ECAI 2025
I. Lynce et al. (Eds.)

doi:10.3233/FAIA251463

Harnessing Large Language Models for Efficient Crowd Management in Large-Scale Events

Blerina Spahiu[a,*], Marco Cremaschi[a], Andrea Maurino[a] and Giuseppe Vizzari[a]

[a]University of Milan - Bicocca, Italy
ORCID (Blerina Spahiu): https://orcid.org/0000-0002-6958-8215, ORCID (Marco Cremaschi): https://orcid.org/0000-0001-7840-6228, ORCID (Andrea Maurino): https://orcid.org/0000-0001-9803-3668, ORCID (Giuseppe Vizzari): https://orcid.org/0000-0002-7916-6438

Abstract. Managing large public events involves significant challenges, including transportation congestion, attendee coordination, and enhancing the overall event experience. This paper introduces a novel approach that integrates Large Language Models (LLMs) and generative AI to address these issues effectively. We present a framework that leverages enriched knowledge graphs comprising diverse datasets, such as geographic information, transportation systems, environmental factors, and more, to recommend personalised, context-aware itineraries. By employing LLMs, we aim to manage attendee flow, stagger departure times, and guide individuals through engaging points of interest. Additionally, we incorporate generative AI to design gamified content, such as interactive quizzes and puzzles, tailored to user preferences. These gamification elements not only provide entertainment but also encourage staggered event departures, mitigating post-event congestion. The experimental study conducted in Milan demonstrated the effectiveness of the proposed system: AI-generated itineraries closely matched expected travel times, with minimal deviations of 2–5 minutes. Moreover, responses to the user experience questionnaire reflected high levels of usability, engagement, and overall satisfaction, reinforcing the potential of this approach for improving post-event mobility and attendee experience.

1 Introduction

Public events, such as concerts, sports games, and conventions, serve as powerful catalysts for cultural exchange, social inclusion, and economic growth [19]. These events attract many visitors, boosting local businesses and stimulating economic activity [22]. In addition, they encourage cultural and social interaction and often drive infrastructure development, leading to lasting improvements in transportation, accommodation, and other essential elements, positively shaping the city's overall development (e.g., the creation of ATM M5 for the EXPO 2015 in Milan[1], the extension of line 11 in Paris[2], during Olympic games).

Although large events bring numerous benefits, they also pose significant challenges for host cities, such as increased traffic congestion, parking shortages, and pressure on public transportation systems, which can inconvenience both residents and visitors [10]. Moreover, they can contribute to environmental issues, including increased pollution, waste, and energy consumption [7]. Cities like Milan have created Action Plans to manage event mobility, aiming to ease congestion, cut emissions, and promote walking and cycling[3].

Large-scale events typically result in congestion during two key time periods: before the event begins and after it ends. The pre-event rush often involves attendees all arriving within a narrow window, leading to bottlenecks in transportation, parking, and venue entry. Research shows that time affects the marginal utility difference between origins and destinations, influencing travel schedules, confirming that many event attendees have flexible schedules, making early departures effective [10, 6, 5]. Similarly, large crowds leaving events at once cause peak congestion, resulting in pedestrian bottlenecks, long transit waits, and inefficiencies in public transportation. Traditional crowd management solutions, such as increasing transit capacity and temporary traffic redirection, have shown limited success [23]. Although effective in some scenarios, these approaches are often costly, require extensive coordination, and fail to provide a truly adaptive response to the dynamic nature of crowd behavior. This highlights the need for innovative solutions that can not only distribute crowds more effectively but also enhance the overall experience of attendees [21].

Recent advances in AI, particularly Large Language Models (LLMs), offer promising new approaches to this challenge. These systems, trained on vast amounts of text data, can understand context, generate personalised recommendations, and communicate in natural language. Models like GPT-4, Claude, and LLaMA have demonstrated remarkable capabilities in processing complex information and providing tailored responses across diverse applications [11]. By leveraging these capabilities in the context of event management, we can develop more adaptive and engaging solutions for crowd distribution.

This paper introduces a novel framework that leverages LLMs and generative AI to enhance event management, with a primary focus on post-event engagement strategies, although the same approach holds strong potential for pre-event planning. While the system is designed to support both phases, our current implementation and evaluation centre on the post-event context. In this phase, we use LLMs to generate personalised itineraries around nearby Points of Interest (POIs) to help naturally disperse crowds. Additionally, we leverage LLMs for gamification, transforming potentially frustrating delays into engaging and entertaining experiences.

* Corresponding Author. Email: blerina.spahiu@unimib.it
[1] https://www.firstonline.info/en/expo-2015-a-milano-apre-la-prima-tratta-della-nuova-metro-5/
[2] https://www.adlittle.com/en/insights/viewpoints/paris-2024-olympics
[3] https://ckan.mobidatalab.eu/dataset/action-plan-on-large-events-management

In this paper we make the following contributions: (i) integrate and enrich datasets of different domains (geographic information, transportation systems, points of interest, etc.) into a big knowledge graph; (ii) utilise LLMs to recommend personalised, context-aware itineraries, guiding attendees through POIs; (iii) explore the potential of generative AI to create engaging gamification elements, such as quizzes and puzzles, tailored to individual preferences, and (iv) publicly release the enriched knowledge graph and the code for itinerary generation and gamification, and the code of the application[4]. This framework not only enhances the event-going experience but also contributes to more efficient crowd management, ultimately paving the way for smarter, more enjoyable large-scale events.

