Automating the Classification of Smart Tourism Tools with LLMs

A Taxonomy-Driven Approach

Diogo Cosme[1]
ISTAR, Instituto Universitário de Lisboa (ISCTE-IUL), Portugal

Fernando Batista
INESC-ID, Lisbon, Portugal
Instituto Universitário de Lisboa (ISCTE-IUL), Portugal

António Galvão
CENSE, School of Science and Technology, NOVA University Lisbon, Caparica, Portugal

Fernando Brito e Abreu
ISTAR, Instituto Universitário de Lisboa (ISCTE-IUL), Portugal

Abstract

Smart Tourism Tools (STTs) are digital solutions that support tourism stakeholders in designing, managing, promoting, and enhancing tourism experiences. As the number and diversity of these tools increase, consistent classification becomes essential for organizing them in observatory-like repositories and making them easier to search, compare, and reuse. However, previous attempts to classify STTs using Large Language Models (LLMs) showed that broad taxonomy descriptions can lead to subjective and inconsistent interpretations. This paper proposes an improved STT classification framework that operationalizes the original taxonomy as a set of binary classification questions. Each taxonomy category is represented by a yes/no question, making the classification process more explicit, traceable, and easier to analyze. The taxonomy is also refined by renaming selected action domains and categories and removing one low-utility category. An exploratory evaluation was conducted using 100 randomly selected English-language STT descriptions and three state-of-the-art LLMs. A randomly selected subset of 20 examples was validated by a human expert. The results suggest that the proposed framework can support LLM-based classification with substantial alignment to expert judgment, while also revealing remaining challenges related to over-classification and ambiguous category boundaries. The paper contributes a more transparent and reproducible framework for classifying STTs and provides initial evidence of its feasibility for semi-automatic classification in smart tourism observatories.

2012 ACM Subject Classification Computing methodologies → Natural language processing

Keywords and phrases Language Models (LMs), Multi-Label Learning, Domain-Specific NLP, Taxonomy Alignment, Smart Tourism Systems

Digital Object Identifier 10.4230/OASIcs.SLATE.2026.18

Category Short Paper

Supplementary Material *Text (Prompts)*: `https://doi.org/10.5281/zenodo.19741250`

Funding This work was partly supported by project 2024.07579.IACDC/2024, funded by FCT – Fundação para a Ciência e a Tecnologia, I.P., under PRR – Plano de Recuperação e Resiliência funding, investment RE-C05-i08 – Ciência Mais Digital and, partially supported by national funds through FCT, ISTAR-Iscte Pluriannual Project UID/04466/2025, and INESC-ID projects UID/50021/2025 and UID/PRR/50021/2025.

[1] Corresponding author

15th Symposium on Languages, Applications and Technologies (SLATE 2026).
Editors: Fernando Batista, Eugénio Ribeiro, Ricardo Ribeiro, and André L. Santos; Article No. 18; pp. 18:1–18:11
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Smart tourism has become an increasingly important area of research and practice, driven by the growing use of digital technologies to support destinations, tourism providers, and tourists. Digital solutions can improve the design, management, promotion, and consumption of tourism experiences, while also contributing to more sustainable and data-informed tourism development. However, as the number and diversity of such solutions increase, it becomes more difficult to organize, compare, and reuse information about them in a consistent way.

The concept of Smart Tourism Tools (STTs) was introduced to describe digital solutions that support tourism stakeholders and contribute to smart tourism development. For example, STTs can include booking systems for tourism operators or enable digital experiences within a tourism activity. To support the identification and organization of these tools, a taxonomy of STTs was proposed to classify solutions according to their role in the tourism ecosystem, including their contribution to tourism experiences, marketing, and the design, management, and operation of tourism products [9].

