Point of Interest Recommendation: Pitfalls and Research Directions

ALEJANDRO BELLOGIN*, Universidad Autonoma de Madrid, Madrid, Spain
LINUS DIETZ, Department of Informatics, King's College London, London, United Kingdom of Great Britain and Northern Ireland
FRANCESCO RICCI, Free University of Bozen-Bolzano, Bolzano, Italy
PABLO SÁNCHEZ, Instituto de Investigación Tecnológica (IIT), Universidad Pontificia Comillas, Madrid, Spain

Point of interest (POI) recommender systems (RSs) aim at enriching tourists' visit experiences by suggesting context-dependent and preference-matching attractions and services at specific locations in a tourist destination, such as restaurants, parks, and cultural and historical attractions. Tourism, unlike some more common recommendation domains, such as music and video, requires a more structured and high-stakes decision process: tourists invest significant time, money, and effort to search, choose, and consume the selected POIs. Despite extensive research contributions in this area, based on our analysis of the relevant literature and our experience in designing, building, and testing POI RSs, we conclude that some fundamental issues of POI RSs are still unresolved, limiting the applicability of these systems in real-world scenarios. In this reflection article, we briefly summarize the research field and identify important pitfalls that challenge scientific progress and impact on businesses. To address these challenges, we outline research directions that may help move the field forward.

Therefore, the first contribution of this reflection article is a critical assessment of the current state of the art on POI RSs, and the identification of key shortcomings in three main dimensions: users and system log datasets, recommendation algorithms, and system evaluation methods. We highlight critical limitations, such as the lack of standardized benchmark datasets, flawed assumptions in the problem definition and model design, and inadequate treatment of biases in user behavior learning and system performance. The second contribution is a structured research agenda that, starting from the identified issues, introduces important directions for future research related to multistakeholder design, context awareness, data collection, trustworthiness, novel interactions, and real-world evaluation. We offer the proposed research directions, while not being exhaustive, as a contribution to address the identified pitfalls of POI RSs.

CCS Concepts: • **Information systems** → **Recommender systems**; *Evaluation of retrieval results*; *Electronic commerce*; *Users and interactive retrieval.*

Additional Key Words and Phrases: Recommender Systems, Tourism, Data, Evaluation, Algorithms

1 Introduction

Tourism offers a rich and multifaceted application scenario for the development of recommender systems (RSs). By assisting tourists in processing information collected on web portals and social networks, and choosing alternative products and services, RSs can both enhance tourists' travel experiences, and impact on the economic, social, and environmental dimensions of the visited destinations. In this reflection article, we focus on a type of tourism

*All authors contributed equally to this research and are listed in alphabetical order.

Authors' Contact Information: Alejandro Bellogin, Universidad Autonoma de Madrid, Madrid, Spain; e-mail: alejandro.bellogin@uam.es; Linus Dietz, Department of Informatics, King's College London, London, United Kingdom of Great Britain and Northern Ireland; e-mail: linus.dietz@kcl.ac.uk; Francesco Ricci, Free University of Bozen-Bolzano, Bolzano, Italy; e-mail: francesco.ricci@prof.senior.unibz.it; Pablo Sánchez, Instituto de Investigación Tecnológica (IIT), Universidad Pontificia Comillas, Madrid, Spain; e-mail: psperez@icai.comillas.edu.

ACM 2770-6699/2026/5-ART
https://doi.org/10.1145/3816430

RSs, which generate point of interest (POI) recommendations. POI recommendations are context-dependent and tourists' preference-matching attractions and services at specific locations in a tourist destination, such as restaurants, parks, and cultural and historical attractions. Reflecting on a targeted and concise assessment of the state of the art and its research gaps, we derive a few non-exclusive research directions aimed at addressing these pitfalls, which we believe currently hinder the wider adoption of POI recommender systems. The proposed research lines are also informed by a reflection on how POI RSs should be designed and optimized to really help travelers in their decision-making processes. Moreover, we focus on how POI RSs can manage the trade-offs between tourists' preferences satisfaction and the long-term sustainable development of the visited destinations [13, 124].

RSs are becoming more and more common in daily life, and POI recommendation is no exception. POI catalogs are featured on websites and applications used for navigation, travel planning, and booking. Based on tourists' expressed preferences and the analysis of their past choice behavior, RSs help them find landmarks, restaurants, museums, and even hidden attractions, making it easier to explore unfamiliar destinations [130, 143]. However, tourism-related decision-making is quite different from the processes involved in other, more traditional recommendation domains, such as books or movies [27]. The stakes are higher in tourism as travelers invest more time, money, and emotional involvement to search and select the desired products and services. The success of an RS depends, not only on the right match of the collected and learned tourist preferences with the recommended items, but also on a wide range of highly dynamic contextual factors [10, 63, 83]: the specific visit intent/goals of the traveler, the traveler's personality and mood [24], the specific atmosphere of a venue in a particular situation [149], the season and the weather at the visit time [10], and the composition of the traveler's group [6, 140]. These factors, combined with the variability of traveler and travel types, make it difficult to design effective RSs, i.e., that ensure a smooth and effective user-system interaction and ultimately contribute to making satisfactory travelers' decisions. Moreover, such variability and dynamicity make it complex to properly evaluate POI RSs with the necessary rigor and significance, by means of offline or online experiments. As a result, research in this area has often overlooked real usage scenarios, considering oversimplified contexts, producing results whose practical value is difficult to interpret. That ultimately complicates the transfer of research RS prototypes to operational solutions [35, 162].

Hence, in this reflection paper, we aim to identify important shortcomings in the research on POI RSs. Our goal is to contribute to the development of best practices in the academic community by critically examining common problems across three main interconnected dimensions: user behavior and system log datasets, recommendation algorithms, and RSs evaluation methods. Based on our expertise and supported by a targeted discussion of the state of the art, we have identified 20 pitfalls that we consider both widespread and critical. Although by no means exhaustive, these pitfalls highlight key challenges whose resolution would substantially strengthen the impact of research in this field. In fact, we believe that by providing such an organized list of POI RSs' shortcomings, we can raise awareness on critical challenges, so that the field can be better understood and more relevant and significant research directions can be prioritized.

The first pitfall category relates to the availability and utilization of datasets. We observe that offline evaluations of POI RSs are conducted relying almost exclusively on outdated, incomplete, and biased data sets, which are obtained mainly from Location-Based Social Networks (LBSNs). In contrast, user studies, which have been used to collect more detailed features of tourists' preferences and behaviors, have other limitations: they involve small samples of subjects and generate tiny data sets, which are useful for addressing specific research questions, but cannot be used to test general system applicability and effectiveness. We believe that such data-related issues limit our understanding of the true effect of POI RSs on real users.

Second, we observe that core POI recommendation algorithms often prioritize accuracy metrics, which, while important, can overshadow other valuable goals, such as contextual relevance, diversity, and fairness. This narrow focus tends to favor algorithms that prioritize popular places, which are easily predicted as relevant for any tourist, hence reducing personalization and limiting the discovery of other minor but still relevant POIs. In fact, recommending well-known places adds little value, because travelers are also easily made aware of popular places by alternative information channels, such as newspapers, social networks, and web portals, which are simpler to implement and maintain. We recognize that what makes POI RSs especially useful, and also sustainable, is their potential to foster the virtual exploration of a destination, and the discovery of novel POIs, matching travelers' intent and preferences, that might not have been easily found by using other information search tools [15].

Third, we provide a critical analysis of how POI RSs are evaluated. We stress again that many studies rely on a narrow selection of performance metrics in offline testing environments and, therefore, fail to capture the variety and complexity of the true tourists' behavior. Moreover, the lack of standardized data pre-processing strategies leads to experimental outcomes that are hard to interpret and match to the dynamic and context-sensitive nature of travelers' behavior. Finally, we criticize the research focus on studies aimed at incremental improvements of well-known general techniques, missing the more substantial challenges posed by the reproducibility and the proper measurement of algorithmic performance [33].

Based on the identified pitfalls, we propose a research agenda comprising a selection of research directions aiming to address the identified limitations of current POI RSs. Like the identified pitfalls, these research directions are not exhaustive. They are proposed here as those we believe can best help address the challenges highlighted. The research lines are: multistakeholder system design, context-aware recommendations, collection of tourist behavior data, trustworthiness in system design, novel interaction modalities, and real-world evaluation. In more detail, the proposed research agenda calls for the adoption of richer and more reliable datasets, together with recommendation algorithms that can adapt to dynamic contextual factors, such as user intention, social group dynamics, and destination environment (e.g., weather, traffic, transportation). In conclusion, we consider that POI RSs research needs to better frame the POI recommendation task, target broader multistakeholder objectives, and adopt more reliable evaluation procedures, which exploit more significant data and are transparent about data pre-processing.

We wrote this paper because we were dissatisfied with some current practices in POI recommendation that we think limit the impact of the scientific work. We are university researchers based in Europe, with academic interests primarily rooted in computer science, specifically in the core methodological foundations of recommender systems. Our scientific background shapes how we view challenges and prioritize research opportunities in POI recommendation. Due to this, this reflection paper puts less emphasis on tangential aspects of related and relevant disciplines, such as tourism management, human-computer interaction, or business and policy aspects of tourism ecosystems.

The remainder of the paper is structured as follows. Section 2 introduces the application problem, stating the specific characteristics and challenges of POI recommendation in tourism contexts. Section 3 briefly reviews the state of the art, summarizing the main approaches and research topics in the field. Section 4 discusses the limitations and open issues related to the datasets commonly used in POI recommendation works. Section 5 focuses on algorithmic limitations, highlighting aspects such as evaluation biases, lack of contextual modeling, and ignoring multistakeholders, while Section 6 addresses widespread shortcomings in evaluation practices. Section 7 presents our research agenda to address the previously mentioned issues. Finally, Section 8 summarizes the lessons learned and concludes the paper.

2 Application Problem

Point of interest recommendations can play a crucial role in supporting travelers in the whole decision-making process, in particular, by helping them discover relevant venues and better organize their visits throughout the different stages of the trip. Since multiple players are involved, we initially give an overview of the key stakeholders and how they are connected. Moreover, to properly define the POI recommendation problem requires

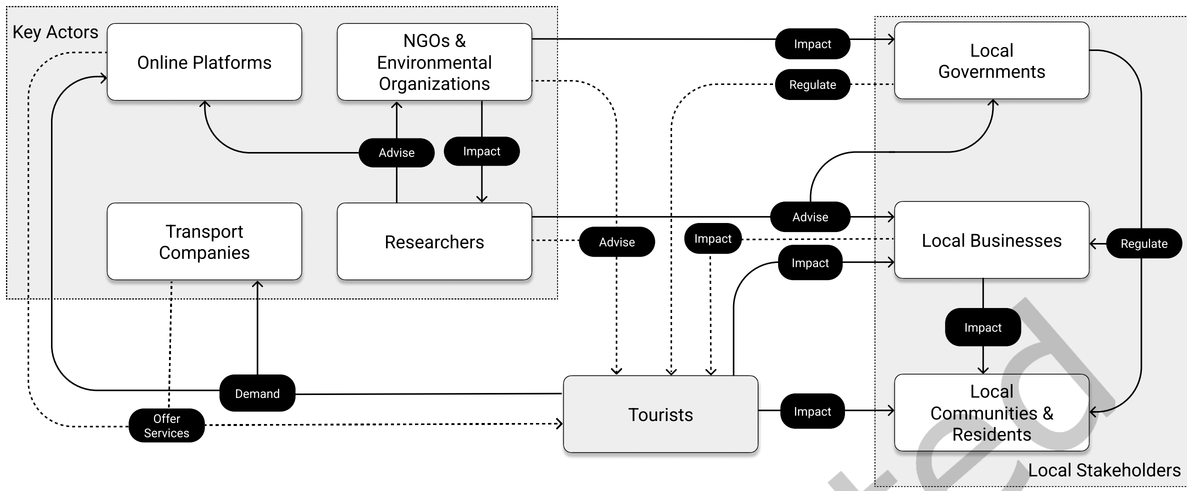

Fig. 1. Relationships between the main stakeholders of the tourism market. The relationships are based on [85, 138, 163]. For clarity in presentation, we used dashed lines to highlight all influencing aspects on the tourists.

differentiating between tourist information processing and decision-making activities that are conducted before the travel (*pre-trip planning*) and during the trip (*in-trip recommendation*) [142].

2.1 Stakeholders Involved in POI Recommendation

Tourists have diverse necessities and constraints, such as preferences for certain types of venues, time or budget constraints, the kind of suggestions they want, or they require to meet the preferences of others in their group. These needs and constraints affect the choices they make and are often influenced by promotion activities performed by a wide range of stakeholders in the tourism market, such as destination management organizations (DMOs), local governments, booking platforms, tour operators, airline companies, travel agencies, or even social network influencers [85, 138, 163], as depicted in Figure 1. Each stakeholder has its own economic, social, and political objectives, which could be complementary, but some players may also have goals that conflict with the interests of tourists, depicted in the center of the figure. Hence, the tourism recommendation scenario must be framed as a *multistakeholder scenario* [1, 22, 101] and an effective RS must properly balance the goals of the stakeholders involved when constructing its recommendations.

Early work on multistakeholder recommendation focused only on the immediate stakeholders: tourists seeking personalized and satisfying experiences, and businesses and e-tourism platforms seeking revenue through bookings. Today, local businesses are also influenced by community expectations for more sustainable regional development, which can lead to regulations such as visit-time limits at overcrowded POIs or restrictions on certain types of establishments. Communities expect governments to support economic growth while preserving social and environmental resources. Many DMOs are now seeking strategies to promote lesser-known areas to help local businesses grow while reducing overcrowding in more popular venues[77, 114–116, 139]. Herein, the goal is to achieve a fair exposure of all the areas of a larger region, so that even the less popular ones can attract visitors and compete with the few already popular ones [13]. In fact, some tourists are explicitly expressing interest in sustainable travel information and suggestions, and are ready to accept extra costs or limitations, when the goal is to improve the sustainability of their trip [48, 69].

When designing POI RSs, it is not enough to match user preferences and constraints with the available POIs. Other goals also matter, such as the overall business development, the spatial dispersion of tourists, the well-being of local communities, and the general sustainability of the involved processes [117], while maintaining the tourists' satisfaction at the center of the recommendation process. POI RSs research must keep this complex ecosystem in mind and ideally should be able to advise all the connected stakeholders on how to offer or consume services that are to the benefit of everyone.

Moreover, to better frame the POI recommendation problem, it is important to recognize that what we refer to as a tourist is often a group of tourists. Families, friends, couples, or ad hoc organized groups plan their visits and experience together the chosen POIs. This dimension increases the complexity of the recommendation task, as within a group, its members can exhibit divergent preferences and constraints [70]. Developing POI RSs that consider both the potentially conflicting objectives and the dynamics of group decision-making is necessary to generate fair and inclusive recommendations that satisfy all group members.