The remainder of this paper is organised as follows: Section 2 reviews the related literature. Section 3 outlines the methodology used to develop the framework, including details on dataset integration and model fine-tuning. Section 4 presents the application of the framework in a real-world context in Milan, along with the results and findings from a user satisfaction survey. Section 5 discusses the limitations of our approach, and Section 6 concludes the paper.

2 Related Work

Managing large-scale events has been widely studied due to their logistical challenges and social benefits associated with these gatherings. The management of crowds at large events has traditionally relied on a combination of physical infrastructure planning, human resources, and static modelling techniques [2].

Cities have historically employed measures such as temporary road closures [14], dedicated lanes [9], and increased public transportation frequency to address event-related congestion. These approaches, while necessary, often fail to adapt to real-time conditions and typically require substantial resources for implementation [16]. Physical crowd control measures, including barriers, queuing systems, and designated entry and exit points, provide structure, but they lack flexibility in response to dynamic crowd behaviours [1]. Similarly, static prediction models based on historical data [3], offer general planning guidance but often cannot account for unexpected variations in attendance patterns or external factors affecting crowd movement. Machine learning algorithms have demonstrated greater accuracy than traditional methods in forecasting crowd movements, utilising variables such as ticket sales, weather, and social media activity. However, they face challenges in capturing the nuances of human behaviour [20].

Gamification has proven effective in influencing mobility behavior, with studies like [4] showing a 25% increase in daily steps through Pokémon GO. Building on this, our approach integrates interactive challenges into walking itineraries to sustain engagement and encourage users to follow suggested routes rather than heading directly to metro stations. Public transport systems have implemented gamification to encourage off-peak travel. Singapore's Travel Smart Rewards [18] program offered incentives to commuters who adjusted their schedules to avoid rush hours. This successfully reduced peak-hour congestion. While their approach relied on economic incentives, this research focuses on engagement-driven motivation, using gamification elements rather than monetary rewards to guide pedestrian movement. Additionally, gamification in pedestrian navigation apps has been used to crowdsource data on walkable routes, enabling the creation of pedestrian-friendly maps through incentive systems and peer-reviewed contributions [17]. While effective for data collection and encouraging exploration, this approach focuses on static route mapping rather than adapting to real-time, event-driven mobility patterns.

[4] https://github.com/unimib-datAI/eventure

The application of artificial intelligence to crowd management represents the cutting edge of research in this field. LLMs have only recently begun to be explored in this context. Preliminary research by [19] introduces the LLM-MPE framework, which leverages LLMs to process unstructured textual data about public events, such as descriptions from online sources, and combines it with historical mobility data to predict crowd movements. The study demonstrates that LLM-MPE outperforms traditional models in forecasting human mobility during events, offering interpretable insights into its predictions.

Despite these advances, significant gaps remain in current research. Integration of different technological approaches into comprehensive systems has been limited, with most studies focusing on isolated aspects of crowd management. Additionally, personalization of crowd management strategies based on individual preferences and behaviors remains underdeveloped in the literature [8]. The combination of AI-driven recommendations with gamification elements represents a particularly promising but underexplored direction. While both approaches have shown merit independently, their integration into cohesive systems that address both pre-event planning and post-event dispersal has received minimal academic attention.

This study builds upon these advancements by introducing an LLM-driven itinerary generation framework that not only optimizes pedestrian routes but also actively mitigates congestion near metro stations after large events. Unlike prior studies, this approach integrates cultural and gamified elements directly into the itinerary, providing users with contextual information and interactive engagement at POIs. This ensures that pedestrian movement is both strategically distributed and enriched with meaningful experiences, making urban exploration more dynamic and sustainable.

3 Methodology

To implement the proposed framework, we adopted a multi-layered methodology that combines different datasets, their processing and enrichment, POIs extraction and their distance computation, itinerary generation, and gamification strategies.

3.1 Dataset Collection

To build a comprehensive view of Milan's urban context, data was gathered from multiple sources, including Wikidata[5], the City of Milan's open data portal[6], and Open-Meteo[7]. A total of 10 datasets were integrated for this analysis.

A key dataset includes 4,830 POIs extracted from Wikidata. This source was selected for its frequent updates, high level of detail, and seamless compatibility with other services, making it ideal for dynamic urban applications. Complementary data from the Municipality of Milan's open data portal reflects various aspects of urban life and infrastructure.

- The tree location dataset[8] contains 255,075 georeferenced records, each with 16 attributes, covering all trees managed by the city.
- The WiFi antenna dataset[9] includes 597 entries with 14 properties, offering insights into the city's open wireless infrastructure.

[5] https://www.wikidata.org/
[6] http://www.datiopen.it/it
[7] https://open-meteo.com/
[8] https://dati.comune.milano.it/dataset/ds2484_infogeo_alberi_localizzazione
[9] https://dati.comune.milano.it/dataset/ds69-ammcomunale-antenne-open-wifi-localizzazione

- The historic shops dataset[10] documents 460 commercial establishments with at least 50 years of continuous activity, each described by 12 attributes.
- The bench dataset[11] lists 29,996 georeferenced benches with 11 properties. Similarly, the picnic table dataset[12] includes 466 elements, also described by 11 attributes. Finally, the fountain (Vedovelle) dataset[13] contains 674 public water sources, each with 8 descriptive fields, useful particularly in warmer seasons.
- The NIL (Nuclei di Identità Locale, Local Identity Units in english) dataset[14] comprises 88 zones with 13 attributes, capturing neighborhood-level characteristics and their connections to infrastructure and services.
- Weather data from Open-Meteo[15], selected parameters include temperature, precipitation (amount and probability), WMO weather codes, wind speed, and day/night indicators, chosen for their relevance to user needs in weather forecasts.