This classification task is particularly relevant to the EUROSTIT observatory[2], which aims to collect and organize information about STTs from different sources [9]. Since many of these solutions are described in heterogeneous formats, such as catalogues, reports, and promotional texts, manual classification can be time-consuming and difficult to scale. Large Language Models (LLMs) offer a promising way to support this process, since they can interpret natural-language descriptions and assign taxonomy categories with limited task-specific training.

However, previous attempts to classify STTs with LLMs revealed important limitations [6]. The original taxonomy was useful as a conceptual framework, but its category descriptions left room for interpretation. As a result, both human and machine annotators could apply the taxonomy differently, especially when solution descriptions were short, ambiguous, or related to more than one category. This reduced the consistency and explainability of the classification process.

To address this limitation, this paper proposes an improved STT classification framework. The original taxonomy is refined by renaming some action domains and categories, removing one low-utility category, and operationalizing each remaining category through a binary classification question. Instead of asking an annotator or LLM to directly infer the most appropriate category from a broad description, the proposed framework asks whether each category-specific condition is satisfied. This questionnaire-based approach aims to reduce ambiguity, make classification decisions more traceable, and support systematic error analysis.

The paper makes three main contributions. First, it refines the original STT taxonomy to improve clarity and usability. Second, it proposes a questionnaire-based classification framework that transforms taxonomy categories into binary decision questions. Third, it presents an exploratory evaluation of the framework using three state-of-the-art LLMs and expert validation on a randomly selected subset of classified examples.

The remainder of the paper is organized as follows. Section 2 reviews related work on smart tourism, STT classification, taxonomy-based classification, and LLM-based text classification. Section 3 presents the improved classification framework and the revised taxonomy. Section 4 describes the exploratory evaluation with LLMs. Section 5 discusses the main findings and limitations. Finally, Section 6 concludes the paper and outlines directions for future work.

[2] European Observatory of Smart Tourism Initiatives and Tools

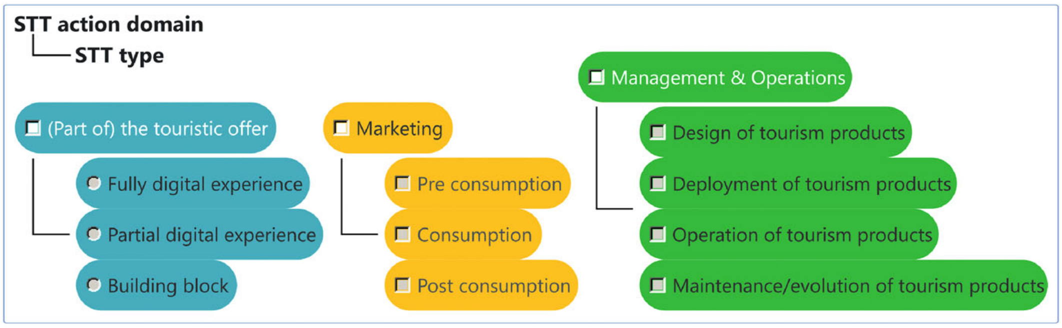

Figure 1 Original taxonomy of Smart Tourism Tools (STTs) proposed in [9].

2 Related Work

Smart tourism has emerged from the increasing integration of information and communication technologies, data-driven services, and interconnected stakeholder ecosystems in tourism. Foundational work on smart tourism emphasizes the role of digital infrastructures, big data, smart technologies, and value co-creation in transforming tourism experiences and destination management [10]. Related work on smart tourism destinations also highlights the importance of technological platforms, real-time information exchange, and personalization for improving tourism services and visitor experiences [2]. More recently, the concept of Smart Tourism Tools (STTs) has been proposed to capture digital solutions that support tourism stakeholders while contributing to sustainability-oriented smart tourism development [9].

The STTs taxonomy [9] was created to address the lack of comprehensive and structured approaches for categorizing smart tourism solutions. The initial version of the taxonomy was developed from the opinions of more than 300 tourism experts collected through a survey. It organized STTs around three main action domains: (i) *(Part of) the touristic offer*, (ii) *Marketing*, and (iii) *Management & Operations*. Figure 1 illustrates the original taxonomy and its respective action domains.