2.2 Pre-trip Planning Support

In a tourist's customer journey, *pre-trip planning* refers to the high-level selection of which city or region to visit, and the search for information on the destination's main services and attractions, i.e., which accommodation to book and the most valuable POIs to visit once the user is there. It can be performed over a long time frame, e.g., weeks and months. In this phase, tourists usually interact with various systems, such as booking platforms, transportation and navigation apps, destination portals, etc., exchanging and processing a considerable amount of information. The notion of *destination* lies in the core of this phase and defines its goal. However, the concept of *destination* is not limited to a geographical place [41]. It is a concept defined by the cultural background of the traveler, the prior knowledge of the destination, and the distance of the traveler to the destination [44, 57]. A destination is also dynamically defined and redefined by tourists' behaviors. Places or activities that were not considered attractive may become appealing through content shared on social media, tourism websites, and other communication platforms. Targeted media campaigns and the spread of information can influence visitors' interests and attention. Moreover, the reasons for visiting a given POI can differ from one tourist to another and be related to factors, such as the tourist's cultural background, prior knowledge, and specific travel goals, whether to study, relax, satisfy curiosity, or something else [55, 113, 119].

Therefore, RSs designed to support *pre-trip planning* must go beyond traditional information filtering, which is the most common approach in the available solutions. That is, instead of simply selecting items, they should act as intermediaries, offering personalized content that matches the different cultural and emotional values of the destination with the traveler's interests [142]. Hence, effective POI RSs need to show that the recommendations satisfy the underlying motivations and preferences of users through clear descriptions and explanations [147]. Moreover, it is important to segment the venues of a larger territory into tourists' targets, i.e., to define and describe destinations whose visits can better match the wide variety of tourists' needs [44, 62, 112]. Thus, POI RSs should become moderators of complex information exchange processes between suppliers and consumers to identify underused areas, suggest promotional opportunities for emerging places, and support marketing strategies based on traveler demand and behavior, especially when the destination is new to the user [143].

2.3 In-trip Recommendation

In contrast to the pre-trip recommendation, *in-trip recommendation* is strongly influenced by the real-time context of the visit to a selected destination, including the availability of POIs, the position of the tourist and its distance to alternative POIs, time and weather constraints, accompanying group dynamics, and the evolving preferences of the tourist [42, 132]. Hence, in-trip POI RSs should support the real-time, context-dependent, decision-making processes that tourists perform when visiting a selected destination. Context can refer to multiple concepts, depending on the application scenario and the discipline of analysis [2]. We argue that in the case of POI recommendation, relevant contextual information falls into three main categories. First, *POI-related context* includes factors such as distance from the traveler's position, estimated travel and visit times, POI opening hours, and ongoing events in the POI's area. Second, *the user-specific context* refers to current interests, needs, and constraints of the traveler, such as fatigue, hunger, mood, or group preferences [23, 64, 72, 104, 127]. Third, *external context* includes factors not directly related to the tourist but still influential, such as weather conditions, air pollution, temporary disruptions, urban noise, or crowd density [132, 149]. The combination of these three types of contextual factors can significantly influence tourists' decision-making process, potentially leading to changes in their original pre-trip planned targets and itinerary. For example, a group may plan to visit two museums in a row, but after the first visit, they could decide to rest in a nearby bar before continuing [65].

3 State of the Art

In this section, we briefly survey some important results from the state of the art of POI recommendation. We organize the discussion around three core dimensions: data, algorithms, and system evaluation. The aim is to establish the necessary background to identify existing research gaps and directions which can address these gaps; these are the two main contributions of this work. This state-of-the-art survey is not comprehensive; a much larger space is necessary, which is beyond the scope of this reflection article. The reader can consult recent systematic surveys [130, 160], which offer more comprehensive and detailed analysis of the literature.

3.1 Data Used in POI Recommendation

Datasets based on Location-Based Social Networks (LBSNs) such as Foursquare, Gowalla, or Yelp, have been widely used in the POI recommendation literature, as they provide large-scale check-in data from users with diverse profiles, including different types of tourists and locals [130]. In some cases, these LBSNs also provide information regarding the POIs, such as their type (e.g., park, historical building) and reviews, as well as user data, including gender and followers. However, despite their usefulness and usage, these datasets present some limitations that must also be considered and will be discussed in Section 4. In addition to LBSNs, some works use photo-based platforms such as Flickr, where users store, organize, and publicly share POI images, together with rich metadata. TripBuilder [26] is an example of a dataset based on Flickr used in tourism RSs: it includes reconstructed POI visit trajectories in three major Italian touristic cities. These data logs were created in a fully automated process by using geo-tagged photos from Flickr, and subsequently identifying the featured POIs and gathering additional information by using Wikipedia. This Flickr data was also used in [73], where the authors dealt with errors in both timestamps and geographical coordinates of the originally extracted trajectories. This issue was also pointed out by [43]. However, Flickr data provides good indicators of the spatial regions preferred by tourists and POI popularity, by assuming that photo sharing reflects user interest. Among the publicly available Flickr datasets, the Yahoo Flickr Creative Commons 100 Million Dataset (YFCC100M) [146] stands out for its scale and usage. It has served as the basis for multiple derived datasets in tourism and POI recommendation [88, 89].

Other recent data sources are emerging outside the domain of LBSNs. For example, the YJMob100K dataset [166] contains large-scale anonymized human mobility trajectories collected from 100,000 mobile phone users over 90 days, covering both regular and emergency scenarios. This dataset, originally intended for the prediction of urban mobility, may offer promising opportunities to evaluate POI and trajectory recommendation models under realistic and changing conditions. Tourist-oriented data can be much less noisy when interactions with POIs are explicitly recorded, for example, through visit actions with visitor cards. Although such data is rare, a notable case is when tourists log their activities using a visitor card or city pass, as done in the Italian city of Verona [102].

Finally, when real data is scarce or incomplete, synthetic datasets have been proposed to fill this gap. For example, Merinov et al. propose a mechanism that generates individual-level user profiles from aggregated population-level tourism data, and simulates POI visits using a discrete choice model with divergence minimization. A recent trend in the community for creating synthetic datasets consists of using large language models (LLMs) to emulate users. In this case, instead of generating full trajectories or proposing a new recommendation algorithm, Banerjee et al. focus on generating realistic and diverse travel queries used by an RS based on LLMs [14].

In summary, while traditional LBSNs datasets provide large-scale check-ins and rich metadata, they suffer from quality drawbacks regarding the check-ins [172]. Other sources based on photos also contain errors regarding the coordinates or the temporal information [73], although they might provide better signals of the tourists' preferences [88, 89]. Other datasets based on mobility and visitor-card logs capture more realistic preferences and trajectories; however, their use is not so widespread. In Table 1 we present a selection of datasets commonly used for addressing the POI recommendation problem and other related tasks, such as next-POI or trajectory recommendation. The table also includes details like access links and key data attributes. It is worth noting that these datasets vary in popularity and usage within the research community. Finally, when real data is scarce, synthetic datasets (sometimes obtained using LLMs) are used to represent user profiles and travel queries. However, some authors have shown that if synthetic datasets are not carefully designed, they may end up exhibiting systematic biases such as popularity bias [34]. In Section 4, we discuss more in depth these problems in the Pitfalls 1 to 7.

3.2 Recommendation Algorithms and Models

Researchers targeting tourism and POI RSs have proposed a wide variety of algorithmic approaches for their development, ranging from traditional nearest neighbors and matrix factorization techniques to more recent strategies based on neural networks. Classical models–especially those based on content and user profile similarity–are often used because of their interpretability and simplicity. These approaches, together with matrix factorization, try to predict user preferences based on previous check-ins or ratings by identifying similar patterns or latent interactions in the user-POI interaction or rating matrix. Content-based information, which is normally obtained from POI attributes (e.g., categories, textual descriptions, or geographical information) and user profiles, has also been explored as a source of side information in collaborative-based approaches. For example, Gao et al. proposed in [52] a matrix factorization approach with content information, while in [30], Chang et al. proposed a content-aware POI RS based on neural networks (using convolutional neural networks, multilayer perceptrons, and an attention mechanism). In fact, the use of these classical techniques has been somewhat displaced due to advances in neural network-based architectures, which, in recent hybrid proposals, can also use nearest neighbors or matrix factorization algorithms as part of their model [130].

Even more recently, the rapid progress of LLMs in recommender systems has also gained attention in both tourism and POI RSs [37]. For example, Li et al. address the problem of next-POI recommendation by proposing a new method that uses pretrained language models to better exploit the rich contextual data from LBSNs [87]. Unlike previous approaches that rely on numerical representations, the proposed framework maintains the original structure of the data and relies on contextual information. Wang et al. present SeCor [157], a new approach to next-POI recommendation (i.e., which POI to visit after having visited an initial sequence of POIs) that combines collaborative and semantic information using a multi-modal strategy. It addresses limitations of previous models in capturing complex spatio-temporal patterns, while also mitigating LLM-induced hallucinations. By combining user–POI interaction embeddings with rich semantic representations, their model generates more robust and expressive hybrid encodings that improve recommendation performance.

Reinforcement learning has also been adopted as a viable solution to the challenges of POI recommendation, such as the cold-start problem. Massimo and Ricci introduce a reinforcement learning tourism RS called *"QEXP"*

Table 1. Datasets used in publications dealing with POI recommendation and related tasks (ordered alphabetically). We include the links of the datasets if they are available to the public and other important characteristics (we use NA if the information regarding a specific column is not available). We refer to the reader to [130, 132] for more details regarding some of the datasets.

Dataset	**Timeframe**	**Size & Type**	**Area**	**Venue details**	**Additional information**
Brightkite	2008 - 2010	4.49M (check-ins)	Worldwide	Coordinates	User friends
Context Trails	2017 - 2018	1.3M (check-ins)	Tokyo, Petaling Jaya, New York	Categories, coordinates, price, schedule	Trajectories, weather
Foursquare Global-Scale	2012 - 2013	33.3M (check-ins)	Worldwide	Categories, coordinates	NA
Google Local	2002 - 2021	666M (reviews)	United States	Categories, coordinates, descriptions, price, schedule	Photos, text
Gowalla	2009 - 2010	6.4M (check-ins)	Worldwide	Coordinates	User friends
Semantic Trails 2018	2017 - 2018	11.9M (check-ins)	Worldwide	Categories	Trajectories
Trip builder	2007 - 2012	133K (check-ins)	Pisa, Florence, Rome	Categories, coordinates	Photos, trajectories
VeronaCard	2014 - 2023	2.7M (check-ins)	Verona	Categories, coordinates, popularity, time visit	Crowdedness, weather
Yelp 2025	2005 - 2022	6.99M (reviews)	11 metropolitan areas	Categories, coordinates, schedule	Photos, review text, stars, user friends
YJMob100K	75 days	141M (records)	Japan	Categories	Emergency scenarios
YFCC100M	2004 - 2014	99.2M (photos), 0.8M (videos)	Worldwide	NA	Photos, videos

that uses a model based on the behavior of POI visits extracted from POI visits logs. A related approach introduces a hierarchical reinforcement learning framework named *"HRL-PRP"* designed to improve POI recommendation by better capturing complex user behavior [164]. In contrast to traditional models, their proposal revises user profiles through a two-level decision process: a high-level component decides whether the user's profile should be updated, while a low-level component selects which noisy POIs from the user's history should be removed.

Cross-domain and transfer learning approaches have also been used to mitigate training data sparsity and improve recommendations. For instance, Zheng et al. exploited information from GPS logs, POI databases, and the Web to obtain location-activity, location-feature, and activity-activity correlations to improve recommendation performance [173], whereas Sánchez and Bellogín demonstrate that augmenting data with nearby cities' check-ins improves accuracy in cities with fewer interactions [129]. Similarly, Zhang and Wang use clustering over city regions to support recommendations in unfamiliar cities [170].

Conversational recommender systems are also particularly useful in travel and tourism [71]. For example, Nguyen and Ricci proposed a novel approach to address the group POI recommendation problem [103]. Their mobile system tracks and exploits users' interactions during group discussions to recommend POIs and offer other helpful suggestions to guide the group to reach consensus. Li et al. present a reinforcement learning approach that integrates geographical patterns with dialogue interactions for next-POI recommendation [86]. Zhang et al. propose a conversational translation approach for the next-POI recommendation problem under uncertain check-ins [171]. Their method combines a recommendation module that captures both sequential information and recent preferences from conversations, with a conversational module to improve recommendation quality while minimizing conversational turns.

Another major research line has adopted operations research techniques to solve optimization problems in itinerary generation and route planning. Herzog et al. give an overview of the different problem formulations [63], including the "Time Dependent Team Orienteering Problem with Time Windows" [54] to model route planning for travelers who wish to visit various POIs using public transport, the "Vacation Planning Problem" [150], in which a traveler wants to explore an extensive geographical area, or the "Tourist Trip Design Problem with Travel Instructions" integrating public transport data [7].

In summary, classical similarity-based and matrix factorization approaches —sometimes incorporating content-based features— are still used. However, most current works make use of neural networks in the recommendation pipeline, often integrating content information as well. In some cases, multi-modal approaches are also considered in order to better exploit contextual data. Other research lines include reinforcement learning, cross-domain, and transfer learning, in order to address the cold-start problem or to mitigate sparsity, and conversational RSs that help users in deciding their next visits through a dialogue. Finally, operations research techniques are used for itinerary generation and route planning, considering constraints related to available time windows and public transport.

It is worth noting that most of the current POI recommendation approaches focus only on building solutions that optimize accuracy metrics [130] (i.e., the accuracy in predicting the POIs that the tourist will actually visit), ignoring other relevant evaluation dimensions, which are discussed in the next section. Besides, other problems arise; for example, the generated recommendations may be biased, affecting all the involved stakeholders [131], and models may not adapt well to changing contextual conditions or evolving user preferences [81]. Moreover, if the recommendation process is not transparent enough, the user may easily mistrust the system. In Section 5, specifically through Pitfalls 11 to 20, we provide a more detailed analysis of these challenges, related to algorithm design.

3.3 Evaluation Practices

The evaluation of POI RSs has traditionally relied on offline experiments, i.e., by using the datasets mentioned above in Section 3.1, and measuring the performance only of the core recommendation algorithm, in terms of how well it generates recommendations which appear to be suited to the logged users, according to some hold out test data. That is, an available dataset of users consumptions (e.g., check-in or reviews) is split into training and test sets, either temporally or at random. In particular, the split can be done per user, based on each user's check-in history, or globally, treating all interactions collectively. Then, recommendation algorithms are trained on the training set and evaluated for their ability to predict which venues users interacted with (checked-in or reviewed) as recorded in the test set. Common evaluation metrics include Precision, Recall, Normalized Discounted Cumulative Gain (nDCG), and Hit Ratio, usually measured at a fixed cutoff [60]. Due to the often sequential nature of the POI recommendation problem, some authors have proposed metrics that consider the order of visited POIs, such as the pairs-F_1 proposed by Chen et al. [31] or sequential adaptations by Sánchez and Bellogín of the previously mentioned metrics [128].