These datasets were selected for their ability to provide a rich, multi-layered understanding of Milan's urban landscape. They capture key aspects such as public space usage, cultural heritage, environmental assets, digital infrastructure, and neighborhood identity, elements that directly impact mobility patterns, citizen well-being, and urban planning. Together, they support data-driven insights for designing more accessible, sustainable, and user-centered solutions for large-scale events and everyday city life.

Finally, DBpedia[16] is considered as the dataset for the generation of quizzes.

3.2 Data Processing and Enrichment

To ensure high data quality, interoperability, and semantic consistency, the framework implements a robust preprocessing and enrichment pipeline that systematically transforms raw datasets into structured, meaningful spatial information.

The preprocessing phase starts with cleaning and normalizing data through various transformation tasks. This includes splitting compound fields into discrete components (e.g., separating street names from civic numbers), and disaggregating categorical fields such as POI types to better align with a defined taxonomy. Geospatial attributes like latitude and longitude are converted from strings to numeric types to restore semantic meaning. OpenRefine[17] is used extensively for faceted filtering, allowing the detection and normalization of inconsistent entries (e.g., malformed values like "Y?not" or civic numbers like "BOC") and duplicates (e.g., variations of "Viale De Gasperi").

A key part of the pipeline is the deduplication and harmonization of POI types, achieved through clustering techniques. The framework combines the Nearest Neighbour algorithm with Levenshtein distance and the Key Collision method using the Fingerprint function to identify and unify semantically similar POI labels.

To enrich records lacking complete address data, geocoding is performed using the Geoapify Geocoding API[18]. Out of 1,048 spatial records, 911 were successfully enriched with missing street names, civic numbers, and postal codes. This geospatial enrichment notably improved datasets such as NIL boundaries, Vedovelle fountains, and Wikidata POIs, enabling precise mapping and cross-referencing (e.g., multiple POIs linked to "Via Pescara 41").

The pipeline also supports semantic enrichment by tagging each entity with at least one category to enable classification and advanced search. To reconcile and standardize categories across heterogeneous datasets, the framework uses OpenRefine with the Reconcile-csv extension. Through fuzzy matching and taxonomy-based mapping, raw labels (e.g., "theatre") are aligned with standardized categories (e.g., "Theatre and Cinema"). When ambiguous matches arise, a human-in-the-loop approach is applied for validation. Clustering algorithms are used to further group semantically similar labels (e.g., "hotel" and "Hotel") and consolidate them into unified terms like "Hotel."

Once reconciled and validated, these enriched attributes replace their original counterparts, transforming noisy raw data into a harmonized, query-ready spatial knowledge base.

3.3 Distance Matrix Computation

The ability to accurately compute walking distances and travel times is crucial in our approach.

OpenStreetMap[19] (OSM) is a collaborative mapping platform that offers freely accessible geographic data, curated by a global community of volunteers [12]. To calculate walking distances between POIs, the Open Source Routing Machine[20] (OSRM) is used, ensuring that routes follow pedestrian paths rather than roads intended for vehicles. OSRM's architecture consists of three main components: preprocessing, which converts OSM data into a routable graph; customisation, which tailors the graph to specific travel modes (e.g., walking); and query execution, which computes the most efficient routes based on this optimized graph. For this walking-specific routing implementation, geographical data from Geofabrik[21] has been utilised, specifically the `nord-ovest-latest.osm.pbf` dataset covering northern Italy, including Milan. The data processing workflow begins with extracting pedestrian paths using the `foot.lua` profile while filtering out unsuitable routes. The process continues with partitioning the dataset to optimise query performance and refining the routing graph with pedestrian-specific attributes such as footpaths, crossings, and walking speeds. Docker is used to containerise the OSRM setup, enabling a reproducible and dependency-free environment for preprocessing geographic data and deploying the routing engine efficiently.

Once operational, the server's API enables the computation of walking distances and travel times between POIs through the `/route/v1/foot` endpoint, providing distance in meters and duration in seconds. For multiple POI calculations, the matrix API (`/table/v1/foot`) generates comprehensive distance and travel time matrices suitable for accessibility studies and network analysis.

3.4 LLM Fine-Tuning for Itinerary Generation

The fine-tuning pipeline for itinerary generation using a language model employs a structured, multi-stage approach designed to teach the model how to plan, validate, and reason about travel itineraries

[10] https://dati.comune.milano.it/dataset/ds2858-economia-botteghe-storiche-2024
[11] https://dati.comune.milano.it/dataset/ds2748_infogeo_panchine_localizzazione
[12] https://dati.comune.milano.it/dataset/ds2749_infogeo_tavoli_picnic_localizzazione
[13] https://dati.comune.milano.it/dataset/ds502_fontanelle-nel-comune-di-milano
[14] https://dati.comune.milano.it/dataset/ds964-nil-vigenti-pgt-2030
[15] https://open-meteo.com/
[16] https://www.dbpedia.org/
[17] https://openrefine.org/
[18] https://www.geoapify.com/geocoding-api/
[19] https://wiki.openstreetmap.org/wiki/API
[20] https://project-osrm.org/docs/v5.24.0/api/#
[21] https://www.geofabrik.de/

within a specific set of rules. The process begins with model selection and setup. Mistral [15], an open-weight LLM is chosen for its strong reasoning capabilities and its free availability through platforms like Hugging Face[22]. The implementation utilised PyTorch[23] for underlying deep learning operations, Hugging Face Transformers[24] for model management, and Unsloth[25] to optimise memory and performance during training.

