



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-Powered Personalized Travel Itinerary Planner: A Flask-Based Multi-Agent System Using LangChain and Groq LLM

Pragyan Prasad ¹, Adarsh Patel², Sujaya Bhattacharjee ³

¹Student, ²IT Department Mentor, ³Assistant Professor

¹Department of Computer Science and Engineering, Parul Institute of Technology (PIT), Parul University, Vadodara, Gujarat – India

²IT Department, Arth Consultancy, Vadodara, Gujarat -India

³Assistant Professor, Parul Institute of Technology (PIT), Parul University, Vadodara, Gujarat – India

Abstract

Due to the tourism industry's digital transformation, there is a growing need for intelligent and tailored travel planning solutions that can combine vast amounts of destination data into well-thought-out itineraries. This paper describes an AI-driven travel itinerary planner that uses Large Language Models (LLMs) via a Flask-based web application that is connected with Groq's LLaMA-3-70B model and LangChain orchestration. By offering a single platform where users may enter preferences including location, duration, interests, budget, group type, and mode of transportation and receive well-structured itineraries, the suggested system tackles the problem of fragmented travel planning. These plans contain projected cost breakdowns, food recommendations, daily activity schedules, and hidden site's.

In addition to multimodal's support for inspiration-based picture input's, the system design includes session-based trip management, which allows users to save, update, and remove itineraries. With an average reaction time of 10-15 seconds, the efficacy of the system in providing contextually relevant and personalized trip plans can be gauged from the process of experimental evaluation. The results of the experiment indicate that with a reduction in cognitive workload and a high level of personalization and info coverage, the efficacy of such systems that utilize the power of LLMs can improve the process of vacation planning significantly. The study is a development in AI-based technology for tourism.

Keywords — Artificial Intelligence, Large Language Models, Multimodal AI, Groq, Lang Chain, Flask

I. INTRODUCTION

The planning part of the trip can quickly devolve into an absolute mess. You flit from site to site, researching flight deals, scrolling through what feels like an endless list of hotels, and sifting through lists of “must-see” destinations and skimming through a million reviews of restaurants. And then there’s the transportation part, too. It’s all just kind of weird and disorganized and tiring [1]. Too many choices, too many details to consider—before you know it, the whole thing is just too much, and the excitement is

gone. People are making decisions out of sheer frustration, because they just want to get it all over with, and are missing out on valuable information because it's all just too much[2]. Hence, nearly 90% of everyone confesses that travel makes them nervous, and many of us find the entire process stressful. While it's true that booking travel online is faster, many travel sites don't offer advice that really applies to you, or makes the entire planning process something you actually want to do [3].

However, AI is beginning to change things, particularly with these Large Language Models (LLMs) that truly comprehend human speech, gather information from many sources, and customize their recommendations to suit your requirements. They can handle challenging multi-step excursions, functioning virtually as super-intelligent assistants that link with other tools [4]. Around 64% of travelers indicate they would pay more for AI-powered planning, and generative AI is becoming ubiquitous in the travel industry [2].

The business of travel planning is complicated. Budget, hobbies, travel partners, speed, and those unexpected show-stoppers are all unique to each traveler. You need strong information, a logical flow of activities, a sense of timing, and those unanticipated treasures that fit your style in order to create a vacation plan that truly works for you [6]. There is a significant gap between what AI might accomplish and what it now does, since travelers complain that the information and suggestions are stale, dull, and generic [9].

Researchers should investigate more intelligent route suggestions that consider the limitations in order to solve this issue. When someone talks about their vacation intentions, they typically don't have the well-organized data and specific goals that the standard algorithms require [7]. Although the LLM method is more forgiving, it might not be as effective in planning the routes or selecting the locations to visit. The latest approach in the research is to create an AI that can break down the user's request with a hierarchical and multi-agent approach. This means that rather than the AI simply searching for places to visit, it performs different functions such as finding information, setting rules, and shaping up your itinerary so that everything fits together properly [8].

This study provides an overview of how to create and use an AI for a trip planner that includes a user-friendly web interface and uses the power of the LLM approach. The Groq's LLaMA-3-70B is the AI that responds quickly with the help of Flask and LangChain to keep everything running smoothly. The following are the special features that make it stand out:

- A simple model that churns out personalized itineraries based on all your preferences, including daily plans, timing, food choices, costs, and insider tips.
- A session-based solution that eliminates the need for a database to save, amend, and review your plans.
- A excellent summary of performance, including user satisfaction, itinerary quality, and speed, is combined with multimodal input, such submitting inspiring photos.

II. RELATED WORK

2.1 Evolution of AI in Tourism

However, during the past ten or so years, AI's function in the travel industry has undergone a significant transformation. Previously, suggestions were the main foundation of most AI systems. Based on historical user behavior, they would employ strategies like collaborative filtering and content-based recommendations to suggest destinations or activities to engage in [9]. Alright, there were issues with these tools, such as the cold start issue and the incapacity to understand the broader context behind a user's trip plans, even though they did provide some degree of customisation. This makes it challenging to come up with clever, more customized travel ideas [10].