However, although practical and reproducible, this type of evaluation does not fully reflect the dynamic and interactive nature of the real-world recommendation task, and the available test data is not sufficient to properly validate the recommendations, as they cannot fully simulate the possible user feedback to the recommendations. To address the inherent limitations of offline evaluation methods, user studies have also been adopted to better understand user perception, satisfaction, and behavior, while assessing the system-generated recommendations. For example, Hazwani et al. show in their user study how to mitigate the cold-start problem by showing that people who participate in storytelling experiences are more willing to provide explicit preferences [61]. In addition, Hofschen et al. analyzed the divergence between the expected utility of users before visiting a POI and their experience after the visit, collected through a web-based interactive survey. The results showed that these two types of utility often lead to different preference patterns [66].

Other studies focus on the perceived usefulness and novelty of recommendations. For example, some studies found that although algorithms can generate highly novel recommendations, many users find it difficult to appreciate items that are not familiar with [95, 123]. Complementing these findings, research on group decision-making in tourism has explored how group preferences evolve during collaborative destination planning [40]. The findings highlight the need for systems that support dynamic group interactions and help build consensus.

Beyond user studies, simulation-based evaluations are also gaining traction as a way to assess recommender systems in scenarios that are difficult to observe directly, such as the tool proposed by Piliponyte et al. that allows DMOs to simulate the impact of an online promotion campaign [118]. This simulation uses recommendation techniques to select the destinations that are promoted to each tourist. Simulations of interactions in a multistakeholder RS can be optimized to benefit both users and destinations by maximizing tourist usefulness, while fostering unpopular venues [101].

Hence, most of POI recommendation works still evaluate their proposals with offline methods based on the classical train-test splits of an available data set, and the computation of accuracy metrics like Precision, Recall, or nDCG. While the evaluation of additional dimensions such as novelty and diversity is becoming more common in other application domains, these aspects are still largely overlooked when recommending POIs. Furthermore, many details regarding evaluation strategies are often insufficiently reported, which limits the reproducibility of the obtained results [130]. Classical offline evaluation methods also fail to capture the interactive nature of the real recommendation task; thus, user studies are employed instead. In such studies, it is essential not only to capture user satisfaction but also to gather a more fine-grained assessment of the interaction, including their opinions on the explanations provided for the recommendations. In Section 6, we analyze the extent of these issues in Pitfalls 16 to 20.

4 Pitfalls in Widely Adopted Datasets

In this section, we examine key challenges related to the use of existing data sets in POI recommender systems. Our analysis is organized by data source, where the two most prevalent primary sources in current research, location-based social networks and user studies, are discussed in detail. We also briefly touch on more unconventional or emerging data sources that have been or could be used to support the development and analysis of POI RSs.

4.1 Issues with Location-Based Social Network Datasets

LBSN datasets, discussed in Section 3.1, present several challenges that must be considered when used to develop POI recommender systems. As LBSN data is the most commonly used source for offline evaluation of POI recommendation models, we exhaustively analyze it.

Pitfall 1: Outdated data leads to mismatches with today's reality. Many publicly available LBSN datasets are outdated, as some of them were collected more than a decade ago [16, 130]. The lack of published and recent datasets is likely to be attributed to tightened data protection legislation in the Western World, such as GDPR [50], as robust anonymization of individual mobility is practically unattainable without the loss of essential information typically required to train recommendation models [153, 156]. This temporal gap presents significant challenges: POIs may have closed, relocated, or their experience may have changed substantially. As a result, models trained on such data may end up recommending venues that no longer exist or are no longer relevant, undermining the system's real-world applicability [47]. In addition, user behavior recorded in these datasets may no longer represent actual mobility trends, particularly in urban environments where user preferences frequently change.

Critically, to the best of our knowledge, no publicly available POI recommendation dataset has been collected and released following the COVID-19 pandemic. This absence presents a significant gap, as the pandemic has had a lasting impact on urban mobility [134, 165].

Pitfall 2: Mismatches between training data and target RS users impede reliability of model performance. Several studies have revealed a mismatch between the training data used in POI recommender systems and the diverse behaviors of their target users [129, 133]. For example, research on the Foursquare global-scale check-in dataset has shown that most users exhibit highly localized behavior, with check-ins concentrated in a single city [129]. This bias limits the dataset's utility for modeling tourist activity, which often implies cross-city or international travel. Moreover, reducing users to a binary classification of "locals" and "tourists" oversimplifies the complexity of real-world behavior. Each group can be further segmented based on the patterns available in their check-ins [43]. Among locals, distinctions emerge based on the types of venues they visit, while tourist behavior may vary significantly depending on the scope of their travel, i.e., if they are regional, domestic, or even intercontinental visitors, affecting the performance of different recommendation models [133]. A common pitfall in this context is that POI recommender systems designed for travelers are trained and evaluated with interactions provided predominantly by locals, resulting in unreliable results.

Regarding temporal aspects, Noulas et al. analyzed 12 million Foursquare check-ins, finding a significant disparity in check-in patterns between weekdays and weekends. However, in both cases, check-ins at residential venues increased constantly throughout the day, suggesting a behavior more typical of residents than of tourists [106]. Some authors have shown that user POI visiting patterns differ across temporal states, such as working vs leisure time or weekdays vs weekends. Hence, treating check-in data as temporally homogeneous, not accounting for variations in temporal contexts, may affect temporal regularities, reducing the effectiveness of POI recommendation models [21, 121].

Finally, there is concern that LBSN data may not accurately represent ground truth. For example, a study of the Murcia region in Spain [4] revealed discrepancies between venues registered in Foursquare and those detected on-the-ground observations. Their study shows divergences of more than 80% in the retail sector and more than 30% in food and entertainment venues. Another aspect of unreliable data is the posts/check-ins performed by bots, which should be detected and excluded from the analysis [108].

Pitfall 3: Incomplete data increases sparsity. The User-POI interaction matrices in LBSNs are extremely sparse. Most users check in at only a few venues, while many POIs receive few or no check-ins. This imbalance makes it difficult to learn actionable user preferences and intensifies the cold-start problem for both users and locations [16]. To illustrate the extent of this sparsity, standard benchmarks in traditional recommendation tasks such as the Netflix and MovieLens20M datasets have interaction densities slightly above 1.5% and 0.5%, respectively. In contrast, LBSNs datasets like Foursquare and Gowalla have densities as low as 0.003% and 0.005% [130]. Furthermore, LBSN data is inherently incomplete as it is based on voluntary contributions from users. Unlike domains such as music or video streaming, where every user interaction is automatically logged by the system, LBSNs depend on users actively recording their visits, such as checking in or posting content. This self-selection introduces important bias and coverage gaps [53, 136], as we further discuss in Pitfall 4. To address this problem, it may be beneficial to enhance existing datasets with additional contextual information using cross-domain techniques. For example, an increase in visitor activity at a particular place may be due to external factors, such as a major event (e.g., a sports match, concert, or cultural festival), while favorable weather conditions may also lead to increased mobility compared to rainy days.

Pitfall 4: The types of logged visited POIs lead to bias in the datasets. Importantly, missing data is not missing at random (MNAR). This means that the probability of missing data depends on hidden causes, such as user preferences, behaviors, platform design, and the type of users who use these platforms. This non-randomness introduces different types of biases, which can significantly affect the performance and fairness of the RS [159]. For instance, Wang et al. observed substantial discrepancies between the check-ins recorded on Foursquare and the actual mobility of the users, due to privacy considerations, lack of interest, or simply forgetting to check in in the venues [155]. Some POIs, especially those that are considered socially sensitive or niche (e.g., adult entertainment locations or particular nightlife spots), may be underrepresented in LBSN datasets. This scarcity of check-ins is not necessarily due to the absence of visitors but rather to users' reticence to publicly associate themselves with such places. Similar behavior has been observed in other domains, such as when users omit certain music tracks from their listening history to maintain a curated public image [46]. These types of bias, introduced by oversharing certain content or self-censorship, raise serious limitations to developing accurate and equitable POI recommender systems. When the underlying data is not only incomplete but also systematically biased, models may not capture the real preferences and behaviors of users. Discrepancies can also arise from both extraneous check-ins (e.g., users checking only to earn rewards [155]) and missing check-ins at sensitive locations such as healthcare facilities or political venues [5].

Pitfall 5: Platform-specific interactions bias the temporal and spatial patterns of user behavior. Online platforms exhibit different interaction patterns, which can alter how user behavior is observed and interpreted. For example, review-based platforms may offer rich information on POIs [168], but often lack temporal accuracy. This is because reviews are typically posted some time after the original visit, hiding when the interaction actually occurred.

Similarly, research that relies on photo-sharing platforms like Flickr often uses geotagged images to infer user trajectories. In such cases, researchers attempted to map geotagged photos to real-world POIs to reconstruct users' travel paths [43, 89]. Like reviews, there are challenges in the geographical or temporal inaccuracy: more than one POI may exist at the recorded coordinates (because of a low resolution or granularity, or simply because several venues exist very close to each other), users may upload content long after it is captured, and many of the available datasets from the pre-smartphone era have unreliable timestamps of creation [146]. Due to these uncertainties, it becomes difficult to accurately reconstruct tourist trajectories. Furthermore, geotags are not always automatically recorded but are often manually added, which introduces further uncertainty in location data. Finally, the content shared on social networks is often event-driven [78]. Users are more likely to post photos during festivals, holidays, or peak tourist seasons, which incorporates temporal patterns [17]. This seasonal bias makes it difficult to extrapolate findings to less frequent periods or to formulate year-round generalizations about tourist behavior.

Pitfall 6: Prevalent data collection practices raise privacy and ethical issues. The data collection methods for many of the datasets present additional ethical issues. The use of LBSN data without explicit user consent, combined with the ability to re-identify individuals based on location patterns [36, 126], is incompatible with GDPR legislation [50]. Many of the freely available datasets were collected prior to entry within the GDPR, and contain sensitive and difficult to anonymize location information about users who probably did not agree to their data being processed [148]. Furthermore, data collection that breaches platform terms of service can raise legal issues related to users' rights and may expose researchers who publish such datasets to legal action. Currently, it seems that much research relies on datasets that do not meet high ethical standards. To mitigate the issues linked to individual mobility datasets, the research community will need to rethink its practices and likely shift toward application-based data-sharing platforms for sensitive data, as is common in fields such as medicine or psychology [18, 79, 125].

Pitfall 7: Limited knowledge of user context reduces personalization. LBSNs datasets often do not contain rich contextual and demographic details, such as users' age, gender, travel purpose, or personal preferences. This lack of information limits the ability to generate personalized POI recommendations that take into account important distinctions, such as whether a user is a tourist or a local, or whether a traveler is visiting a location for leisure, business, or other reasons. While some research has explored alternative data sources to enrich the user's context in terms of weather [149] or by analyzing user-generated photos to infer travel context [141], such approaches are not commonly integrated into traditional LBSN platforms.

In addition, the influence of group dynamics is frequently neglected [40, 49]. It is often unclear whether a registered check-in represents individual patterns or is influenced by group contexts, such as traveling with friends, family, or colleagues. A single user may show different behavior depending on the social environment, leading to heterogeneity in activity patterns [154].

4.2 Issues with Data from User Studies

User studies are commonly used during the development and evaluation of recommender systems. As in the social sciences, there is a continuum of control level in these experimental settings. At one end, there are lab studies, which offer strong control over experimental variables. At the other end, there are field studies, in which user choices are observed through interactions with a recommender system in (pseudo) real-world conditions. It is important to note that the LBSN datasets mentioned earlier do not fall into this category, as the interactions they contain are not shaped by a recommender system but instead reflect user mobility and behavior.

In the context of recommender system research, user studies in laboratory settings often compare different interaction paradigms and are used to provide an early indication of model performance. In such cases, interaction data is collected under controlled conditions, specifically designed to evaluate the effects of alternative methods or models [60, 65, 169]. While such controlled experiments can reveal users' general opinions on recommendation algorithms and offer extra insights, for example, through think-aloud methods [105], their artificial tasks and environments limit how well the results generalize [80, 120]. In addition, the participant pool often shows sampling bias [90] and participants' attention towards the task may vary [107].

A smaller number of studies have collected observational data by capturing user behavior or preferences without the influence of a deployed recommender system [40, 42, 66]. These studies are rare but valuable, as they offer insight into users' natural decision-making processes. Still, they face the issue that users have no real "skin in the game," as they do not need to invest time or money in actually visiting the places they select. However, even these observational studies typically highlight only specific facets of tourist behavior or the capabilities of recommender systems.

Publicly available field studies in POI recommendation remain scarce because they require sending people on actual trips, which is prohibitively costly for most research institutions. In the following, we summarize the main limitations of datasets derived from user studies.

Pitfall 8: Limited users' sample impedes comparability with offline studies. User studies that evaluate research prototypes often have a limited number of participants, raising concerns about the scalability and generalization of the findings. Typical sample sizes are of the magnitude of 100-200 participants [8, 40, 42, 94]. Although these studies provide useful information, their small sample sizes limit the statistical power and robustness to train modern recommender models developed with respect to large-scale datasets.

Pitfall 9: Sampling bias skews participant demographics. While social media data introduces its own set of biases (cf. Pitfall 2 & 3), user studies are often even more demographically skewed due to the challenges of participant recruitment. In many cases, detailed demographic information is insufficient, and there is a well-documented bias in computer science research toward recruiting WEIRD participants (Western, Educated, Industrialized, Rich, and Democratic) due to their accessibility to researchers [90, 137]. This over-representation limits the generalization of the findings and raises concerns about the external validity of user studies in POI recommendation research.

Pitfall 10: Short-term observations limit temporal generalizability. Consistent with the limitations discussed above, user studies typically capture only short-term behavior, offering a snapshot of user interactions over a limited period. As a result, user studies are typically ill-suited for examining long-term tourism dynamics, such as evolving preferences, seasonal patterns, or repeated visit behavior [169]. Besides, while user studies have been widely adopted in traditional RSs, they remain largely unexplored in the POI recommendation domain [160], mostly due to the substantial cost of recruiting participants to use research prototypes in a non-laboratory setting. One of the few works that conducted user studies in this domain is [95], where the authors performed a user study to analyze the user-perceived novelty and interest for the recommendations.

5 Pitfalls in Algorithm Design

As discussed in Section 3.2, a wide variety of algorithms have been proposed for POI recommendation. In this section, we focus on the general challenges that most algorithms face in this recommendation domain.

Pitfall 11: Emphasis on accuracy metrics in algorithm design despite unreliable measurements. The incompleteness of LBSN datasets (cf. Pitfalls 3 and 4) presents a significant challenge for any algorithm aiming to perform well in practice. Since recorded information consists of voluntary check-ins, POI RSs cannot capture the full scope of user behavior, as it is possible in general recommendation systems, where datasets stem from interactions with a deployed recommender system [29]. This limited information forces models to learn only a partial view of real-world behavioral patterns, leading to unreliable outcome metrics. Moreover, many POI recommendation approaches formally optimize accuracy, despite the fact that the datasets themselves do not fully reflect user behavior and are inherently unreliable. Hence, even when accuracy is targeted, this contributes to the observation of lower estimated values, compared to other RS application areas [95]. This issue will also be discussed in the evaluation pitfalls part (Section 6), but here we emphasize the importance of a more cautious selection of the target accuracy metric, which is optimized in the model.