The dataset used for fine-tuning was carefully designed to include a wide range of structured examples, built upon modified distance matrices. Initially, a distance matrix was derived from the POIs of the KG in Milan[26]. To ensure the model would not memorise specific POIs or locations, the matrix was altered by remapping each POI to a different location in Rome while preserving walking durations. Additional generalised matrices were created to diversify the dataset further, replacing POI names with abstract labels while retaining the logical structure of the connections. These matrices ensured that the model learned the structure of itinerary planning rather than relying on memorisation of fixed spatial data.

An example of such a generalized matrix is shown below:

```
From metro: GrandCentral, to art_gallery: MetroGallery - Walking time = 5 min
From museum: ScienceMuseum, to monument: PoetryGarden - Walking time = 5 min
[...]
From metro: UnionSquare, to sculpture: WindSpirit - Walking time = 9 min
```

Using these matrices, several types of training examples were generated. One set focused on itinerary validation and reasoning. In this task, the model was given a prompt explaining all the rules that define a valid itinerary, such as avoiding duplicate POIs, ending at a transportation hub, preserving bidirectional walking durations, and retrieving times only from a predefined list of connections. Along with the prompt, each example included a distance matrix and a sample itinerary that was either valid or intentionally incorrect. The model was tasked with assessing whether the itinerary was correct and providing a justification. This helped the model learn to reason through rules and detect logical inconsistencies in itinerary structure. A sample prompt-response pair is shown in Listing 1.

Listing 1: An example of the prompt-response

```
PROMPT:
Validate these itineraries respecting the following rules:

[RULES]
1. START: Starts with a non-transport location
2. CONTINUITY: Each stop must connect to the next
3. TRANSPORT END: Every itinerary MUST end at a transport location
4. NO INTERMEDIATE TRANSPORT: Transport locations only allowed as final stop
5. LENGTH: Itineraries can have 3-6 stops
6. UNIQUE: No location can be used twice in the same itinerary
7. DISTINCT: Intermediate stops should be different across itineraries
8. BIDIRECTION: The connections are bidirectional
9. WALKING TIME: Walking time must be retrieved from the connection list

[BIDIRECTIONAL CONNECTIONS - IMPORTANT]
Every connection in this system works both ways:
1. If you can walk from A →B in X minutes
2. You can ALWAYS walk from B →A in the same X minutes
3. Example from the connections list:
  - If "Museum: Louvre →Church: Notre Dame = 10 min" exists
  - Then "Church: Notre Dame →Museum: Louvre = 10 min" is automatically valid
4. Key Points:
  - Same walking time in both directions
  - No need for explicit listing of reverse routes
  - Use ANY connection in reverse with confidence
  - The walking time stays EXACTLY the same in either direction

[CONNECTION LIST]
From sculpture: WindSpirit, to museum: CityHistory - Walking time = 5 min
From museum: NaturalHistoryMuseum, to monument: VictoryColumn - Walking time = 4
    min
From monument: VeteransMemorial, to theater: Trianon - Walking time = 5 min
From museum: ModernArtMuseum, to monument: VictoryColumn - Walking time = 4 min
[...]

Itinerary 6:
Start: (CityHistory, museum)
Stop 1: (FreedomTower, monument) - Walking Time: 7 min [reverse]
Stop 2: (Trianon, theater) - Walking Time: 7 min [forward]
Stop 3: (TimesSquare, metro) - Walking Time: 4 min [reverse]
Total Time: 18 min

Response:
The itinerary is valid. It respects all rules:
1. ✓Starts with a non-transport location
2. ✓All stops are connected according to the connection list
3. ✓Ends at a transport location
4. ✓No transport locations used as intermediate stops
5. ✓Has 4 stops (within 3-6 range)
6. ✓No location is used more than once
7. ✓Intermediate stops are different from other itineraries
8. ✓Uses bidirectional connections correctly
9. ✓Walking times match the connection list
```

Another set of training examples focused on retrieving specific walking durations between two POIs. Here, the model received a prompt outlining bidirectionality rules and time constraints, along with a modified matrix and a query such as, "What is the walking time between POI A and POI B?" The model had to extract the correct time directly from the matrix. Similarly, a third type of training example asked the model to retrieve all connections associated with a given POI, again enforcing strict rules like exact name matching and correct formatting of bidirectional links. The dataset, generated via Python scripts, included 5,500 examples balanced between valid and invalid itineraries. It was converted into a JSON format compatible with LLM fine-tuning, structured as two-turn conversations (`[prompt, response]`).

The fine-tuning process begins with initialising the Mistral model with optimised parameters. The model is loaded using the FastLanguageModel Api[27], ensuring efficient memory utilisation by leveraging 4-bit quantisation. This approach helps in reducing computational overhead while maintaining model performance. To enable efficient fine-tuning while preserving memory efficiency, LoRA (Low-Rank Adaptation) [13] is applied to the model. This significantly reduces the number of trainable parameters while maintaining expressiveness and adaptability. Fine-tuning achieves reasonable training speeds of 800-1,000 tokens/sec on the T4 GPU, enabling adaptation of LLMs to specialised domains with minimal computational resources. To validate the distance matrix, selected results were compared with Google Maps[28], showing consistent differences of 1–2 minutes. This variation may be due to the exact reference points used for each POI in the two systems, as well as Google Maps' ability to account for factors such as downhill or uphill terrain.

3.5 Gamification: Multiple-Choice Question Generation with Mistral-7B

To reduce congestion near metro stations and promote exploration, gamification is used to engage users with interactive, educational content. This section describes how multiple-choice questions (MCQs) are generated for POIs from the previously generated itineraries.