In addition to its conceptual contribution, the STTs taxonomy was designed to support the EUROSTIT observatory, which aims to function as a centralized repository of smart tourism solutions and to improve the visibility of these solutions among Destination Management Organizations, tourism providers, and other tourism stakeholders [9]. Classifying STTs according to this taxonomy provides concrete benefits to these stakeholders. For instance, it empowers Small and Medium-sized Enterprises (SMEs) to bridge the digital divide by allowing them to filter out irrelevant software; a small hotel can easily locate a *Consumption tool*, such as an in-room upselling app, without sifting through back-office logistics software. Similarly, it aids Destination Management Organizations (DMOs) in making strategic, low-risk investments; a DMO tackling overcrowding can use the *Operation* category to quickly discover GDPR-compliant crowd-monitoring sensors for heritage sites. Finally, for the observatory itself, a LLM-based classification framework enables automated market intelligence, allowing the platform to continuously ingest and categorize new digital solutions as they emerge, ensuring the repository remains an up-to-date compass for European smart tourism.

EUROSTIT content is extracted from several sources, including PDF catalogs created by tourism-related organizations [8,16]. However, such catalogs are often difficult to search, compare, and reuse directly, especially when their content is not structured or easily indexable. As the number of collected solutions grows, the value of the observatory increasingly depends on the ability to organize STTs consistently according to a meaningful taxonomy. If organized, search is facilitated, namely when using the chatbot our team is developing for this observatory [1].

From a Machine Learning perspective, classifying STTs is a multi-label text classification problem because a single solution may belong to more than one taxonomy category. Multi-label classification has been widely studied, and previous surveys emphasize the specific challenges of modeling label dependencies, evaluating partially correct predictions, and handling imbalanced or ambiguous labels [20]. The STTs classification problem also has a hierarchical component, since labels are organized under broader action domains. Hierarchical classification is known to introduce additional challenges because prediction errors may occur both at the level of broad parent categories and at the level of more specific child categories [17]. These characteristics make the STT classification more complex than the standard single-label document classification.

Traditional text-classification approaches often rely on supervised learning over annotated datasets, including deep-learning methods that learn representations from text and associate them with predefined categories [15]. However, in emerging or highly specialized domains such as smart tourism, large labeled datasets are rarely available. This makes purely supervised approaches difficult to apply in the early stages of taxonomy development. The use of LLMs offers an alternative because these models can perform classification through prompts without requiring a large domain-specific training set [7, 13, 14]. Recent work has shown that LLMs can be effective for text classification tasks, especially when the prompts are carefully designed and when the task benefits from language understanding and reasoning [4, 12, 19]. At the same time, these studies also show that LLM-based classification remains sensitive to task formulation, prompt design, model choice, and the availability of examples.

The use of LLMs in tourism is also gaining attention. Recent studies discuss the role of generative AI, GPT-like systems, and LLMs in tourism and hospitality, including applications related to personalization, service support, content generation, and decision-making [11]. Other tourism-related AI systems, such as chatbots and cultural-heritage assistants, demonstrate the potential of conversational and language-based technologies for improving tourist experiences [3, 18]. Nevertheless, the specific problem of using LLMs to classify smart tourism solutions according to a domain taxonomy remains underexplored.

The first attempt to automate STT classification using LLMs was presented in [6]. In that approach, the taxonomy description and a small number of examples were provided as context, and the LLM was asked to infer the most appropriate classification from each solution description. Although this demonstrated the potential of LLMs for supporting the observatory, several limitations were identified. The task remained highly subjective, the interpretation of category boundaries was not always consistent, and disagreements occurred even among human annotators. As a result, it was difficult to obtain stable classification quality and acceptable inter-annotator agreement.