Pitfall 12: Rigidity and lack of adaptability of the algorithms reduce contextual relevance. Often, the recommendation algorithms address a formal problem that does not match exactly the application problem, as discussed in Section 2. A common consequence of that is the lack of a desirable property: dynamic adaptability to the evolving state of individual user intents and contextual factors [81]. In fact, tourists' general motivations for visiting a destination—whether leisure, business, or cultural exploration–may significantly influence the preferences. Moreover, while visiting a destination, it is common for users to include a variety of types of POIs in their itinerary, depending on their context, for instance, searching for a place to relax when tired. Hence, algorithms that lack a proper mechanism to adapt general personalization techniques to incorporate dynamic contextual information do not provide effective support to users and will be perceived as static and not responsive. To fill this gap, there is a need to develop systems that can respond to contextual changes and facilitate decision-making during the trip, not only before the trip starts. For example, in [109], the authors propose a framework that allows users to modify itineraries directly on a map to better adapt their personal interests, which in the long-term might only gradually shift, but can vary widely in the short run, or be very time-sensitive [91]. Advanced models that integrate multi-objective optimization, taking into account factors like user satisfaction, diversity, and contextual relevance, have shown promise in addressing these challenges [32]. However, the true problem remains to design an effective and unobtrusive context monitor component and to adapt recommendations to the obtained information.

Pitfall 13: Popularity bias in recommendation algorithms compromises fairness, diversity, and sustainable tourism promotion. Popularity bias is a common problem in RS, and is even more pronounced in POI RSs. The result is that well-known venues are disproportionately recommended, eclipsing less popular but potentially interesting POIs. While popular venues are worth seeing for many tourists and recommending them increases confidence in the RS, excessively focusing on these POIs is a clear limitation. In fact, recommending popular POIs might even exacerbate the massification of central venues, negatively impacting the quality of tourists' experiences and the well-being of the local community. Moreover, exposure bias, where certain items are presented more frequently to users regardless of their relevance, might only increase the severity of the problem [131]. Such biases not only limit the diversity of recommendations, but also raise concerns about fairness, as popular POIs may not be well suited to niche types of travelers [167]. Hence, addressing these challenges requires developing algorithms that can mitigate such biases while still providing accurate and personalized recommendations to all types of users.

Pitfall 14: Disregarding multistakeholder setting hinders fair and sustainable competition. Current algorithms, developed by academic research, often target the end-user, that is, the user who visits the recommended POIs, ignoring other stakeholders such as service providers or the local community. This focus can lead to unacceptable outcomes for the broader ecosystem [12], and can even produce unintended negative consequences for the target users of the system, such as overtourism and the lack of coverage of many small businesses. Some works have shown that it is possible to incorporate the interests of different multistakeholders while preserving the utility of the recommendations. Balakrishnan and Wörndl show that end users are sensitive to changes in the priorities of the recommendations and accept the inclusion of the interests of other stakeholders, as long as there is transparency and the recommendations maintain some utility level for them [8]. Another investigated solution proposes to tune a stakeholder's balance parameter, which can simultaneously improve user satisfaction while reducing overcrowding in popular locations [101]. That study highlights that a balanced promotion of sustainable POIs can align different stakeholder goals more effectively.

Pitfall 15: Lack of transparency and explainability compromises user trust and limits system adoption. As a multistakeholder environment, POI recommendations affect not only user decisions but also destinations' economies. In e-tourism, user trust is already weakened because users know they are customers of a commercial platform that, while serving their needs, also optimizes recommendations for its own goals. The platform's goals, namely balancing profit and user satisfaction, can sometimes diverge [98]. Relying on black-box models without providing any explanation to users may cause them not to trust the system and, therefore, to stop using it. Although recent approaches based on LLMs allow for more instinctive interaction with the system using natural language [14], the validity and truthfulness of the recommendations are not easy to verify. Hence, since visiting a POI affects both the time and the economic cost of the user, it is a fundamental requirement to verify the explainability of the models.

6 Pitfalls in the Evaluation

In this section, we discuss important challenges that should be faced when evaluating POI recommendation approaches; some of them are direct consequences of the previously discussed issues regarding data and algorithms.

Pitfall 16: Incomplete, and misaligned with user goals, performance metrics overlook trust and diversity. There is often a mismatch between the evaluation metrics used and the actual users' tasks and objectives. For example, while information retrieval metrics like precision or recall are adequate to measure the system's ability to retrieve POIs that match user queries, they will not effectively evaluate the system's ability to help users explore a destination's offerings or to plan a complete visit involving multiple POIs. Besides, in the tourism domain, it is more important to support, and therefore evaluate, the ability of the system to allow the discovery of novel POIs, rather than enabling look-up information search, which focuses on the retrieval of items, based on the user's explicit queries, which are sometimes impossible to formulate for the user [59, 142].

This is related to the multifaceted nature of the application domain: travelers want to capture the variety of destinations' offerings, hence they need help to reflect on, compare, and assess the recommended POIs. Therefore, POI RSs should also be evaluated on their ability to represent the diversity of a destination's attractions, not just their ability to identify individual POIs that match the user's tastes.

At the same time, a solid evaluation of a tourist RS should take into account the trust of the user towards the system, as recommendations in this domain can often be considered as advertisements, especially those present in operational systems, which are created by marketing departments. Moreover, users usually invest a large amount of money and time in arranging a holiday, so receiving poor recommendations has a significant negative impact. Therefore, it is essential that the systems are reliable to be useful in this domain [20].

Pitfall 17: Neglect of user-centric factors reduces usability, trust, and contextual relevance in real-world applications. Tourist online experience is closely related to the system interactivity, where ergonomics, usability, and accessibility are needed to provide useful systems for users [51, 144]. Therefore, evaluations must assess and measure these aspects. Without proper validation of these aspects, it is risky to launch a POI RS in an operational environment. In this context, incorporating explanations and transparency in user interactions should be another dimension to consider in the evaluation, especially when conducting user studies. This would increase trust in the system and make tourism experiences more adaptive and comprehensible [84].

Moreover, important aspects for tourists, such as distance to POIs, price, and user experience at the destination environment, are frequently omitted in both recommendation algorithms and their evaluations. Incorporating these factors into the evaluation is essential to improve user satisfaction and provide contextually relevant suggestions [111]. Herein, there is a disconnect between urban studies and personalized recommendations. Only a few studies bridged the experience of the urban environment of a certain destination or the surroundings of a POI and the travelers' individual preferences [42, 68].

Pitfall 18: Lack of user and item segmentation leads to generic recommendations that poorly match alternative traveler types. Distinguishing between different types of users (such as locals and tourists) or items (popular or niche) is also crucial in the evaluation stage of the system [133]. However, most current research does not consider users or items based on their behavior or characteristics, leading to generic recommendations that may not align with individual user needs or other multistakeholder goals.

In particular, user segmentation plays an important role in addressing the cold-start problem, which is common across many recommender system domains. New users and new items often lack sufficient interaction data, making it difficult to generate meaningful recommendations without leveraging additional information. Some approaches attempt to mitigate this problem by incorporating auxiliary information or taking advantage of group behaviors [145], but a universally effective solution is still missing. In fact, standard evaluation procedures do not explicitly consider these conditions as part of the holistic scenario where tourism RSs should continue providing *good* suggestions.

Pitfall 19: Inconsistent evaluation strategies with a few metrics hinder fair comparisons and obscure real-world effectiveness. The lack of standardized data splitting methods can lead to data leakage and unreliable performance evaluations [74]. Given the implications of random vs. temporal, user-based vs. global data splitting methods, it is essential for studies to clearly specify and motivate their data splitting strategies and other evaluation settings to ensure reproducibility and allow fair comparisons of alternative models [99, 130].

At the same time, most of the current literature prioritizes ranking accuracy metrics like Precision, Recall, and nDCG, and often ignores other critical beyond-accuracy dimensions, such as novelty, diversity, user satisfaction, and fairness [130, 161]. For example, algorithms trained to optimize accuracy may repeatedly suggest popular POIs, which are also liked by many tourists, but this also produces a lack of variety and ultimately user disengagement. These limited evaluation perspectives result in poorly diversified recommendations and users who are unaware of new or less popular venues, hence providing a biased picture of how the algorithm would perform in a real, multistakeholder environment.

Pitfall 20: Lack of reproducibility hinders scientific progress through comparative evaluation. Although it is becoming increasingly common to publish the source code of the core RSs algorithm in public repositories, the reproducibility of the published results is still problematic. Specifically, in the POI recommendation domain, Sánchez and Bellogín analyzed 310 contributions between 2011 and 2020 finding that only 13 provided the code of their models [130]. It should also be noted that sometimes even the bare code is not sufficient to determine whether the experiments conducted are significant and reproducible [33]. Hence, it is necessary to ensure that the code is easily adaptable and modifiable to experiment with other datasets or under different conditions, so that the overall evaluation remains comparable and fully reproducible.

7 A Research Agenda to Address the Pitfalls of Current POI Recommender Systems

As we have discussed in the article, the very practical target of building more useful POI recommender systems is still facing a number of diverse obstacles. In this section, we highlight a selection of research directions, which have already produced some results, and we believe can further contribute to overcoming the identified pitfalls. It is important to note that while the identified research directions are broad, they are not exclusive, and have been selected after the analysis of the cited literature and a discussion among the authors. We believe that moving the research along these directions can contribute to the implementation of more useful, realistic, and reliable POI recommender systems. The identified research directions are: multistakeholder design, context-awareness, data collection, trustworthiness design, novel interactions, and real-world evaluations.

Research Direction 1: Multistakeholder Design. As previously mentioned, a proper definition of the POI recommendation task requires the identification of multiple stakeholders' objectives [1]. Commercial platform owners are aware of this essential requirement and have therefore implemented recommendation techniques, which are only partially transparent to the academic community. These techniques balance platforms' profit with end user personalization, i.e., matching offers with observed tourists' preferences and intents [56]. Conversely, public organizations, i.e., national and regional DMOs, lack the competence and the technologies to implement an effective transition from the classical role of tourist information providers to the emerging need of becoming tourists' flow managers. In fact, today, an important focus and driver of the research on POI recommendation should be the wide goal of destination sustainability [97]. In this perspective, many DMOs are struggling when facing overtourism and are still failing to enhance the development of less popular and peripheral POIs and locations, which is a vital goal for these organizations. To help such organizations and build more useful POI recommender systems, research should focus on systems that, while still enabling the marketing of the main destination's POIs, can also temporally and spatially load-balance tourists in the entire region, and therefore support all destination service providers in a fairer way. Some research lines, focusing on multistakeholder RSs in tourism, have already targeted this goal [15, 101], but more research should be focused specifically on: optimization of effective policies to more uniformly distribute tourists [152], crowdedness prediction techniques [151], collaboration platforms for the promotion of small tourism services [92, 110], and nudging techniques for behavioral change [97]. This research direction can address pitfalls related to recommendation algorithm selection and optimization. Moreover, building multistakeholder systems requires the collection of more diverse behavior data, covering multiple segments of users and POIs, and overcoming the current research focus on biased data sets produced by social network usage. This is not an easy task, as privacy concerns are strong in this application domain. However, we believe that the joint satisfaction of multiple stakeholders' goals and the implementation of more transparent RSs can also promote a more trustworthy and extensive collection of behavioral data.

Research Direction 2: Context-Awareness. Notwithstanding the extensive state-of-the-art treatment of context dependency of tourist preferences and behavior [3, 11], a limited usage of context has been made so far in deployed POI recommender systems [127, 151]. Context is particularly important during trip planning and re-planning, as the main contextual dimensions that have been considered are dynamic: weather [24], user's mood and personality [23, 75], and group composition [39, 104]. However, these dimensions have often been used independently and confined to a second tier of importance, i.e., the personalization model was still mainly driven by item-related preferences extracted from observed behaviors. We need to elaborate a more comprehensive view and better use the multiple contextual factors that influence tourists' preferences and behaviors, which can even be more important than specific individual preferences [2, 10, 25].

Although the physical context factors of location, time, and weather are easy to measure and use, the research should focus more on the analysis of traveler intentions and mood [23, 39, 68]. We need more research on the extraction of the true tourists' intents from observational data, in order to better understand what experiences tourists are looking for, and at the same time, we must develop better solutions to extract context and intent information from the user–system interaction, i.e., by mining text content or by interacting with the user with natural language interfaces. To this end, we believe some recent research works are of interest; for example, real-time analysis of tourists' discussions [76], behavioral clustering of observed tourists' behaviors [96], identification of the best context for a visit, or methods to detect context importance [10].

The other aspect of context is the effect of specific urban spaces on travelers. Herein, interdisciplinary findings from urban sciences quantifying the *"vibe"* and other characteristics of tourism destinations and areas surrounding individual POIs need to be considered in the recommendation models [44, 45, 82, 135].

Hence, context management can contribute to addressing pitfalls related to data and algorithms. Novel and more comprehensive context-dependent preference, behavior data, and environment should be unobtrusively collected and analyzed, and recommendation algorithms that prioritize the highly dynamic contextual state of the tourist should be integrated in the RS.

Research Direction 3: Data Collection. The pitfalls related to data acquisition and algorithms show that state-of-the-art POI RSs are systematically constrained by preference models learned from limited and biased observational logs of user behavior. Commercial stakeholders are in a better position, as they can rely on more complete and massive logs of real service purchases and tourists' movements at destinations. However, the scope of their recorded interactions is often limited to a single type of POIs or services, mostly accommodation or event bookings. DMOs are better positioned to access various data sources: transportation, hotel reservations, culturally related tickets, or information requests and dialogues with tourist information points. The analysis of these data sets could be challenged by their heterogeneity and size. Hence, developing techniques and frameworks that facilitate these types of analysis is an important research direction. Currently, marketing destination companies have started to analyze these data sources, with the help of commercial enterprises offering customer relationship management and AI-based data mining tools, but, at the same time, this data is very rarely shared with academic research [152]. Therefore, the development and dissemination of these datasets is of primary importance, and specific actions must be identified yo promote and enable data-sharing activities. National and international initiatives must be started to support even smaller destinations in planning effective data analysis strategies. We also believe in the potential to generate pseudo-synthetic data that combines real observations with debiasing and data augmentation techniques [49, 100]. These data sets may be used to explore new opportunities for personalization, nudging, and promotion, without jeopardizing the marketing activity of a destination and raising privacy concerns (Pitfall 6).

