semantic map for mobile robots using human semantics, IEEE Transactions on Automation Science and Engineering 14 (2017) 1193–1202.
[25] C. Liu, Y. Xu, L. Xu, Q. Huang, C.-W. Fu, Deepfloorplan: Structural understanding of floorplans using deep learning, Pattern Recognition 107 (2021) 107528.
[26] L. Zhao, H. Cai, C. Zhang, Data-driven approach for room type classification from floor plans, in: International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2019, pp. 523–530.
[27] M. Do, A. G. Schwing, Graph-based modeling of building layouts for scene synthesis, in: Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 10687–10696.
[28] P. Liu, et al., A survey of large language models, arXiv preprint arXiv:2303.18223 (2023).
[29] W. L. Hamilton, R. Ying, J. Leskovec, Representation learning on graphs: Methods and applications, IEEE Data Engineering Bulletin 40 (2017) 52–74.
[30] Y. Liu, H. Zhou, L. Li, W. X. Zhao, J.-R. Wen, Promptkg: Prompt-based knowledge graph construction, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022.
[31] M. Yasunaga, X. Ren, J. Leskovec, Linkbert: Pretraining language models with document links, arXiv preprint arXiv:2203.15827 (2022).
[32] T. Wang, X. Wang, R. Zhang, et al., Deep reasoning with knowledge graph for social relationship understanding, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
[33] L. Gilpin, D. Bau, B. Yuan, et al., Explaining explanations: An overview of interpretability of machine learning, in: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), IEEE, 2018, pp. 80–89.
[34] E. Commission, Proposal for a regulation on a european approach for artificial intelligence (artificial intelligence act), 2021. Available at: https://artificialintelligenceact.eu/.