These limitations suggest that the main challenge is not only the choice of model, but also the way the taxonomy is operationalized for classification. A taxonomy description alone may leave too much room for interpretation, especially when categories overlap or when solution descriptions are short and marketing-oriented. Therefore, the present work complements the STTs taxonomy with a questionnaire-based classification framework. Instead of asking the classifier to directly select categories, each category is represented by a binary question. This design aims to reduce ambiguity, improve traceability, and make classification disagreements easier to analyze. The next section describes the taxonomy refinement process and the resulting questionnaire-based framework.

3 Revised STT Taxonomy and Classification Protocol

3.1 Conceptual Taxonomy

The revised STT taxonomy is a conceptual framework for organizing Smart Tourism Tools according to their primary functional role. The taxonomy is composed of three action domains, each containing a set of categories representing distinct classes of functionality related to tourism. In addition to introducing classification questions, the taxonomy was refined to improve clarity and usability. Some action domains and categories were renamed, and one category was removed.

- The action domains previously called *(Part of) the touristic offer* and *Management & Operations* were renamed, respectively, *Tourism Fruition* and *Design, Management & Operations of Tourism Products* to better reflect the nature of the solutions that fall under these domains.
- The *Building Block* category was removed from the taxonomy because it corresponded to a specific type of solution that was uncommon and, therefore, not useful for classification.
- The categories *Design of tourism products, Deployment of tourism products, Operation of tourism products,* and *Maintenance/Evolution of tourism products*, were renamed to *Design, Deployment, Operation,* and *Maintenance/Evolution*, respectively, to make them more concise and easier to understand.

The resulting taxonomy contains three action domains: *Design, Management & Operations of Tourism Products*, *Marketing*, and *Tourism Fruition.* This taxonomy exists independently of the annotation procedure and defines the conceptual constructs used to describe the space of STTs.

3.2 Operationalization as a Classification Protocol

The main objective of the revised framework is to reduce subjectivity and improve the consistency of STT classification. To support more consistent and reproducible annotation, the original taxonomy was retained as the conceptual classification structure, while a questionnaire-based protocol was introduced to operationalize category assignment.

In this protocol, for each taxonomy category, a corresponding binary classification question was defined to guide annotation decisions. When the answer to a decision question is “yes”, the corresponding conceptual category is assigned to the solution. The questions were derived from the descriptions of the original taxonomy categories and designed to make the operational criteria as explicit as possible. This approach transforms the classification task from an open-ended interpretation of category meanings into a series of focused operational decisions. As a result, the framework is intended to support more transparent, reproducible, and explainable classification.

The use of binary questions aims to minimize ambiguity and reduce the range of possible interpretations during classification. In the previous framework, it was often difficult to determine why a given solution had been assigned to a specific category, since the annotator, whether human or machine, could rely on implicit assumptions or different interpretations of the conceptual taxonomy. The questionnaire-based protocol mitigates this issue by making operational criteria explicit. Each category is linked to a specific question, which makes it easier to understand the rationale behind a classification decision. This also facilitates error analysis: when a category is assigned incorrectly, the corresponding question can be examined to determine whether the issue lies in the wording of the question, the interpretation of the solution description, or the boundaries of the conceptual category itself. Therefore, the protocol not only supports classification but also provides a mechanism for refining the taxonomy over time.

3.3 Classification Questions

The resulting taxonomy contains three action domains and their corresponding categories. To support a consistent application of the taxonomy, each category is paired with a binary classification question used during annotation. The questions shown in Table 1 do not define the taxonomy itself; rather, they operationalize the taxonomy categories into observable decision criteria that can be consistently applied by human annotators or LLMs. For instance, a mobile app that provides AR-guided museum visits and location-aware recommendations would likely satisfy the *Partial Digital Experience* question because it enhances a real-world tourist experience through digital augmentation.

Table 1 Classification Questions for Smart Tourism Tools Taxonomy.