Research Direction 4: Trustworthiness Design. A major limitation of current POI recommender systems is their lack of transparency and, more broadly, trustworthiness [158]: tourists are often unsure of the quality and appropriateness of the received recommendations, hence they often opt for the more popular attractions, which are easier to recommend and to be recognized as valuable targets. In fact, visiting a POI often involves time and money, and poor recommendations can negatively affect tourists' experience. This explains why tourists tend to favor popular places and invest time reading reviews, which signal social validation of these POIs. These simple approaches to decision-making have been shown to increase user confidence in the choice [70]. However, we believe that confidence and trust should also be achieved when less mainstream recommendations are generated by new types of trustworthy RSs. Tourists must be confident that the recommendations are genuinely constructed to match their needs and preferences. Although such aspects are now required under European legislation, e.g., through the Digital Services Act, their actual availability and utility remains limited. To address these goals, we need algorithms and testing approaches that can transparently ensure that the involved stakeholders are satisfied with the generated recommendations, i.e., the multiple criteria that the stakeholders have independently defined and should be jointly maximized by the RS [1]. Moreover, consumers should be made aware of such a mediation of competing goals by explicit messages of the GUI and the available explanations, so that they can make more sensible decisions and eventually even opt out of an untrustworthy RS. This requires a non-trivial level of transparency, adapted to be understandable and usable by various types of tourists. This research line is relevant for many pitfalls discussed in the article. Clearly, the recommendation algorithms must be modified to become really multistakeholder and trustworthy. But, a trustworthy design of the recommender system requires also novel evaluation metrics and procedures, which can be primarily performed by audit institutions, but also by client applications, on behalf of the end user, to decide whether to rely or not on the provided recommendation service.

Research Direction 5: Novel Interactions. Nowadays, before and during a trip, tourists rely on a multitude of specialized applications, each dealing with specific components and aspects of the trip, such as accommodations, transport, restaurants, or local attractions. While these tools offer useful services, no application can integrate all of them at once, and hence they require users to manually switch from app to app and compare, integrate, and judge information collected from these various sources. In fact, this information fragmentation makes it difficult to assess and enable the smooth interconnection of elements of the trip. For example, a recommended attraction may not be conveniently reached from the user's hotel, with available transportation options, or due to the user's time constraints. These difficulties can result in suboptimal itineraries that do not take into account the general needs and preferences of the traveler. To meet these challenges, the next generation of POI recommender systems must take a more holistic approach to trip planning, still prioritizing usability to completeness. This involves integrating various trip components, such as transportation, accommodations, activities, and restaurants, into a coherent framework that takes into account the user's context, preferences, and constraints. Using comprehensive data sources and advanced algorithms, these systems should be able to generate personalized end-to-end itineraries that increase user satisfaction and streamline the trip planning process. A technical approach to glue the output of multiple systems comes from the usage of LLMs in conversational systems [28, 37, 38, 58, 103]. The future generation of RSs may be perceived as agents that, relying on multiple information and recommendation components, offer the necessary middleware to integrate these components and can be used as a unique access point to trip planning. Some recent works on AI supporting group discussion go exactly in this direction [93, 122]. The research we propose on new forms of interaction can address many pitfalls, but primarily the mismatch between user needs and current recommendation algorithms, which are often technology-driven and precision-focused, while paying too little attention to the best practices of information search and discovery that tourists naturally follow.

Research Direction 6: Real-World Evaluations. Traditional evaluation methods for tourism RSs often rely on the same offline metrics that are used in other application domains and are easy to measure with the available data sets. Although certainly useful, these metrics only score system performance on the basis of historical user-system interaction data. However, they cannot assess the overall quality of the recommendations, which is influenced by multiple factors, such as ephemeral preferences, contextual factors, or the specific individual dynamics of the decision-making process. To fill this gap, Living Labs have emerged as a promising approach to real-world evaluation of tourism recommender systems. Living labs are user-centered innovation ecosystems that facilitate co-creation and experimentation in real-world environments [9]. In living labs, diverse stakeholders (recall Figure 1) can design, test, and create different solutions that benefit multiple stakeholders. The collaborative nature of living labs, thanks to continuous feedback and adaptation, ensures that the RS remains responsible for the changing needs and preferences of the users. In general, we need to experiment with new forms of online evaluations where it could be simpler and more efficient to run targeted tests of specific system properties.

Notwithstanding the potential increase in the efficiency of online tests, the need to perform offline experiments remains. We believe that the recent practice of performing calibrated simulations of tourist behaviors can enhance the ability of offline evaluations to select good system candidates to be tested online [101, 104]. Hence, in conclusion, this research line can produce a significant impact in addressing the pitfalls related to the evaluation of the RSs, but can also foster the development of more significant recommendation services.

8 Lessons Learned and Conclusions

We summarize the identified relationship between pitfalls and research directions in Table 2. Here we distinguish the match between pitfalls and directions in primary and secondary, to better separate what we believe is the primary and secondary way to approach a potential solution of the pitfall.

In general, closing the gap between research and practice through Real-world Evaluations (Research Direction 6) emerges as the most important area of improvement. It is the primary direction for six pitfalls and a secondary direction seven times. Multistakeholder Design, Data Collection, and Novel Interactions also address a large number of pitfalls. However, novel interactions are the primary direction only for Pitfall 5 (Platform Biases). This suggests that new ways to support travelers are necessary to address a broad range of issues, even if they are not always the primary remedy.

Interestingly, almost all directions address all categories of pitfalls, except Research Direction 2 (Context-Awareness), which does not resolve any evaluation-related issues. As expected, improvements in data collection practices (Pitfall 3) mainly target pitfalls related to datasets, but also have indirect effects on evaluation practices. For example, one of the most serious problems in POI recommendation–Pitfall 13 (Popularity Bias)–has its root in the nature of available data sets, and we posit that algorithmic design would be different if richer and realistic datasets were available. For this reason, we believe that several pitfalls may require multiple research directions working in coordination to improve current practices.

Some necessary improvements lie outside the scope of our proposed research agenda. For example, improving reproducibility depends not only on the methods and data used but also on establishing high publication standards for documenting procedures and sharing code and data [67]. Additional steps, such as selecting fair baselines and conducting reproducibility and replication studies to monitor algorithmic progress [19, 33], are also needed.

To conclude, POI recommendation is a research area that has received much attention, not only due to its apparent simplicity but also for its clear usefulness. However, we have claimed that the majority of the proposed directions have failed to address the important needs and demands of the stakeholders involved. We have therefore initiated the paper with a discussion of the POI recommendation problem, before surveying state-of-the-art algorithms and evaluation approaches. We then identified important characteristics and challenges that need to be addressed to produce more useful and truly deployable systems. We have structured our analysis along three axes: datasets, algorithms, and evaluation approaches.

By means of this analysis, we have claimed that while LBSNs are the main source of information for POI RSs, they are composed of sparse, outdated, and highly biased data. For this reason, we argue that it is necessary to complement this information and consider more recent, complete, and even complementary data sources, such as those collected by DMOs and those related to weather, events, or blogs. Regarding algorithms, we have found evidence that most of them are focused on the optimization of user-related precision metrics, ignoring other stakeholders and beyond accuracy dimensions, such as novelty, fairness, or diversity. Moreover, current

Table 2. **Pitfalls mapped to potential research directions.** "✓✓" indicates the column as a primary research direction to the pitfall (row), while "✓" represents a secondary, i.e., more indirect connection.

Cat.	**Pitfall**	**Research direction**					
		Multi-stake-holder Design	Context-Awareness	Data-Collection	Trustwor-thiness Design	Novel-Interactions	Real-World Evalua-tions
LBSN Data	(1) Outdated Data	✓	✓	✓✓		✓	✓
	(2) Data Mismatches			✓			✓✓
	(3) Incomplete Data	✓	✓	✓✓		✓	
	(4) POI Category Bias	✓✓		✓		✓	
	(5) Platform Bias	✓	✓	✓		✓✓	
	(6) Data Collection Ethics			✓	✓✓		
	(7) Limited User Context		✓✓	✓	✓	✓	
User Study Data	(8) Limited Scalability			✓		✓	✓✓
	(9) Sampling Bias	✓		✓✓		✓	✓
	(10) Short-term Observations		✓	✓			✓✓
Algo-rithms	(11) Accuracy Emphasis	✓			✓	✓	✓✓
	(12) Algorithmic Rigidity	✓	✓✓		✓	✓	
	(13) Popularity Bias	✓		✓✓	✓		✓
	(14) Lack of Multistakeholder Aspects	✓✓			✓		✓
	(15) Lack of Transparency				✓✓	✓	✓
Eval-uation	(16) Incomplete Measurements	✓			✓✓	✓	✓
	(17) Neglect of User-centric Factors				✓	✓	✓✓
	(18) Lack of User and Item Segmentation	✓✓		✓			✓
	(19) Inconsistent Evaluation Strategies	✓		✓			✓✓
	(20) Reproducibility Issues			✓			✓✓

solutions lack adequate support for in-trip context-dependent re-planning. In fact, it is not enough to recommend relevant items, but also to adapt the specific user profile and consider other factors such as distance, price, and popularity of the POI, so that the user is aware of the variety of POIs that can be visited. Regarding the evaluation dimension, we have evidenced that more interaction with users must be collected to better train and test the designed solutions.

Finally, we have proposed a structured research agenda that, starting from the identified issues, introduces important directions for future work. These potential future lines are related to multistakeholder design, context-awareness, data collection, trustworthiness, novel interactions, and real-world evaluation. In particular, we consider the creation of datasets that incorporate not only data from LBSNs but also information from other stakeholders and contexts as vital. Moreover, the addressed recommendation problem must take into account the interests of all relevant stakeholders in the tourism domain. Because of this variety of goals and data, newer models that take into account a wider range of destination data and produce more sustainable and trustworthy recommendations, for the benefit of the entire ecosystem of the tourism economy, are a primary target for future research.

It is worth concluding that our contribution in this reflection article can provide neither a complete analysis of the research gaps nor a definitive program of research on POI RSs. We aimed at offering to researchers and practitioners a starting point for deepening the reflection on this subject, and to contribute to the progress of this important applied research area, which can benefit the social and economic development of many global regions, hopefully, in a more sustainable and fair way.

Acknowledgments

This work was in part supported by grant PID2022-139131NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe."

References

[1] Himan Abdollahpouri and Robin Burke. 2022. Multistakeholder Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 647–677. doi:10.1007/978-1-0716-2197-4_17

[2] Gediminas Adomavicius, Konstantin Bauman, Alexander Tuzhilin, and Moshe Unger. 2022. Context-Aware Recommender Systems: From Foundations to Recent Developments. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 211–250. doi:10.1007/978-1-0716-2197-4_6

[3] Gediminas Adomavicius, Bamshad Mobasher, Francesco Ricci, and Alexander Tuzhilin. 2011. Context-Aware Recommender Systems. *AI Mag.* 32, 3 (2011), 67–80.

[4] Taras Agryzkov, Pablo Martí, Leandro Tortosa, and José-Francisco Vicent. 2017. Measuring urban activities using Foursquare data and network analysis: a case study of Murcia (Spain). *International Journal of Geographical Information Science* 31, 1 (May 2017), 100–121. doi:10.1080/13658816.2016.1188931

[5] Fatma S. Alrayes, Alia I. Abdelmoty, Waleed El-Geresy, and George Theodorakopoulos. 2020. Modelling perceived risks to personal privacy from location disclosure on online social networks. *International Journal of Geographical Information Science* 34, 1 (2020), 150–176. doi:10.1080/13658816.2019.1654109

[6] Patrícia Alves, Helena Martins, Pedro M. Saraiva, João Carneiro, Paulo Novais, and Goreti Marreiros. 2023. Group recommender systems for tourism: how does personality predict preferences for attractions, travel motivations, preferences and concerns? *User Modeling and User-Adapted Interaction* 33, 5 (2023), 1141–1210. doi:10.1007/S11257-023-09361-2

[7] Victor Anthony Arrascue Ayala, Kemal Cagin Gülsen, Anas Alzogbi, Michael Färber, Marco Muñiz, and Georg Lausen. 2017. A Delay-Robust Touristic Plan Recommendation Using Real-World Public Transportation Information. In *2nd RecSys Workshop on Recommenders in Tourism (CEUR Workshop Proceedings, Vol. 1906)*. CEUR-WS.org, 9–17.

[8] Gokulakrishnan Balakrishnan and Wolfgang Wörndl. 2021. Multistakeholder Recommender Systems in Tourism. In *RecSys Workshop on Recommenders in Tourism (CEUR Workshop Proceedings, Vol. 2974)*. CEUR-WS.org, 39–53.

[9] Pieter Ballon and Dimitri Schuurman. 2015. Living labs: concepts, tools and cases. *Info* 17, 4 (2015). doi:10.2139/ssrn.2642754

[10] Linas Baltrunas, Bernd Ludwig, Stefan Peer, and Francesco Ricci. 2012. Context relevance assessment and exploitation in mobile recommender systems. *Personal and Ubiquitous Computing* 16, 5 (2012), 507–526. doi:10.1007/S00779-011-0417-X

[11] Linas Baltrunas, Bernd Ludwig, and Francesco Ricci. 2011. Context relevance assessment for recommender systems. In *16th International Conference on Intelligent User Interfaces (IUI'11)*. ACM, 287–290. doi:10.1145/1943403.1943447

[12] Ashmi Banerjee, Paromita Banik, and Wolfgang Wörndl. 2024. A review on individual and multistakeholder fairness in tourism recommender systems. *Frontiers Big Data* 6 (May 2024). doi:10.3389/fdata.2023.1168692

[13] Ashmi Banerjee, Gourab K. Patro, Linus W. Dietz, and Abhijnan Chakraborty. 2020. Analyzing 'Near Me' Services: Potential for Exposure Bias in Location-based Retrieval. In *International Workshop on Fair and Interpretable Learning Algorithms (FILA'20)*. IEEE. doi:10.1109/bigdata50022.2020.9378476

[14] Ashmi Banerjee, Adithi Satish, Fitri Nur Aisyah, Wolfgang Wörndl, and Yashar Deldjoo. 2025. SynthTRIPs: A Knowledge-Grounded Framework for Benchmark Data Generation for Personalized Tourism Recommenders. In *48th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2025) *(SIGIR'25)*. ACM, 3743–3752. doi:10.1145/3726302.3730321

[15] Paromita Banik, Ashmi Banerjee, and Wolfgang Wörndl. 2023. Understanding User Perspectives on Sustainability and Fairness in Tourism Recommender Systems. In *31st ACM Conference on User Modeling, Adaptation and Personalization*. ACM, New York, NY, USA, 241–248. doi:10.1145/3563359.3597442

[16] Jie Bao, Yu Zheng, David Wilkie, and Mohamed Mokbel. 2015. Recommendations in location-based social networks: a survey. *GeoInformatica* 19, 3 (Feb. 2015), 525–565. doi:10.1007/s10707-014-0220-8

[17] Filipe Batista e Silva, Mario Alberto Marín Herrera, Konštantín Rosina, Ricardo Ribeiro Barranco, Sérgio Freire, and Marcello Schiavina. 2018. Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management* 68 (Oct. 2018), 101–115. doi:10.1016/j.tourman.2018.02.020