For each POI in the KG, we retrieved, where possible, information from additional attributes such as *instanceOf*, *startTime*, *material*, and *creator*, along with descriptive metadata like textual descriptions, providing deeper context for each POI. Further enrichment was achieved by querying DBpedia for related abstracts for these POIs. To ensure accurate matches, Wikidata labels were refined by replacing spaces with underscores and removing phrases like *"Monument to"*, e.g., *"Monument to Giuseppe Parini"* became *"Giuseppe Parini"*, retrieving abstracts relevant to the individual honoured by the monument. The final dataset includes key attributes for each POI, such as label, type, heritage status, materials, description, attribution (e.g., creator, date), and DBpedia abstracts. Missing values are handled using `("")`, ensuring consistent formatting and avoiding processing errors.

After retrieving enriched POI data, the Mistral language model is used to generate fact-based multiple-choice questions. To optimize

[22] https://huggingface.co/
[23] https://pytorch.org/
[24] https://huggingface.co/docs/transformers/index
[25] https://docs.unsloth.ai/
[26] Please refer to this link for more info for the fine-tunning of the model: https://github.com/unimib-datAI/eventure/blob/main/supplementaryMaterial.pdf
[27] https://github.com/huggingface/trl
[28] https://www.google.com/maps

performance and manage computational resources, the model runs in float16 precision on T4 GPUs, reducing memory usage and increasing speed with minimal quality loss. For each POI, a tailored prompt is constructed using relevant attributes, such as description, abstract, type, creator, and construction details, to provide the model with accurate context. The prompt is carefully designed to produce high-quality educational questions with a clear format: a question, multiple answer choices (including plausible distractors), and the correct answer marked with an asterisk. Each question is also paired with a factual statement to enhance learning. Once prepared, the prompt is tokenised and passed to the model, which generates up to 2000 tokens per response. Outputs include the POI label, type, the question, answer options, and a related fact. Robust exception handling ensures failed generations (e.g., due to timeouts or model errors) are skipped without disrupting the overall process. This pipeline produces consistent, informative, and accessible content that supports cultural heritage engagement through LLM-generated educational material.

4 Case Study Evaluation

Major events in Milan, such as concerts, festivals, and fashion shows, often result in significant pedestrian congestion, particularly around key locations like Piazza del Duomo. For the scope of this study, only classes like metro stops and POIs such as museums, theatres, monuments, and sculptures have been retrieved from the created KG. To reduce the computational request both for the distance matrix creation and for the LLM, an area around Duomo of 600 meters has been selected to simulate the area around an event (as the crow flies), retrieving a total of 33 POIs, including 3 metro stops. The result is a matrix of distances and durations between each POI. The initial 33 POIs generated a total of 528 connections. Analysis of the distance matrix revealed a minimum walking distance of 14 meters, a maximum of 1,428 meters, and an average of 629 meters. The corresponding average travel time was 454 seconds (~7.5 minutes), with a maximum of 1,028 seconds (~17 minutes).

The predominant distance between POIs ranges from 400 to 700 meters with the highest frequency occurring around 600 meters. This distribution aligns effectively with the study's objectives of facilitating pedestrian movement between POIs end routes to the nearest metro station. The observed distance ranges are optimal for this purpose, as longer paths would be unnecessary given the focus on local pedestrian circulation patterns.

This study serves as a prototype to evaluate the feasibility of AI-generated itineraries in managing post-event pedestrian congestion. While the experiment involves a limited number of participants (10), the goal is to test the effectiveness of an AI-driven system in dynamically generating itineraries that encourage attendees to explore cultural landmarks before heading to metro stations. By assessing this controlled scenario, it is possible to determine whether this approach could be scaled to accommodate larger crowds in real-world applications.

4.1 Itenerary Generation

To simulate a post-event scenario, the experiment assumes that an event has occurred in Piazza del Duomo, with attendees attempting to exit the area while avoiding congestion. To experiment, it is necessary to define the variables that will be used by the model. The selected parameters are shown in Table 1.

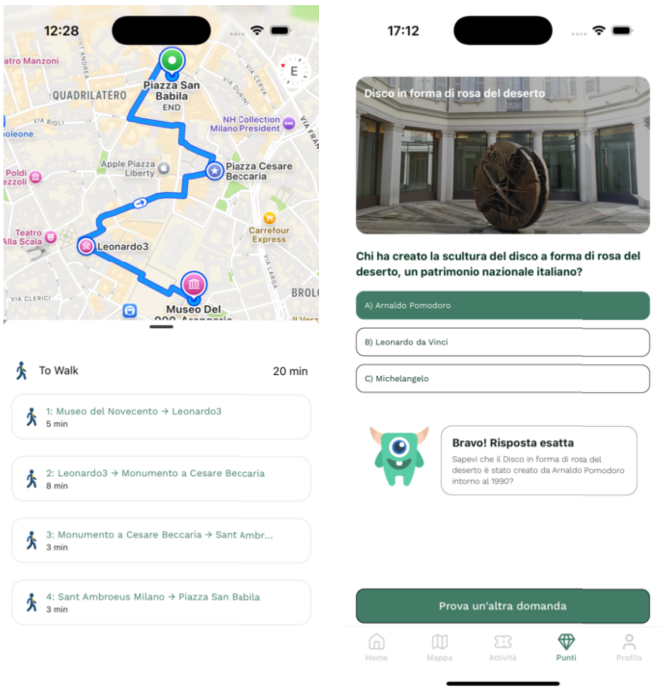
The starting POI is set as *Museo del Novecento*, a well-known cultural site located within walking distance of the square. This POI is chosen due to its central location within the area covered by the supposed event in 'Piazza del Duomo', ensuring that most POIs would be easily accessible from it. Next, the number of itineraries and POI stops per itinerary is defined. Three itineraries were chosen for the study, assuming that a group of people can complete the itineraries together. The time difference between itineraries was set to 5 minutes to create a staggered arrival schedule, particularly useful if multiple itineraries conclude at the same metro station. Lastly, the distance matrix was filtered to include only visits to *Monuments*, *Museums* and *Sculptures*, simulating a scenario where users explicitly choose to visit these types of POIs based on their preferences. An example of the model-generated itinerary is shown below.