Action Domain	Categories & Questions	
	Category	**Question**
Design, Management & Operations of Tourism Products	Design	Does the STT help Tourism Providers design and build their offerings?
	Deployment	Does the STT help Tourism Providers deploy their offerings?
	Operation	Does the STT help Tourism Providers manage and operate their offerings?
	Maintenance/ Evolution	Does the STT help Tourism Providers collect and/or analyze data to improve their tourism offerings?
Marketing	Pre Consumption	Does the STT help Tourism Providers promote their services before the tourism experience?
	Consumption	Does the STT help Tourism Providers promote additional offers or services during the tourism experience?
	Post Consumption	Does the STT allow tourists to share their experiences, provide feedback, or act as influencers?
Tourism Fruition	Fully Digital Experience	Is the STT a fully immersive experience within a virtual environment?
	Partial Digital Experience	Is the STT a technological enabler and/or enhancer of the tourist's experience in the real world?

4 Exploratory Evaluation of LLM-based Classification

To assess whether the proposed questionnaire-based framework can be applied automatically by LLMs, we conducted an exploratory evaluation with three state-of-the-art models: *anthropic/claude-sonnet-4.6*, *openai/gpt-5.4*, and *google/gemini-3.1-pro-preview*. The goal of this evaluation was not to establish a definitive benchmark, since no large-scale human-annotated ground truth dataset is currently available, but rather to examine whether the models' classifications were sufficiently aligned with expert judgment to support further development of the approach.

The evaluation dataset consisted of 100 Smart Tourism solution descriptions randomly selected from the available collection. All descriptions used in this experiment were written in English, so no translation step was required. Each model was asked to classify every solution according to the questionnaire-based taxonomy presented in Section 3. Since the taxonomy is multi-label, each solution could be assigned to more than one category. The output format also allowed the models to return no category when the description did not correspond to an STT or when no taxonomy category was applied. The system prompt and classification prompt templates used in the evaluation are available in [5].

From the 100 classified descriptions, a subset of 20 examples was randomly selected for expert validation. For each of these 20 examples, the outputs of the three models were independently assessed by a human expert, resulting in 60 evaluated model outputs. The expert assigned a suitability score using a five-point Likert-type scale: 0, 0.25, 0.5, 0.75, and 1.0. A score of 1.0 indicates that the classification was considered fully suitable, while a score of 0 indicates that the classification was incorrect or unsuitable. Intermediate values were used when the output was partially correct, for example, when the model assigned some relevant categories but also omitted an important one or included a questionable category.

This scoring approach was adopted because the task is not always reducible to a single objectively correct answer. Some solution descriptions are short, ambiguous, or written in broad commercial language, which can make more than one classification plausible. For example, a communication tool that manages customer requests may be interpreted as supporting operations if it is used during service delivery, but it may also have a marketing function if it is used to engage potential visitors before consumption. In such cases, a strict binary correct/incorrect metric would not capture the reasonableness of the model's interpretation. The expert score, therefore, evaluates the suitability of the overall classification rather than only exact label matching.

Figure 2 presents the distribution of the expert scores across the 20 evaluated examples. The best average score was obtained by *google/gemini-3.1-pro-preview* (0.8625), followed by *anthropic/claude-sonnet-4.6* (0.8250) and *openai/gpt-5.4* (0.7250). The median score was 1.0 for all three models, indicating that more than half of the evaluated outputs for each model were considered fully suitable by the expert. However, the standard deviations and the number of zero-score cases show that the models still produced important errors in some cases. Overall, the results suggest that the models were often able to apply the questionnaire-based framework in a way that aligned with expert judgment. Across the 60 evaluated outputs, the overall mean score was 0.8042, and the overall median was 1.0. Moreover, 48 out of the 60 evaluated outputs received a score of at least 0.75, suggesting that most classifications were considered suitable or mostly suitable.