[18] Howard Bauchner, Robert M. Golub, and Phil B. Fontanarosa. 2016. Data Sharing: An Ethical and Scientific Imperative. *JAMA* 315, 12 (March 2016), 1238. doi:10.1001/jama.2016.2420

[19] Alejandro Bellogín and Alan Said. 2021. Improving accountability in recommender systems research through reproducibility. *User Modeling and User-Adapted Interaction* 31, 5 (Nov. 2021), 941–977. doi:10.1007/s11257-021-09302-x

[20] Kassegn Berhanu and Sahil Raj. 2020. The trustworthiness of travel and tourism information sources of social media: perspectives of international tourists visiting Ethiopia. *Heliyon* 6, 3 (2020), e03439. doi:10.1016/j.heliyon.2020.e03439

[21] Veronika Bogina, Tsvi Kuflik, Dietmar Jannach, Mária Bieliková, Michal Kompan, and Christoph Trattner. 2022. Considering temporal aspects in recommender systems: a survey. *User Modeling and User-Adapted Interaction* 33, 1 (July 2022), 81–119. doi:10.1007/s11257-022-09335-w

[22] Samantha Boom, Jelmer Weijschede, Frans Melissen, Ko Koens, and Igor Mayer and. 2021. Identifying stakeholder perspectives and worldviews on sustainable urban tourism development using a Q-sort methodology. *Current Issues in Tourism* 24, 4 (2021), 520–535. doi:10.1080/13683500.2020.1722076

[23] Matthias Braunhofer, Mehdi Elahi, and Francesco Ricci. 2014. *User Personality and the New User Problem in a Context-Aware Point of Interest Recommender System*. Springer, 537–549. doi:10.1007/978-3-319-14343-9_39

[24] Matthias Braunhofer, Mehdi Elahi, Francesco Ricci, and Thomas Schievenin. 2014. Context-Aware Points of Interest Suggestion with Dynamic Weather Data Management. In *Information and Communication Technologies in Tourism 2014* (2013). Springer, 87–100. doi:10.1007/978-3-319-03973-2_7

[25] Matthias Braunhofer and Francesco Ricci. 2016. Contextual Information Elicitation in Travel Recommender Systems. In *Information and Communication Technologies in Tourism 2016* (2016). Springer, 579–592. doi:10.1007/978-3-319-28231-2_42

[26] Igo Ramalho Brilhante, José Antônio Fernandes de Macêdo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. 2013. Where shall we go today?: planning touristic tours with tripbuilder. In *22nd ACM International Conference on Information & Knowledge Management (CIKM'13)*. ACM, 757–762. doi:10.1145/2505515.2505643

[27] Robin D. Burke and Maryam Ramezani. 2011. Recommender Systems Handbook. Springer, Boston, MA, USA, Chapter Matching Recommendation Technologies and Domains, 367–386. doi:10.1007/978-0-387-85820-3_11

[28] Mark Anthony Camilleri and Ciro Troise. 2023. Chatbot recommender systems in tourism: A systematic review and a benefit-cost analysis. In *8th International Conference on Machine Learning Technologies*. ACM, New York, NY, USA, 151–156. doi:10.1145/3589883.3589906

[29] Rocío Cañamares and Pablo Castells. 2018. Should I Follow the Crowd?: A Probabilistic Analysis of the Effectiveness of Popularity in Recommender Systems. In *41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 415–424. doi:10.1145/3209978.3210014

[30] Buru Chang, Yookyung Koh, Donghyeon Park, and Jaewoo Kang. 2020. Content-Aware Successive Point-of-Interest Recommendation. In *SIAM International Conference on Data Mining*, Carlotta Demeniconi and Nitesh V. Chawla (Eds.). SIAM, 100–108. doi:10.1137/1.9781611976236.12

[31] Dawei Chen, Cheng Soon Ong, and Lexing Xie. 2016. Learning Points and Routes to Recommend Trajectories. In *25th ACM International on Conference on Information and Knowledge Management (CIKM'16)*. ACM, New York, NY, USA, 2227–2232. doi:10.1145/2983323.2983672

[32] Lei Chen, Guixiang Zhu, Weichao Liang, and Youquan Wang. 2023. Multi-objective reinforcement learning approach for trip recommendation. *Expert Systems with Applications* 226 (2023), 120145. doi:10.1016/J.ESWA.2023.120145

[33] Maurizio Ferrari Dacrema, Simone Boglio, Paolo Cremonesi, and Dietmar Jannach. 2021. A Troubling Analysis of Reproducibility and Progress in Recommender Systems Research. *ACM Transactions on Information Systems* 39, 2 (April 2021), 1–49. doi:10.1145/3434185

[34] Savvina Daniil, Manel Slokom, Mirjam Cuper, Cynthia Liem, Jacco van Ossenbruggen, and Laura Hollink. 2025. On the challenges of studying bias in Recommender Systems: The effect of data characteristics and algorithm configuration. *Information Retrieval Research* 1, 1 (Feb. 2025), 3–27. doi:10.54195/irrj.19607

[35] Manoj Reddy Dareddy. 2016. Challenges in Recommender Systems for Tourism. In *Workshop on Recommenders in Tourism (CEUR Workshop Proceedings, Vol. 1685)*, Daniel R. Fesenmaier, Tsvi Kuflik, and Julia Neidhardt (Eds.). CEUR-WS.org, 59–61. https://ceur-ws.org/Vol-1685/paper11.pdf

[36] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports* 3, 1 (March 2013). doi:10.1038/srep01376

[37] Yashar Deldjoo, Zhankui He, Julian J. McAuley, Anton Korikov, Scott Sanner, Arnau Ramisa, René Vidal, Maheswaran Sathiamoorthy, Atoosa Kasrizadeh, Silvia Milano, and Francesco Ricci. 2024. Recommendation with Generative Models. *CoRR* abs/2409.15173 (2024). doi:10.48550/ARXIV.2409.15173

[38] Amra Delic, Hanif Emamgholizadeh, Thuy Ngoc Nguyen, and Francesco Ricci. 2024. CHARM: a Group Decision Making Support Chatbot. In *29th International Conference on Intelligent User Interfaces*. ACM, 7–10. doi:10.1145/3640544.3645220

[39] Amra Delic, Hanif Emamgholizadeh, Francesco Ricci, and Judith Masthoff. 2024. Supporting Group Decision-Making: Insights from a Focus Group Study. In *32nd ACM Conference on User Modeling, Adaptation and Personalization*. ACM, 301–306. doi:10.1145/3627043.3659538

[40] Amra Delic, Julia Neidhardt, Thuy Ngoc Nguyen, and Francesco Ricci. 2018. An observational user study for group recommender systems in the tourism domain. *Information Technology & Tourism* 19, 1-4 (2018), 87–116. doi:10.1007/S40558-018-0106-Y

[41] Linus W. Dietz. 2023. *Data-driven Destination Recommender Systems*. Ph. D. Dissertation. Technical University of Munich, Munich, Germany.

[42] Linus W. Dietz, Sanja Scepanovic, Ke Zhou, and Daniele Quercia. 2024. Exploratory Analysis of Recommending Urban Parks for Health-Promoting Activities. In *18th ACM Conference on Recommender Systems*. ACM, 1131–1135. doi:10.1145/3640457.3691712

[43] Linus W. Dietz, Avradip Sen, Rinita Roy, and Wolfgang Wörndl. 2020. Mining Trips from Location-based Social Networks for Clustering Travelers and Destinations. *Information Technology & Tourism* 22, 1 (March 2020), 131–166. doi:10.1007/s40558-020-00170-6

[44] Linus W. Dietz, Mete Sertkan, Saadi Myftija, Sameera Thimbiri Palage, Julia Neidhardt, and Wolfgang Wörndl. 2022. A Comparative Study of Data-driven Models for Travel Destination Characterization. *Frontiers in Big Data* 5 (April 2022). doi:10.3389/fdata.2022.829939

[45] Linus W. Dietz, Sanja Šćepanović, Ke Zhou, André Felipe Zanella, and Daniele Quercia. 2025. Understanding the potential of urban parks to promote well-being. *Nature Cities* 2, 12 (Nov. 2025), 1205–1216. doi:10.1038/s44284-025-00345-4

[46] Eric A. Drott. 2018. Music as a Technology of Surveillance. *Journal of the Society for American Music* 12, 3 (July 2018), 233–267. doi:10.1017/s1752196318000196

[47] Krittika D'Silva, Kasthuri Jayarajah, Anastasios Noulas, Cecilia Mascolo, and Archan Misra. 2018. The Role of Urban Mobility in Retail Business Survival. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3, Article 100 (Sept. 2018), 22 pages. doi:10.1145/3264910

[48] José Luis Durán-Román, Pablo Juan Cárdenas-García, and Juan Ignacio Pulido-Fernández. 2021. Tourists' willingness to pay to improve sustainability and experience at destination. *Journal of Destination Marketing & Management* 19 (2021), 100540. doi:10.1016/j.jdmm.2020.100540

[49] Hanif Emamgholizadeh, Amra Delic, and Francesco Ricci. 2024. Predicting Group Choices from Group Profiles. *ACM Transactions on Interactive Intelligent Systems* 14, 1 (2024), 7:1–7:27. doi:10.1145/3639710

[50] European Parliament and Council of the European Union. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council*. OJ L 119, 4.5.2016, p. 1–88. https://data.europa.eu/eli/reg/2016/679/oj

[51] Francisco V. Cipolla Ficarra. 2010. Human-Computer Interaction, Tourism and Cultural Heritage. In *1st International Workshop Human-Computer Interaction, Tourism and Cultural Heritage (Lecture Notes in Computer Science, Vol. 6529)*. Springer, 39–50. doi:10.1007/978-3-642-18348-5_5

[52] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2015. Content-Aware Point of Interest Recommendation on Location-Based Social Networks. In *29th AAAI Conference on Artificial Intelligence*, Blai Bonet and Sven Koenig (Eds.). AAAI Press, 1721–1727. doi:10.1609/AAAI.V29I1.9462

[53] Michael D. Garber, Katie Labgold, and Michael R. Kramer. 2022. On selection bias in comparison measures of smartphone-generated population mobility: an illustration of no-bias conditions with a commercial data source. *Annals of Epidemiology* 70 (June 2022), 16–22. doi:10.1016/j.annepidem.2022.03.003

[54] Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, Grammati E. Pantziou, and Nikolaos Vathis. 2015. Heuristics for the time dependent team orienteering problem: Application to tourist route planning. *Computers & Operations Research and their Application to Problems of World Concern* 62 (2015), 36–50. doi:10.1016/J.COR.2015.03.016

[55] Juergen Gnoth. 1997. Tourism motivation and expectation formation. *Annals of Tourism research* 24, 2 (1997), 283–304. doi:10.1016/s0160-7383(97)80002-3

[56] Dmitri Goldenberg and Javier Albert. 2022. Personalizing Benefits Allocation Without Spending Money: Utilizing Uplift Modeling in a Budget Constrained Setup. In *16th ACM Conference on Recommender Systems*. ACM, New York, NY, USA, 464–465. doi:10.1145/3523227.3547381

[57] Dmitri Goldenberg, Sarai Mizrachi, Adam Horowitz, Ioannis Kangas, Or Levkovich, Alessandro Mozzato, Maud Schwoerer, Michele Ferretti, Panagiotis Korvesis, and Lucas Bernardi. 2021. I Know What You Did Next Summer: Challenges in Travel Destinations Recommendation. In *RecSys Workshop on Recommenders in Tourism (CEUR Workshop Proceedings, Vol. 2974)*. CEUR-WS.org, 7–22.

[58] Elena L. González-Sanz, Iván Cantador, and Alejandro Bellogín. 2025. LLM-based Generation of Personalized, Context-aware City Tourist Itineraries: A User Study with GPT Trip Planner. In *Workshop on Recommenders in Tourism (CEUR Workshop Proceedings, Vol. 4052)*, Julia Neidhardt, Tsvi Kuflik, Amit Livne, Markus Zanker, and Wolfgang Wörndl (Eds.). CEUR-WS.org, 68–73. https://ceur-