Table 1: Defined variables for itinerary generation

Variable	Value
Starting POI	Museo del Novecento
Number of itineraries	3
Number of POIs to be visited	4
Interval (minutes)	5
Additional Information	None
Allowed Types	Monumento, Museo, Scultura

Leg 1: Museo del Novecento → Leonardo3 Museum *6 min*
Leg 2: Leonardo3 Museum → Monument to Cesare Beccaria *7 min*
Leg 3: Monument to Cesare Beccaria → Monument to Saint Ambrose *7 min*
Leg 4: Monument to Saint Ambrose → San Babila metro *2 min*

This itinerary is shown to the user as in Fig. 1a. The model followed all the constraints in the prompt generating exactly three distinct itineraries with a staggered arrival at the metro station of more than 5 minutes, as shown in Table 2.



(a) The itinerary **(b)** Example MCQ
Figure 1: Eventour interface examples.

Table 2: Itineraries Generated for the User Study

Itinerary	Locations	Total Time
Itinerary #1	Museo del Novecento → Cordusio	28 min
Itinerary #2	Museo del Novecento → Cordusio	35 min
Itinerary #3	Museo del Novecento → San Babila	22 min

4.2 Gamification

Following itinerary generation, the data is parsed using Python to extract POI labels, which are then used in SPARQL queries to retrieve additional information from Wikidata and DBpedia. The resulting dataframe includes each POI's label, construction date, creator, material, heritage status, and abstract. This information is then used to generate multiple-choice questions for the gamification component. Examples of the generated MCQ are given below:

POI: Royal Palace of Milan

Which of the following is true about the Royal Palace of Milan?

A) It is a cultural centre and home to international art exhibitions. (*)
B) It was originally designed to include two courtyards but these were later dismantled to make room for the Duomo.
C) It was abandoned for over two years after World War II and its condition further deteriorated.

Fact: More than 1,500 masterpieces are on display annually at the Royal Palace of Milan.

POI: Monument Federico Borromeo

Who was the architect of the Monument Federico Borromeo?

A) Costantino Corti (*)
B) Leonardo da Vinci
C) Michelangelo

Fact: The Monument Federico Borromeo was built around 1865 by Costantino Corti.

4.3 User Study

To evaluate the approach, real-world tests were conducted in Milan with 10 volunteers. Participants provided demographic and interest data for generating personalised itineraries. Table 3 summarizes their backgrounds.

Table 3: Volunteer demographics, backgrounds, and interests

Vol.	Age	Background	Interests
1	29	Bsc in Economics	Museums-Theaters-Monuments
2	28	Bsc in Medicine	Museums-Theaters-Monuments
3	24	Economics Student	Museums-Theaters-Monuments
4	29	Msc in Sports Science	No Preferences
5	28	Msc in Economics	No Preferences
6	58	Bsc in Economics	Museums-Monuments
7	58	Bsc in Economics	Museums-Monuments
8	26	Msc in Sustainability	No Preferences
9	27	Msc in Data Science	Museums-Monuments
10	27	Bsc in Economics	Museums-Monuments

The volunteers were divided into three groups, all starting from Museo del Novecento. Group 1 consisted of three participants interested exclusively in museums, theatres, and monuments. Group 2 consisted of four individuals focused on museums and monuments, while Group 3 comprised three participants with no specific POI preferences. Each group was given a model-generated itinerary and a set of multiple-choice questions to interact with along the route. The actual time taken to complete each itinerary was recorded and compared to the predicted duration from the distance matrix to evaluate the accuracy of the travel time estimates.

For instance, Group 1, preferring museums, theatres, and monuments, followed the generated itinerary:

- Start: Museo del Novecento
- Stop 1: Teatro Gerolamo (5 min)
- Stop 2: Monument to Federico Borromeo (11 min)
- Stop 3: Duomo di Milano Museum (8 min)
- Final Stop: Duomo (Metro station, 5 min)

Total expected time: 29 min

The group of three participants completed the itinerary together in an average of 31 minutes, just 2 minutes longer than the estimated duration by the LLM. The results indicate that the estimated travel times were generally accurate, with deviations ranging from 2 to 5 minutes across the groups. These variations can be attributed to factors such as differences in walking speed and brief stops along the way.

4.4 User Experience Questionnaire

To evaluate the user experience and effectiveness of the proposed system, each participant is asked to complete a questionnaire inspired by the System Usability Scale (SUS). This questionnaire is adapted to specifically assess aspects such as ease of use, engagement, and overall satisfaction with the itineraries and gamification elements.

The results indicate a generally positive reception of the system. Ease of use received the highest rating (mean: 4.7), with low perceived complexity (mean: 1.5). Participants found the gamification elements engaging (mean: 4.0) and felt encouraged to explore the city further (mean: 4.1). The lowest-rated aspect was the itinerary balance (mean: 3.6), suggesting room for improvement in POI distribution. While satisfaction with the selected POIs was relatively high (mean: 4.1), some participants felt the itineraries could better reflect their preferences (mean: 3.9). Figure 2 presents a box plot visualisation of the questionnaire responses.