The evaluation also highlights an important boundary condition: the models must not only assign appropriate taxonomy categories but also recognize cases where no category should be assigned. Three examples in the validation subset were judged as not requiring any taxonomy category and were correctly left unclassified by all three models. This suggests that the models can, in some cases, avoid forcing a classification when the description does not correspond to an STT or does not fit the taxonomy. However, this behavior was not consistent. In the case of *Orbis Booking* (id=11), all three models assigned one or more categories, but the expert rated all three outputs as unsuitable. This indicates that false-positive classifications remain a relevant source of error.

The most common issues observed in the validation subset were over-classification and boundary ambiguity. Over-classification occurred when a model assigned additional categories that were only weakly supported by the description. Boundary ambiguity occurred when a

id	name	anthropic claude-sonnet-4.6	openai gpt-5.4	google gemini-3.1-pro-preview
1	ITSVIZIT	0,75	1	0,75
2	Tracking and Digital Antigen Test	0,75	1	1
3	Qred Nomenclator Smart Posters	0,75	0,5	0,75
4	Accessible Pdf Adaptation	1	0	1
5	Transvia Business Chat	0,5	1	1
6	Q for TOURISM QUALITY	1	1	1
7	Fintech Lab	1	1	1
8	Drafting and Implementation of Covid-19 Contingency Plans	1	1	1
9	2iXR PLATFORM	0,5	0,75	1
10	PLAYVISIT	1	1	1
11	Orbis Booking	0	0	0
12	Seeketing - Mobility and User Behavior	0,25	0,5	0,75
13	Tourist Traceability Sensor	1	1	1
14	Predictive Maintenance of Tourist Assets	1	1	1
15	FS OSINT	1	0	0
16	BEAMBASSADOR	1	0,25	1
17	PHYGI Sensor	1	1	1
18	Smart Occupancy	1	0,75	1
19	APP&TOWN COMPAGNON	1	1	1
20	Chatbot Taro	1	0,75	1
	Avg Score	0,83	0,73	**0,86**

Figure 2 Expert evaluation scores for the LLM-based classification of 20 randomly selected solution descriptions.

solution could plausibly be interpreted as supporting more than one action domain, especially between operational support, marketing, and partial digital experience. These cases confirm the usefulness of the questionnaire-based framework, since each disagreement can be traced back to specific category questions and can therefore inform future refinements of the prompts or of the category definitions.

Given the small size of the expert-validated subset, these results should be interpreted as preliminary. They provide initial evidence that LLMs can apply the proposed framework with substantial alignment to expert judgment, but they do not yet establish generalizable model performance. A larger validation study, ideally involving multiple human annotators and label-level agreement analysis, is required to obtain more robust estimates of classification quality.

5 Discussion

The results suggest that the proposed questionnaire-based framework can make the classification of Smart Tourism Tools more transparent and consistent. By replacing broad category descriptions with binary questions, the framework reduces the space for interpretation and makes each classification decision easier to trace. This is particularly relevant for LLM-based classification, since the model output can be analyzed in relation to specific questions rather than only to final category labels.

Our exploratory evaluation indicates that the tested LLMs were often able to apply the framework in ways that aligned with expert judgment. Most evaluated outputs received high suitability scores, and all three models correctly avoided assigning categories in some cases where no classification was appropriate. This suggests that LLMs can support the semi-automatic organization of STTs in observatory-like repositories.

However, the results also show that classification remains challenging. Some errors were caused by over-classification, where models assigned categories that were only weakly supported by the solution description. Other errors were related to ambiguous boundaries between categories, especially when a tool could plausibly support marketing, operational management, or the tourist experience. These cases show that the framework improves explainability, but does not eliminate the need for expert validation and further refinement. The evaluation should therefore be interpreted as preliminary. The expert-validated sample was limited to 20 randomly selected examples, and only one expert was involved in the validation process. As a result, the findings provide initial evidence of feasibility, but they do not yet establish generalizable performance for the framework or for any specific LLM.