ws.org/Vol-4052/paper6.pdf
[59] Ulrike Gretzel. 2011. Intelligent systems in tourism: A social science perspective. *Annals of tourism research* 38, 3 (2011), 757–779. doi:10.1016/j.annals.2011.04.014
[60] Asela Gunawardana, Guy Shani, and Sivan Yogev. 2022. *Evaluating Recommender Systems.* Springer, 547–601. doi:10.1007/978-1-0716-2197-4_15
[61] Ibrahim Al Hazwani, Tiantian Luo, Oana Inel, Francesco Ricci, Mennatallah El-Assady, and Jürgen Bernard. 2024. ScrollyPOI: A Narrative-Driven Interactive Recommender System for Points-of-Interest Exploration and Explainability. In *32nd ACM Conference on User Modeling, Adaptation and Personalization.* ACM. doi:10.1145/3631700.3665183
[62] Juan M. Hernández, Andrei P. Kirilenko, and Svetlana Stepchenkova. 2018. Network approach to tourist segmentation via user generated content. *Annals of Tourism Research* 73 (2018), 35–47. doi:10.1016/j.annals.2018.09.002
[63] Daniel Herzog, Linus W. Dietz, and Wolfgang Wörndl. 2019. Tourist Trip Recommendations – Foundations, State of the Art and Challenges. In *Personalized Human-Computer Interaction*, Miriam Augstein, Eelco Herder, and Wolfgang Wörndl (Eds.). de Gruyter Oldenbourg, Berlin, Germany, 159–182. doi:10.1515/9783110552485-006
[64] Daniel Herzog and Wolfgang Wörndl. 2019. User-centered evaluation of strategies for recommending sequences of points of interest to groups. In *13th ACM Conference on Recommender Systems.* ACM, 96–100. doi:10.1145/3298689.3346988
[65] Daniel Andreas Herzog. 2020. *A User-Centered Approach to Solving the Tourist Trip Design Problem for Individuals and Groups.* Ph. D. Dissertation. Technische Universität München.
[66] Katharina Hofschen, David Massimo, and Francesco Ricci. 2023. Expected and Experienced Utility of Points of Interest in Tourism Recommender Systems. In *31st ACM Conference on User Modeling, Adaptation and Personalization.* ACM, 50–55. doi:10.1145/3563359.3597405
[67] Matthew Hutson. 2018. Artificial intelligence faces reproducibility crisis. *Science* 359, 6377 (2018), 725–726. doi:10.1126/science.359.6377.725
[68] Katrin Hänsel, Luca Maria Aiello, Daniele Quercia, Rossano Schifanella, Krisztian Zsolt Varga, Linus W. Dietz, and Marios Constantinides. 2025. The Experience of Running: Recommending Routes Using Sensory Mapping in Urban Environments. *International Journal of Human-Computer Studies* 201 (2025). doi:10.1016/j.ijhcs.2025.103512
[69] Sureerut Inmor, Khahan Na-Nan, Kanakarn Phanniphong, Nutt Jaturat, and Matěj Kůstka. 2025. The role of smart green tourism technologies in shaping tourist intentions: Balancing authenticity and sustainability in natural tourism. *Environmental Challenges* 19 (2025), 101171. doi:10.1016/j.envc.2025.101171
[70] Anthony Jameson, Martijn C. Willemsen, and Alexander Felfernig. 2022. Individual and Group Decision Making and Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 789–832. doi:10.1007/978-1-0716-2197-4_21
[71] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2022. A Survey on Conversational Recommender Systems. *Comput. Surveys* 54, 5 (2022), 105:1–105:36. doi:10.1145/3453154
[72] Dietmar Jannach and Markus Zanker. 2020. Interactive and Context-Aware Systems in Tourism. In *Handbook of e-Tourism*, Zheng Xiang, Matthias Fuchs, Ulrike Gretzel, and Wolfram Höpken (Eds.). Springer, Cham, 1–22. doi:10.1007/978-3-030-05324-6_125-1
[73] Priit Järv. 2017. Extracting Human Mobility Data from Geo-tagged Photos. In *1st ACM SIGSPATIAL Workshop on Prediction of Human Mobility.* ACM, 4:1–4:7. doi:10.1145/3152341.3152346
[74] Yitong Ji, Aixin Sun, Jie Zhang, and Chenliang Li. 2023. A Critical Study on Data Leakage in Recommender System Offline Evaluation. *ACM Transactions on Information Systems* 41, 3 (2023), 75:1–75:27. doi:10.1145/3569930
[75] Marius Kaminskas and Francesco Ricci. 2017. Emotion-Based Matching of Music to Places. In *Emotions and Personality in Personalized Services - Models, Evaluation and Applications*, Marko Tkalcic, Berardina De Carolis, Marco de Gemmis, Ante Odic, and Andrej Kosir (Eds.). Springer, 287–310. doi:10.1007/978-3-319-31413-6_14
[76] Esma Karahodža, Amra Delić, and Francesco Ricci. 2025. Conceptual Framework for Group Dynamics Modeling from Group Chat Interactions. In *33rd ACM Conference on User Modeling, Adaptation and Personalization.* ACM, New York City, NY, USA, 5. doi:10.1145/3708319.3733682
[77] Md. Ibrahim Khalil and Mst. Khadijatul Kobra. 2021. Monitoring overtourism: Destination Management Systems as a way forward. *Bangladesh Journal of Public Administration* 29, 2 (2021), 79–87. doi:10.36609/bjpa.v29i2.211
[78] Minkyoung Kim, Lexing Xie, and Peter Christen. 2021. Event Diffusion Patterns in Social Media. *International AAAI Conference on Web and Social Media* 6, 1 (Aug. 2021), 178–185. doi:10.1609/icwsm.v6i1.14248
[79] Emily Kirby, Alexander Bernier, Roderic Guigó, Barbara Wold, Fabiana Arzuaga, Mayumi Kusunose, Ma'n Zawati, and Bartha M. Knoppers. 2024. Data sharing ethics toolkit: The Human Cell Atlas. *Nature Communications* 15, 1 (Nov. 2024). doi:10.1038/s41467-024-54300-3
[80] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. *Evaluating Recommender Systems with User Experiments.* Springer, Boston, MA, 309–352. doi:10.1007/978-1-4899-7637-6_9

[81] Maral Kolahkaj, Ali Harounabadi, Alireza Nikravanshalmani, and Rahim Chinipardaz. 2020. A hybrid context-aware approach for e-tourism package recommendation based on asymmetric similarity measurement and sequential pattern mining. *Electronic Commerce Research and Applications* 42 (July 2020), 100978. doi:10.1016/J.ELERAP.2020.100978

[82] Géraud Le Falher, Aristides Gionis, and Michael Mathioudakis. 2021. Where Is the Soho of Rome? Measures and Algorithms for Finding Similar Neighborhoods in Cities. *International AAAI Conference on Web and Social Media* 9, 1 (Aug. 2021), 228–237. doi:10.1609/icwsm.v9i1.14602

[83] Fátima Leal, Benedita Malheiro, and Juan C. Burguillo. 2018. Context-aware tourism technologies. *The Knowledge Engineering Review* 33 (2018). doi:10.1017/s0269888918000152

[84] Fátima Leal, Bruno Veloso, Benedita Malheiro, and Juan C. Burguillo. 2025. Towards adaptive and transparent tourism recommendations: A survey. *Expert Systems: The Journal of Knowledge Engineering* 42, 1 (2025). doi:10.1111/EXSY.13400

[85] Forina Lestari, Melasutra Md Dali, and Norbani Che-Ha. 2022. The Importance of A Multistakeholder Perspective in Mapping Stakeholders' Roles Toward City Branding Implementation. *Policy & Governance Review* 6, 3 (2022), 264–281. doi:10.30589/pgr.v6i3.601

[86] Changheng Li, Yongjing Hao, Pengpeng Zhao, Fuzhen Zhuang, Yanchi Liu, and Victor S. Sheng. 2021. Tell Me Where to Go Next: Improving POI Recommendation via Conversation. In *26th International Conference on Database Systems for Advanced Applications (Lecture Notes in Computer Science, Vol. 12683)*. Springer, 211–227. doi:10.1007/978-3-030-73200-4_14

[87] Peibo Li, Maarten de Rijke, Hao Xue, Shuang Ao, Yang Song, and Flora D. Salim. 2024. Large Language Models for Next Point-of-Interest Recommendation. In *47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'24)*. ACM, 1463–1472. doi:10.1145/3626772.3657840

[88] Kwan Hui Lim, Jeffrey Chan, Christopher Leckie, and Shanika Karunasekera. 2016. Towards Next Generation Touring: Personalized Group Tours. In *26th International Conference on Automated Planning and Scheduling*, Vol. 26. AAAI, 412–420. doi:10.1609/icaps.v26i1.13775

[89] Kwan Hui Lim, Jeffrey Chan, Christopher Leckie, and Shanika Karunasekera. 2018. Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency. *Knowledge and Information Systems* 54, 2 (2018), 375–406. doi:10.1007/S10115-017-1056-Y

[90] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. In *2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, USA, Article 143. doi:10.1145/3411764.3445488

[91] Xin Liu, Yongjian Yang, Yuanbo Xu, Funing Yang, Qiuyang Huang, and Hong Wang. 2022. Real-time POI recommendation via modeling long- and short-term user preferences. *Neurocomputing* 467 (2022), 454–464. doi:10.1016/J.NEUCOM.2021.09.056

[92] Rui Neves Madeira, Samuel Robalo, and André Cordeiro. 2024. SHIFT towards Sustainable Tourism: Designing a Ubiquitous Platform for Co-Creating Experiences with Tourists and Residents. In *International Conference on Mobile and Ubiquitous Multimedia*. ACM, New York, NY, USA, 467–470. doi:10.1145/3701571.3703387

[93] Manqing Mao, Paishun Ting, Yijian Xiang, Mingyang Xu, Julia Chen, and Jianzhe Lin. 2024. Multi-User Chat Assistant (MUCA): a Framework Using LLMs to Facilitate Group Conversations. *arXiv* (2024). doi:10.48550/arXiv.2401.04883

[94] David Massimo and Francesco Ricci. 2019. Users' Evaluation of Next-POI Recommendations. In *RecSys Recommenders in Tourism Workshop (CEUR Workshop Proceedings, Vol. 2435)*. CEUR-WS.org, 1–8.

[95] David Massimo and Francesco Ricci. 2021. Popularity, novelty and relevance in point of interest recommendation: an experimental analysis. *Information Technology & Tourism* 23, 4 (2021), 473–508. doi:10.1007/S40558-021-00214-5

[96] David Massimo and Francesco Ricci. 2023. Combining Reinforcement Learning and Spatial Proximity Exploration for New User and New POI Recommendations. In *31st ACM Conference on User Modeling, Adaptation and Personalization*. ACM, 164–174. doi:10.1145/3565472.3592966

[97] Noemi Mauro, Livio Scarpinati, Fabio Ferrero, Angelo Geninatti Cossatin, and Claudio Mattutino. 2024. Point-of-Interest Recommender Systems: Nudging towards Sustainable Tourism. In *32nd ACM Conference on User Modeling, Adaptation and Personalization*. ACM, New York, NY, USA, 491–495. doi:10.1145/3631700.3664904

[98] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a Fair Marketplace: Counterfactual Evaluation of the trade-off between Relevance, Fairness & Satisfaction in Recommendation Systems. In *27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. ACM, New York, NY, USA, 2243–2251. doi:10.1145/3269206.3272027

[99] Zaiqiao Meng, Richard McCreadie, Craig Macdonald, and Iadh Ounis. 2020. Exploring Data Splitting Strategies for the Evaluation of Recommendation Models. In *14th ACM Conference on Recommender Systems*. ACM, 681–686. doi:10.1145/3383313.3418479

[100] Pavel Merinov, David Massimo, and Francesco Ricci. 2023. Behaviour-aware Tourist Profiles Data Generation. In *13th Italian Information Retrieval Workshop (CEUR Workshop Proceedings, Vol. 3448)*. CEUR-WS.org, 3–8.

[101] Pavel Merinov and Francesco Ricci. 2024. Positive-Sum Impact of Multistakeholder Recommender Systems for Urban Tourism Promotion and User Utility. In *18th ACM Conference on Recommender Systems*. ACM, 939–944. doi:10.1145/3640457.3688173

[102] Sara Migliorini, Damiano Carra, and Alberto Belussi. 2021. Distributing Tourists among POIs with an Adaptive Trip Recommendation System. *IEEE Transactions on Emerging Topics in Computing* 9, 4 (2021), 1765–1779. doi:10.1109/TETC.2019.2920484

[103] Thuy Ngoc Nguyen and Francesco Ricci. 2018. A chat-based group recommender system for tourism. *Information Technology & Tourism* 18, 1-4 (2018), 5–28. doi:10.1007/S40558-017-0099-Y

[104] Thuy Ngoc Nguyen, Francesco Ricci, Amra Delic, and Derek G. Bridge. 2019. Conflict resolution in group decision making: insights from a simulation study. *User Modeling and User-Adapted Interaction* 29, 5 (2019), 895–941. doi:10.1007/S11257-019-09240-9

[105] Janni Nielsen, Torkil Clemmensen, and Carsten Yssing. 2002. Getting access to what goes on in people's heads? Reflections on the think-aloud technique. In *2nd Nordic Conference on Human-computer Interaction (NORDICHI'02)*. ACM, 101–110. doi:10.1145/572020.572033

[106] Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. 2011. An Empirical Study of Geographic User Activity Patterns in Foursquare. In *5th International Conference on Weblogs and Social Media*, Vol. 5. AAAI, 570–573. doi:10.1609/icwsm.v5i1.14175

[107] Daniel M. Oppenheimer, Tom Meyvis, and Nicolas Davidenko. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology* 45, 4 (2009), 867–872. doi:10.1016/j.jesp.2009.03.009

[108] Mariam Orabi, Djedjiga Mouheb, Zaher Al Aghbari, and Ibrahim Kamel. 2020. Detection of bots in social media: a systematic review. *Information Processing & Management* 57, 4 (July 2020), 102250. doi:10.1016/j.ipm.2020.102250

[109] Keisuke Otaki and Yukino Baba. 2025. Travel itinerary recommendation using interaction-based augmented data. *Expert Systems with Applications* 269 (2025), 126294. doi:10.1016/J.ESWA.2024.126294

[110] Suayb Talha Ozcelik, Meltem Turhan Yondem, Ines Caetano, Jose Figueiredo, Patricia Alves, Goreti Marreiros, Huseyin Bahtiyar, Eda Yuksel, Fernando Perales, and George Suciu. 2024. Transforming Tourism Experience: AI-Based Smart Travel Platform. In *4th European Symposium on Software Engineering*. ACM, New York, NY, USA, 37–45. doi:10.1145/3651640.3651645

[111] Sangwon Park, Yang Yang, and Mingshu Wang. 2019. Travel distance and hotel service satisfaction: An inverted U-shaped relationship. *International Journal of Hospitality Management* 76 (2019), 261–270. doi:10.1016/j.ijhm.2018.05.015

[112] Isabel Paulino, Sergi Lozano, and Lluís Prats. 2021. Identifying tourism destinations from tourists' travel patterns. *Journal of Destination Marketing & Management* 19 (2021), 100508. doi:10.1016/j.jdmm.2020.100508

[113] Philip L. Pearce and Uk-Il Lee. 2005. Developing the Travel Career Approach to Tourist Motivation. *Journal of Travel Research* 43, 3 (2005), 226–237. doi:10.1177/0047287504272020

[114] Harald Pechlaner. 2000. Cultural heritage and destination management in the Mediterranean. *Thunderbird International Business Review* 42, 4 (2000), 409–426. doi:10.1002/1520-6874(200007/08)42:4<409::aid-tie4>3.0.co;2-f

[115] Harald Pechlaner, Elisa Innerhofer, and Greta Erschbamer. 2019. *Overtourism: Tourism management and solutions*. Routledge, London.

[116] Morgan Perry. 2023. Role of Cultural Heritage Preservation in Destination Branding and Tourist Experience Enhancement. *Hospitality and Tourism Journal* 1, 1 (2023), 36–47.