4.5 Application development

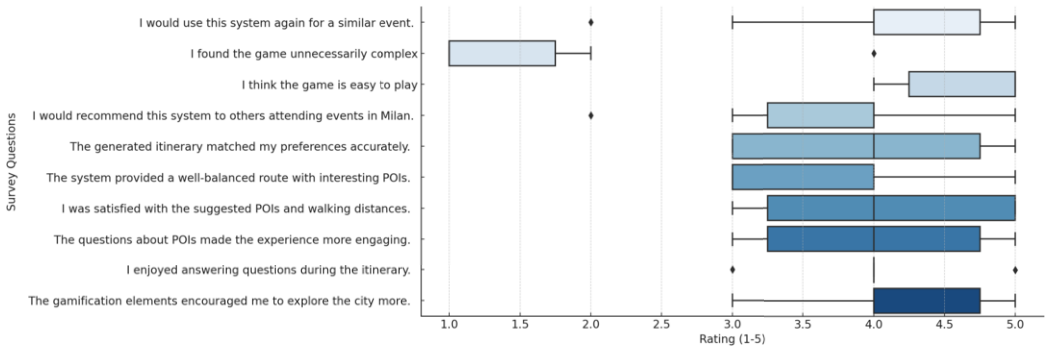
We are currently in the early stages of implementing the mobile application, which will be named *Eventour*. This app will be the practical interface for delivering the personalised itineraries and gamified experiences proposed in this framework. For its implementation, we have selected React Native[29] as the core framework, alongside Expo[30] to simplify and speed up the development workflow. By leveraging React Native, we maintain a single codebase that compiles into fully native iOS and Android apps, eliminating the need to build and maintain two separate codebases.

5 Limitations

While the results of the user experiment were promising, certain limitations must be acknowledged:

- Limited sample size: The user study is conducted with a relatively small number of participants (10), which limits the generalizability of the findings. A sample of this size may not capture the full range of mobility behaviours, preferences, or decision-making patterns observed in broader populations. Additionally, factors such as participants' age, familiarity with the city, and personal interests are not deeply controlled or stratified, which may introduce bias or variability in how the itineraries are evaluated. As a result,

[29] https://reactnative.dev/
[30] https://expo.dev/



Mean	Std Dev	Min	Max
4.0	0.94	2	5
1.5	0.97	1	4
4.7	0.48	4	5
3.8	0.92	2	5
3.9	0.88	3	5
3.6	0.84	3	5
4.1	0.88	3	5
4.0	0.82	3	5
4.0	0.47	3	5
4.1	0.74	3	5

Figure 2: Distribution of responses and participant value perceptions.

the insights drawn from this user study should be interpreted as preliminary and exploratory rather than conclusive.

- Static itineraries: The current implementation generates fixed itineraries based solely on the initial input parameters provided by the user. These itineraries do not account for dynamic contextual factors such as real-time pedestrian flow, temporary closures, weather conditions, or changing user preferences during the journey. As a result, the generated routes may lack flexibility and responsiveness, potentially reducing their practical usefulness in real-world, time-sensitive scenarios.
- Geographical scope: The experiment was limited to a specific area within the city of Milan, which may limit the applicability of the findings to other urban contexts. Characteristics unique to this location, such as its spatial layout, density of POIs, pedestrian infrastructure, and cultural or historical significance, may have influenced both itinerary generation and user perception. As such, the system's performance and effectiveness in other cities or types of environments (e.g., suburban, rural, or highly touristic areas) remain to be tested.
- Gamification depth: Although participants interacted with the gamified elements of the system, the overall depth and complexity of the gamification layer were relatively limited. The existing design lacked a broader range of interactive mechanics, progression systems, or immersive features that could sustain long-term engagement. As a result, the motivational impact may have been short-lived or inconsistent across different user profiles, potentially influencing the level of participation and the quality of feedback gathered during the study.
- Ethical and privacy considerations: The framework, as implemented, raises significant ethical and privacy concerns that were not fully addressed during the study. The use of AI in urban mobility contexts involves the potential handling of sensitive location data, which can implicate user privacy, even if such data is anonymised. Furthermore, the decision-making processes behind itinerary generation may reflect implicit biases in the training data or model logic, potentially leading to skewed or exclusionary suggestions. Additionally, the interface and interaction design may present barriers for non-technical users, limiting the inclusivity and equitable accessibility of the system across diverse user groups.

6 Conclusions and Future Work

This paper explored the application of Generative AI to optimize urban mobility in event-driven scenarios, specifically focusing on mitigating post-event congestion in Milan through personalized itinerary generation and gamified crowd management. The approach combined fine-tuned LLMs, OpenStreetMap-based distance matrices, and a gamification approach to create interactive walking itineraries that guide users to POIs before reaching metro stations. For the implementation, the Mistral 7B model is fine-tuned using LoRA to dynamically generate pedestrian routes and to create multiple-choice questions related to specific POIs by leveraging data from Wikidata and DBpedia. The generated itineraries were tested in a real-world setting, evaluating their impact on both pedestrian flow and user engagement. A usability study using a questionnaire further assessed participants' interaction with the system. The results confirmed that AI-driven itinerary planning can successfully stagger metro arrivals, alleviating congestion while promoting cultural exploration.