5.1 Threats to Validity and Future Work

Although preliminary results suggest that the questionnaire-based framework improves the transparency of the STT classification, several threats to validity must be acknowledged. First, the exploratory evaluation relied on a limited sample size of 20 solution descriptions. This restricts the statistical generalizability of the findings. Second, a single human expert performed the ground truth validation using a subjective suitability score. Although this metric was useful for assessing partial correctness in an exploratory setting, it introduces potential single-annotator bias and precludes the use of standard machine learning metrics (e.g., F1-score or Hamming Loss).

Furthermore, applying independent binary questions to a hierarchical multi-label taxonomy introduces structural risks, such as potential logical inconsistencies between parent domains and child categories, as well as scalability concerns regarding API latency and token consumption when deployed in large-scale observatory repositories.

To mitigate these limitations, future work will involve a comprehensive empirical validation study. We are planning a synchronous offline classification session with a panel of 30 independent tourism domain experts. To ensure statistical robustness, this study will utilize a targeted overlap design: a dataset of 60 STTs will be structured so that each solution is independently evaluated by exactly five experts. This methodology will yield 300 independent human classifications, allowing for the calculation of Inter-Annotator Agreement (IAA) metrics (e.g., Fleiss' Kappa) to quantitatively measure the framework's effectiveness in reducing human ambiguity.

Establishing this multi-annotator "gold standard" will also transition the evaluation from subjective suitability scores to standard multi-label classification metrics. Finally, to address scalability and structural consistency, future iterations of the framework will explore hierarchical gating mechanisms (e.g., routing sub-category queries only if a parent domain is classified positively) and structured batch prompting.

6 Conclusion and Future Work

This paper presented an improved classification framework for Smart Tourism Tools. The original taxonomy was refined by renaming action domains and categories, removing one low-utility category, and operationalizing the remaining categories through binary classification questions. This questionnaire-based approach aims to reduce ambiguity, improve traceability, and support more consistent classification by humans and LLMs.

An exploratory evaluation was conducted using 100 randomly selected English-language solution descriptions and three state-of-the-art LLMs. A randomly selected subset of 20 examples was evaluated by a human expert. The results suggest that LLMs can often apply the proposed framework with substantial alignment to expert judgment, although errors related to over-classification and category-boundary ambiguity remain.

As mentioned before, future work should expand the validation process to a larger sample and include multiple human annotators to measure inter-annotator agreement and obtain more robust estimates of classification quality. Further work should also evaluate the framework at the level of individual binary questions, refine questions that produce systematic disagreement, and investigate whether examples, prompt variants, or hybrid human-in-the-loop workflows can further improve classification reliability.

References

1 Rodrigo Bravo Simões, Adriano Lopes, and Fernando Brito e Abreu. From Natural Language to Structured Queries: An LLM-Powered Chatbot for a Smart Tourism Observatory. In *Proc. of the 15th Symposium on Languages, Applications and Technologies (SLATE)*, Open Access Series in Informatics (OASIcs), Lisbon, Portugal, June 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. To appear. doi:10.4230/OASIcs.SLATE.2026.17.

2 Dimitrios Buhalis and Aditya Amaranggana. Smart Tourism Destinations Enhancing Tourism Experience Through Personalisation of Services. In Iis Tussyadiah and Alessandro Inversini, editors, *Information and Communication Technologies in Tourism 2015*, pages 377–389. Springer, 2015. doi:10.1007/978-3-319-14343-9_28.

3 Mario Casillo, Massimo De Santo, R. Mosca, et al. An ontology-based chatbot to enhance experiential learning in a cultural heritage scenario. *Frontiers in Artificial Intelligence*, 5, 2022. doi:10.3389/frai.2022.808281.

4 Youngjin Chae and Thomas Davidson. Large Language Models for Text Classification: From Zero-Shot Learning to Instruction-Tuning. *Sociological Methods & Research*, 55(2):501–567, 2026. doi:10.1177/00491241251325243.