[117] Greta Piliponyte, David Massimo, and Francesco Ricci. 2023. The Impact of Personalised Advertisement Campaigns on Tourist Choices in South Tyrol: A Sustainable Tourism Perspective. In *31st ACM Conference on User Modeling, Adaptation and Personalization*. ACM, 100–103. doi:10.1145/3563359.3597445

[118] Greta Piliponyte, David Massimo, and Francesco Ricci. 2024. Simulation of recommender systems driven tourism promotion campaigns. *Information Technology & Tourism* 26, 3 (2024), 407–448. doi:10.1007/S40558-024-00283-2

[119] Stanley C. Plog. 1974. Why Destination Areas Rise and Fall in Popularity. *Cornell Hotel and Restaurant Administration Quarterly* 14, 4 (1974), 55–58. doi:10.1177/001088047401400409

[120] Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction* 22, 4 (01 Oct. 2012), 317–355. doi:10.1007/s11257-011-9115-7

[121] Hossein A. Rahmani, Mohammad Aliannejadi, Mitra Baratchi, and Fabio Crestani. 2020. *Joint Geographical and Temporal Modeling Based on Matrix Factorization for Point-of-Interest Recommendation*. Lecture Notes in Computer Science, Vol. 12035. Springer, 205–219. doi:10.1007/978-3-030-45439-5_14

[122] Francesco Ricci and Amra Delić. 2025. Widening the Role of Group Recommender Systems with CAJO. *SIGIR Forum* 59, 1 (June 2025), 1–16. doi:10.1145/3769733.3769745

[123] Francesco Ricci, David Massimo, and Antonella De Angeli. 2021. Challenges for Recommender Systems Evaluation. In *14th Biannual Conference of the Italian SIGCHI Chapter*. ACM, 25:1–25:5. doi:10.1145/3464385.3464733

[124] Rafael Robina-Ramírez, Jesús Torrecilla-Pinero, Ana Leal-Solís, and Juan Antonio Pavón-Pérez. 2024. Tourism as a driver of economic and social development in underdeveloped regions. *Regional Science Policy & Practice* 16, 1 (2024), 12639. doi:10.1111/rsp3.12639

[125] Michael W. Ross, Martin Y. Iguchi, and Sangeeta Panicker. 2018. Ethical aspects of data sharing and research participant protections. *American Psychologist* 73, 2 (Feb. 2018), 138–145. doi:10.1037/amp0000240

[126] Luca Rossi, Matthew Williams, Christoph Stich, and Mirco Musolesi. 2015. Privacy and the City: User Identification and Location Semantics in Location-Based Social Networks. In *9th International Conference on Web and Social Media*, Vol. 9. AAAI, 387–396. doi:10.1609/icwsm.v9i1.14595

[127] Pablo Sánchez. 2019. Exploiting contextual information for recommender systems oriented to tourism. In *13th ACM Conference on Recommender Systems*. ACM, New York, NY, USA, 601–605. doi:10.1145/3298689.3347062

[128] Pablo Sánchez and Alejandro Bellogín. 2020. Applying reranking strategies to route recommendation using sequence-aware evaluation. *User Modeling and User-Adapted Interaction* 30, 4 (2020), 659–725. doi:10.1007/S11257-020-09258-4

[129] Pablo Sánchez and Alejandro Bellogín. 2021. On the effects of aggregation strategies for different groups of users in venue recommendation. *Information Processing and Management* 58, 5 (2021), 102609. doi:10.1016/J.IPM.2021.102609

[130] Pablo Sánchez and Alejandro Bellogín. 2022. Point-of-Interest Recommender Systems Based on Location-Based Social Networks: A Survey from an Experimental Perspective. *Comput. Surveys* 54, 11s (2022), 223:1–223:37. doi:10.1145/3510409

[131] Pablo Sánchez, Alejandro Bellogín, and Ludovico Boratto. 2023. Bias characterization, assessment, and mitigation in location-based recommender systems. *Data Mining and Knowledge Discovery* 37, 5 (2023), 1885–1929. doi:10.1007/S10618-022-00913-5

[132] Pablo Sánchez, Alejandro Bellogín, and Jose L. Jorro-Aragoneses. 2025. Context Trails: A Dataset to Study Contextual and Route Recommendation. In *19th ACM Conference on Recommender Systems*, Mária Bieliková, Pavel Kordík, Markus Schedl, Marco de Gemmis, Sole Pera, Rodrigo Alves, Olivier Jeunen, and Vito Ostuni (Eds.). ACM, 716–725. doi:10.1145/3705328.3748151

[133] Pablo Sánchez and Linus W. Dietz. 2022. Travelers vs. Locals: The Effect of Cluster Analysis in Point-of-Interest Recommendation. In *30th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, 132–142. doi:10.1145/3503252.3531320

[134] Clodomir Santana, Federico Botta, Hugo Barbosa, Filippo Privitera, Ronaldo Menezes, and Riccardo Di Clemente. 2023. COVID-19 is linked to changes in the time–space dimension of human mobility. *Nature Human Behaviour* 7, 10 (July 2023), 1729–1739. doi:10.1038/s41562-023-01660-3

[135] Sanja Scepanovic, Sagar Joglekar, Stephen Law, and Daniele Quercia. 2021. Jane Jacobs in the Sky: Predicting Urban Vitality with Open Satellite Data. *ACM on Human-Computer Interaction* 5, Article 48 (April 2021), 25 pages. doi:10.1145/3449257

[136] Frank Schlosser, Vedran Sekara, Dirk Brockmann, and Manuel Garcia-Herranz. 2021. Biases in human mobility data impact epidemic modeling. doi:10.48550/arxiv.2112.12521 arXiv:2112.12521 [physics.soc-ph]

[137] Ali Akbar Septiandri, Marios Constantinides, Mohammad Tahaei, and Daniele Quercia. 2023. WEIRD FAccTs: How Western, Educated, Industrialized, Rich, and Democratic is FAccT?. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT'23)*. ACM, New York, NY, USA, 160–171. doi:10.1145/3593013.3593985

[138] Yewei Shang, Montserrat Pallares-Barbera, and Francesc Romagosa. 2023. Coordination Mechanism of Stakeholders in Tourism Destinations: A Social Network Analysis Exploration. In *New Perspectives and Paradigms in Applied Economics and Business*. Springer, Cham, 363–381. doi:10.1007/978-3-031-23844-4_26

[139] Egon Smeral. 2019. Overcrowding of tourism destinations: Some suggestions for a solution. In *Overtourism*. Routledge, 163–173.

[140] Xiaoyu Song, Zhizhong Liu, Lingqiang Meng, Dianhui Chu, Jian Yu, and Quan Z. Sheng. 2025. Accurate POI recommendation for random groups with improved graph neural networks and a multi-negotiation model. *Scientific Reports* 15, 1 (March 2025). doi:10.1038/s41598-025-91805-3

[141] Michaël Soumm, Adrian Popescu, and Bertrand Delezoide. 2023. Vis2Rec: A Large-Scale Visual Dataset for Visit Recommendation. In *IEEE/CVF Winter Conference on Applications of Computer Vision*. IEEE, 2986–2996. doi:10.1109/WACV56688.2023.00300

[142] Steffen Staab, Hannes Werthner, Francesco Ricci, Alexander Zipf, Ulrike Gretzel, Daniel R. Fesenmaier, Cécile Paris, and Craig A. Knoblock. 2002. Intelligent Systems for Tourism. *IEEE Intelligent Systems* 17, 6 (2002), 53–64. doi:10.1109/MIS.2002.1134362

[143] Giorgos Stamatelatos, George Drosatos, Sotirios Gyftopoulos, Helen Briola, and Pavlos S. Efraimidis. 2021. Point-of-interest lists and their potential in recommendation systems. *Information Technology & Tourism* 23, 2 (Feb. 2021), 209–239. doi:10.1007/s40558-021-00195-5

[144] Ugljesa Stankov and Ulrike Gretzel. 2020. Tourism 4.0 technologies and tourist experiences: a human-centered design perspective. *Information Technology & Tourism* 22, 3 (2020), 477–488. doi:10.1007/S40558-020-00186-Y

[145] Huimin Sun, Jiajie Xu, Kai Zheng, Pengpeng Zhao, Pingfu Chao, and Xiaofang Zhou. 2021. MFNP: A Meta-optimized Model for Few-shot Next POI Recommendation. In *30th International Joint Conference on Artificial ntelligence (IJCAI'21)*. International Joint Conferences on Artificial Intelligence Organization, 3017–3023. doi:10.24963/IJCAI.2021/415

[146] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. YFCC100M: the new data in multimedia research. *Commun. ACM* 59, 2 (Jan. 2016), 64–73. doi:10.1145/2812802

[147] Nava Tintarev and Judith Masthoff. 2022. Beyond Explaining Single Item Recommendations. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, New York, NY, 711–756. doi:10.1007/978-1-0716-2197-4_19

[148] Leonardo Tonetto, Pauline Kister, Nitinder Mohan, and Jörg Ott. 2024. Ethical and Privacy Considerations with Location Based Data Research. doi:10.48550/arxiv.2403.05558 arXiv:2403.05558 [cs.CY]

[149] Christoph Trattner, Alexander Oberegger, Leandro Balby Marinho, and Denis Parra. 2018. Investigating the Utility of the Weather Context for Point of Interest Recommendations. *Information Technology & Tourism* 19, 1-4 (2018), 117–150. doi:10.1007/s40558-017-0100-9

[150] Nikolaos Vathis, Charalampos Konstantopoulos, Grammati E. Pantziou, and Damianos Gavalas. 2023. The Vacation Planning Problem: A multi-level clustering-based metaheuristic approach. *Computers & Operations Research and their Application to Problems of World Concern* 150 (2023), 106083. doi:10.1016/J.COR.2022.106083

[151] Anna Dalla Vecchia, Sara Migliorini, Elisa Quintarelli, and Alberto Belussi. 2023. A Context-Aware Recommendation System with a Crowding Forecaster. In *31st Symposium of Advanced Database Systems (CEUR Workshop Proceedings, Vol. 3478)*. CEUR-WS.org, 632–640.

[152] Anna Dalla Vecchia, Sara Migliorini, Elisa Quintarelli, Mauro Gambini, and Alberto Belussi. 2024. Promoting sustainable tourism by recommending sequences of attractions with deep reinforcement learning. *Information Technology & Tourism* 26, 3 (2024), 449–484. doi:10.1007/S40558-024-00288-X

[153] Nynke Elske Vellinga and Ekaterina Hailevich. 2025. The legal framework for sharing mobility data: on the road to an EU mobility data space. *Computer Law & Security Review* 58 (Sept. 2025). doi:10.1016/j.clsr.2025.106188

[154] Bin Wang, Tiantian Dong, Yunyao Liu, Jay Kandampully, and Zhanlu Tang and. 2024. Males or females in solo or group travel: how do they impact travel intentions of potential tourists with different self-construals? *Current Issues in Tourism* (2024), 1–20. doi:10.1080/13683500.2024.2417713

[155] Gang Wang, Sarita Yardi Schoenebeck, Haitao Zheng, and Ben Y. Zhao. 2016. "Will Check-in for Badges": Understanding Bias and Misbehavior on Location-Based Social Networks. In *Tenth International Conference on Web and Social Media*. AAAI, 417–426.

[156] Huandong Wang, Yong Li, Chen Gao, Gang Wang, Xiaoming Tao, and Depeng Jin. 2021. Anonymization and De-Anonymization of Mobility Trajectories: Dissecting the Gaps Between Theory and Practice. *IEEE Transactions on Mobile Computing* 20, 3 (March 2021), 796–815. doi:10.1109/tmc.2019.2952774

[157] Shirui Wang, Bohan Xie, Ling Ding, Xiaoying Gao, Jianting Chen, and Yang Xiang. 2024. SeCor: Aligning Semantic and Collaborative Representations by Large Language Models for Next-Point-of-Interest Recommendations. In *18th ACM Conference on Recommender Systems*. ACM, 1–11. doi:10.1145/3640457.3688124

[158] Shoujin Wang, Xiuzhen Zhang, Yan Wang, and Francesco Ricci. 2024. Trustworthy Recommender Systems. *ACM Transactions on Intelligent Systems and Technology* 15, 4 (2024), 84:1–84:20. doi:10.1145/3627826

[159] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random. In *36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, 6638–6647.

[160] Zehui Wang, Wolfram Höpken, and Dietmar Jannach. 2025. A survey on point-of-interest recommendations leveraging heterogeneous data. *Information Technology & Tourism* 27, 1 (2025), 29–73. doi:10.1007/S40558-024-00301-3

[161] Heitor Werneck, Rodrigo Santos, Nícollas Silva, Adriano C. M. Pereira, Fernando Mourão, and Leonardo Rocha. 2021. Effective and diverse POI recommendations through complementary diversification models. *Expert Systems with Applications* 175 (Aug. 2021), 114775. doi:10.1016/j.eswa.2021.114775

[162] Heitor Werneck, Nícollas Silva, Adriano C. M. Pereira, Matheus Carvalho Viana, Alejandro Bellogín, Jorge Martinez-Gil, Fernando Mourão, and Leonardo Rocha. 2022. A reproducible POI recommendation framework: Works mapping and benchmark evaluation. *Information Systems* 108 (Sept. 2022), 102019. doi:10.1016/j.is.2022.102019

[163] Hannes Werthner and Stefan Klein. 1999. ICT and the Changing Landscape of Global Tourism Distribution. *Electronic Markets* 9, 4 (1999), 256–262. doi:10.1080/101967899358941

[164] Yanan Xiao, Lu Jiang, Kunpeng Liu, Yuanbo Xu, Pengyang Wang, and Minghao Yin. 2024. Hierarchical Reinforcement Learning for Point of Interest Recommendation. In *33rd International Joint Conference on Artificial Intelligence (IJCAI'24)*. International Joint Conferences on Artificial Intelligence Organization, 2460–2468. doi:10.24963/ijcai.2024/272

[165] Yanyan Xu, Luis E. Olmos, David Mateo, Alberto Hernando, Xiaokang Yang, and Marta C. González. 2023. Urban dynamics through the lens of human mobility. *Nature Computational Science* 3, 7 (July 2023), 611–620. doi:10.1038/s43588-023-00484-5

[166] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, Kaoru Sezaki, Esteban Moro, and Alex Pentland. 2024. YJMob100K: City-Scale and Longitudinal Dataset of Anonymized Human Mobility Trajectories. doi:10.5281/ZENODO.10836269

[167] Emre Yalcin and Alper Bilge. 2022. Evaluating unfairness of popularity bias in recommender systems: A comprehensive user-centric analysis. *Information Processing and Management* 59, 6 (2022), 103100. doi:10.1016/J.IPM.2022.103100

[168] An Yan, Zhankui He, Jiacheng Li, Tianyang Zhang, and Julian McAuley. 2023. Personalized Showcases: Generating Multi-Modal Explanations for Recommendations. In *46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'23)*. ACM, New York, NY, USA, 2251–2255. doi:10.1145/3539618.3592036

[169] Eva Zangerle and Christine Bauer. 2022. Evaluating Recommender Systems: Survey and Framework. *Comput. Surveys* 55, 8 (Dec. 2022), 1–38. doi:10.1145/3556536

[170] Chenyi Zhang and Ke Wang. 2016. POI recommendation through cross-region collaborative filtering. *Knowledge and Information Systems* 46, 2 (2016), 369–387. doi:10.1007/S10115-015-0825-8

[171] Lu Zhang, Zhu Sun, Jie Zhang, Yiwen Wu, and Yunwen Xia. 2023. Conversation-Based Adaptive Relational Translation Method for Next POI Recommendation With Uncertain Check-Ins. *IEEE Transactions on Neural Networks and Learning Systems* 34, 10 (2023), 7810–7823. doi:10.1109/TNNLS.2022.3146443

[172] Zengbin Zhang, Lin Zhou, Xiaohan Zhao, Gang Wang, Yu Su, Miriam Metzger, Haitao Zheng, and Ben Y. Zhao. 2013. On the validity of geosocial mobility traces. In *12th Workshop on Hot Topics in Networks (HotNets-XII)*. ACM, New York, NY, USA, Article 11, 7 pages. doi:10.1145/2535771.2535786

[173] Vincent Wenchen Zheng, Yu Zheng, Xing Xie, and Qiang Yang. 2010. Collaborative location and activity recommendations with GPS history data. In *19th International Conference on World Wide Web*. ACM, 1029–1038. doi:10.1145/1772690.1772795

Received 12 June 2025; revised 15 December 2025; accepted 26 April 2026