While the proposed framework has shown promising results, there are several directions for future research and improvements. One direction is to test different aspects of the enriched knowledge graph. Future iterations could incorporate real-time data sources, such as pedestrian density analytics from computer vision systems or mobile device tracking, to adjust itineraries dynamically and avoid congested areas. Implementing real-time feedback loops would allow users to report congestion, POI relevance, or gamification engagement, enabling the model to adapt dynamically. Personalized recommendations could also evolve based on past behaviors and explicit user reviews, improving long-term usability and effectiveness. The current approach primarily focuses on walking, but expanding the model to incorporate multimodal transport options such as bike-sharing, e-scooters, and shuttle services could provide more flexible and efficient itinerary recommendations. Future improvements could involve collaborative gamification mechanics, where groups engage in challenges, share itineraries, and compete in urban exploration. Social integration features such as leaderboards, achievement-based rewards, and team-based challenges could further enhance user engagement.

Acknowledgements

This work was partly developed within the Spoke 8 — MaaS and and Innovative services of the National Center for Sustainable Mobility (MOST) set up by the "Piano nazionale di ripresa e resilienza (PNRR)" —M4C2, investimento 1.4, "Potenziamento strutture di ricerca e creazione di "campioni nazionali di R&S" su alcune Key Enabling Technologies" funded by the European Union. Project code CN00000023, CUP: D93C22000410001. This work was also partially supported by the MUR under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy. We thank Cristopher Beneduce for his support with dataset integration, Matteo Anedda for his contributions to the initial implementation of the itinerary generation, and Andrea Feliziani for his assistance in developing the app interface.

References

[1] J. Adrian, A. Seyfried, and A. Sieben. Crowds in front of bottlenecks at entrances from the perspective of physics and social psychology. *Journal of the Royal Society Interface*, 17(165):20190871, 2020.

[2] R. Y. Aldahlawi, V. Akbari, and G. Lawson. A systematic review of methodologies for human behavior modelling and routing optimization in large-scale evacuation planning. *International Journal of Disaster Risk Reduction*, page 104638, 2024.

[3] M. A. Alrowaily. Smart emergency alerting system: A machine learning approach. In *Proceedings of the 2024 2nd International Conference on Communications, Computing and Artificial Intelligence*, pages 8–15, 2024.

[4] T. Althoff, R. W. White, and E. Horvitz. Influence of pokémon go on physical activity. *Journal of Medical Internet Research*, 19(4):e315, 2017.

[5] Y. Bao, H. Yang, Z. Gao, and H. Xu. How do pre-event activities alleviate congestion and increase attendees' travel utility and the venue's profit during a special event? *Transportation Research Part B: Methodological*, 173:332–353, 2023.

[6] Y. Bao, Y. Bao, and L. Kang. Modeling the departure time choices of attendees for special events with cruising-for-parking. *Transportation Research Part E: Logistics and Transportation Review*, 198:104103, 2025.

[7] S. Cerezo-Esteve, E. Inglés, J. Segui-Urbaneja, and F. Solanellas. The environmental impact of major sport events (giga, mega and major): A systematic review from 2000 to 2021. *Sustainability*, 14(20):13581, 2022.

[8] F. De Nijs, G. Theocharous, N. Vlassis, M. M. De Weerdt, and M. T. Spaan. Capacity-aware sequential recommendations. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, pages 416–424, 2018.

[9] L. Fayed, G. Nilsson, and N. Geroliminis. On the utilization of dedicated bus lanes for pooled ride-hailing services. *Transportation Research Part B: Methodological*, 169:29–52, 2023.

[10] G. Giuliano and Y. Lu. Analyzing traffic impacts of planned major events. *Transportation research record*, 2675(8):432–442, 2021.

[11] M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, S. Mirjalili, et al. A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints*, 3, 2023.

[12] M. Haklay and P. Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 7(4):12–18, 2008.

[13] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint*, arXiv:2106.09685, 2021.

[14] T. Iwata, T. Otsuka, H. Shimizu, H. Sawada, F. Naya, and N. Ueda. Finding appropriate traffic regulations via graph convolutional networks. *arXiv preprint arXiv:1810.09712*, 2018.

[15] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. Le Scao, T. Lavril, T. Wang, T. Lacroix, and W. El Sayed. Mistral 7b. *arXiv preprint arXiv:2310.06825*, 2023. doi: 10.48550/arXiv.2310.06825.

[16] A. K. Kanaujiya and V. Tiwari. Crowd management and strategies for security and surveillance during the large mass gathering events: The prayagraj kumbh mela 2019 experience. *National Academy science letters*, 45(3):263–273, 2022.

[17] I. Kapenekakis and K. Chorianopoulos. Citizen science for pedestrian cartography: collection and moderation of walkable routes in cities through mobile gamification. *Human-centric Computing and Information Sciences*, 7(1):10, 2017.

[18] Land Transport Authority (LTA). Travel smart journeys gets a boost: New rewards and expanded opt. 2024.

[19] Y. Liang, Y. Liu, X. Wang, and Z. Zhao. Exploring large language models for human mobility prediction under public events. *Computers, Environment and Urban Systems*, 112:102153, 2024.

[20] J. K. Nguyen, A. Karg, A. Valadkhani, and H. McDonald. Predicting individual event attendance with machine learning: a 'step-forward' approach. *Applied Economics*, 54(27):3138–3153, 2022.

[21] D. Quercia, R. Schifanella, and L. M. Aiello. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 116–125, 2014.

[22] M. B. Talukder, M. Hoque, and T. Miah. The impact of global events on hotel occupancy and tourist flows. In *Adapting to Evolving Consumer Experiences in Hospitality and Tourism*, pages 1–30. IGI Global, 2025.

[23] A. Wang, A. Zhang, E. H. Chan, W. Shi, X. Zhou, and Z. Liu. A review of human mobility research based on big data and its implication for smart city development. *ISPRS International Journal of Geo-Information*, 10 (1):13, 2020.