5 Diogo Cosme. Prompts used in "Automating the Classification of Smart Tourism Tools with LLMs: A Taxonomy-Driven Approach" Paper, 2026. doi:10.5281/zenodo.19741250.

6 Diogo F. M. Cosme. Smart ETL and LLM-based contents classification: The european smart tourism tools observatory experience. Master's thesis, Iscte - Instituto Universitário de Lisboa, 2024. URL: http://hdl.handle.net/10071/34095.

7 Qian Dong, Lei Li, Damai Dai, Ce Zheng, Ke Ma, Rui Li, Yifan Xia, Jingjing Xu, Baobao Wu, Baobao Chang, et al. A Survey on In-Context Learning. *arXiv preprint*, 2024. doi:10.48550/arXiv.2301.00234.

8 European Commission. Leading Examples of Smart Tourism Practices in Europe from the 2025 European Capital of Smart Tourism Competition, 2025. URL: https://smart-tourism-capital.ec.europa.eu/best-practices/european-capital-smart-tourism-best-practices_en.

9 António Galvão, Fernando Brito e Abreu, and João Joanaz de Melo. Toward a Consensual Definition for Smart Tourism and Smart Tourism Tools. In E. Kornyshova, R. Deneckère, and S. Brinkkemper, editors, *Smart Life and Smart Life Engineering: Current State and Future Vision*, pages 153–183. Springer Nature Switzerland, Cham, 2025. doi:10.1007/978-3-031-75887-4_8.

10 Ulrike Gretzel, Marianna Sigala, Zheng Xiang, and Chulmo Koo. Smart tourism: foundations and developments. *Electronic Markets*, 25(3):179–188, 2015. doi:10.1007/s12525-015-0196-8.

11 A. Kontogianni and E. Alepis. AI in Smart Tourism: LLMs & GPTs Leading the Way. In *Proc. of the 15th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE, 2024. doi:10.1109/IISA62523.2024.10786696.

12 Arina Kostina, Marios D. Dikaiakos, Dimosthenis Stefanidis, and George Pallis. Large Language Models For Text Classification: Case Study And Comprehensive Review. *arXiv preprint*, 2025. doi:10.48550/arXiv.2501.08457.

13 Zongqian Li, Yixuan Su, and Nigel Collier. A Survey on Prompt Tuning. *arXiv preprint*, 2025. doi:10.48550/arXiv.2507.06085.

14 Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, 55(9):1–35, 2023. `doi:10.1145/3560815`.

15 Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. Deep Learning–based Text Classification: A Comprehensive Review. *ACM Computing Surveys*, 54(3):1–40, 2021. `doi:10.1145/3439726`.

16 SEGITTUR. Catalogue of Technological Solutions for Smart Tourist Destinations—Third Edition. Smart Tourist Destination, 2024. URL: `https://www.destinosinteligentes.es/wp-content/uploads/2024/02/Segitur_guia_fichas_2024@14703546813.pdf`.

17 Carlos N. Silla and Alex A. Freitas. A Survey of Hierarchical Classification Across Different Application Domains. *Data Mining and Knowledge Discovery*, 22(1):31–72, 2011. `doi:10.1007/s10618-010-0175-9`.

18 Giancarlo Sperlí. A cultural heritage framework using a Deep Learning based Chatbot for supporting tourist journey. *Expert Systems with Applications*, 183:115277, 2021. `doi:10.1016/j.eswa.2021.115277`.

19 Xiaofei Sun, Xiaoya Li, Jiwei Li, Fei Wu, Shangwei Guo, Tianwei Zhang, and Guoyin Wang. Text Classification via Large Language Models. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8990–9005, Singapore, 2023. Association for Computational Linguistics. `doi:10.18653/v1/2023.findings-emnlp.603`.

20 Min-Ling Zhang and Zhi-Hua Zhou. A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2014. `doi:10.1109/TKDE.2013.39`.