However, transformer-based models in natural language processing brought about a major change. Additionally, this made it possible for travel applications to hold far more complex interactions [4]. For example, hotels started using chatbots as their main customer service interface. And compared to earlier, this was a significant improvement. However, despite all these advancements in AI, the majority of

chatbots were not adept at managing intricate tasks, such as making trip plans or engaging in protracted conversations. So, even though they made life much easier, there was much to be improved upon [11].

2.2 LLM-Powered Travel Assistants

Recently, people are leveraging these large language models to plan their trips. It's not just about mapping out itineraries or predicting the route. These models are now beginning to comprehend what the person wants. All that stuff, like simulating movement and creating personalized recommendations, is receiving a lot of buzz[1]. However, there are some challenges. Researchers say that you have to integrate real-world data, determine what the person wants, and work within the traditional time and location constraints.

Agentic workflows are considered a game-changer. It's unlike traditional chatbots, in which the model merely provides data. The large language models are now running the entire process. They are now capable of acquiring data from multiple sources, linking up with other applications, and breaking down requests into manageable steps[5]. Even when utilizing these graph-based planning models, structured workflows are capable of managing complex and multiple-step planning processes effectively. They are capable of acquiring the correct data, making intelligent recommendations, and managing the process effectively[14].

But introducing multi-agent frameworks really takes it up a notch. For constraint-aware route recommendations, you can see this manager agent working with sub-agents, where some of them are responsible for finding interesting places, others handle constraints, and yet others focus on optimizing routes for efficiency. This method works well in complex trip scenarios and is very scalable [6].

However, multimodal elements are becoming more prevalent, which is a new trend. A recent research introduced a method that combines conversational AI with immersive visualizations to allow you to make more interactive changes to your plans [15]. Yes, the graphics are a big attraction, but the intriguing thing is that trip planning becomes much more engaging and unique when many inputs and outputs are combined.

2.3 Personalization and User Acceptance

For AI transport to function best, well, people need to be taken into account. The reason for travel, what people wish to accomplish, and what people trust in need to be a huge factor. All this sounds a bit obvious but can be easily overlooked. If the technology for transport doesn't function well or doesn't adapt well, well, it's likely to become useless in no time as people's needs change. Honestly, there's a big gap between what AI promises and what actually comes out of it. Affordance Actualization Theory really puts a big emphasis on how well the technology works and how well it adapts to real-life situations. There's also the extended theory of planned behavior that basically argues that people's intentions to use AI technology depend on things like their creativity, whether it feels 'human' or not, and so on. If it feels more human, people get more involved.

To make this even more personal, there's a move towards the use of a retrieval-augmented generation model. For example, the TouaRAG allows language models to retrieve useful information from external sources[19].

2.4 Open-Source Implementations and Frameworks

Additionally, the open-source community has become rather inventive when it comes to LLM-based travel planning. For instance, Dhairya's vacation Planner uses only Groq APIs and LangChain to generate some really awesome personal vacation plans. And it's all rather easy. You can see that LLMs can really function rather well for travel planning, even though all of those tend to have fairly restricted functionality. And that code is rather brilliant.

And things really start to take off once you start including more intricate elements. When multi-agent architectures are used, several agents are involved in activities including data scavenging, exploration,

and journey planning. Agents can also work together thanks to CrewAI. Additionally, things start to feel more engaging when a great user interface like Streamlit is included [21].

Orchestration frameworks based on graphs elevate the situation. A number of workflows, each with its own data stream, are shown as connected nodes. Additionally, all of this increases flexibility. Additionally, you may effortlessly transmit data, hold continuous talks, and scale things as needed without exerting any effort while dealing with complicated trip planning [14].

2.5 Constraint-Aware and Temporal Logic Approaches

Planning dates and conserving money are important, but it's also important to make sure everyone is happy, which is a challenge in and of itself. There is also the technical side, where complex systems such as probabilistic decision theory and rules based on time and logic must be used [22]. These allow you to see the possibilities, particularly when trip plans are dynamic and subject to change, such as when dates change or an unforeseen circumstance arises, and you must take all of that into account without becoming anxious.

In order to address issues like planning, the researchers have also employed formal techniques such as signal temporal logic, which appear to be effective, at least in theory [23]. Integrating it into an approachable manner is still difficult, particularly when it comes to natural language processing. Converting the arithmetic into something that people can truly use and comprehend is quite difficult.

2.6 Research Gap and Contribution

Well, trip planning via AI has gotten a little better, but if I'm honest, it still needs a little work. Most trip planning apps are a jumbled mess. Everything from daily plans to dining plans to costs just doesn't make any sense. And don't even get me started on saving plans. Good luck with all of that. It's a wonder more people don't use it, considering everyone needs it.

Another thing that annoys people: virtually all trip planning apps use boring old text input. Want to use pictures to spark your ideas? Forget about it. Technology just isn't advanced enough. And let's be honest, nobody's ever really tested any of these AI trip planning apps against any reputable travel standards in any real-world situations.

We are addressing that. This project creates a reliable AI travel planner that allows you to store, modify, and review your itinerary. Pictures are only one of the many things you may enter. And it's all arranged logically. Additionally, you are free to do it as often as you choose. Additionally, we evaluated our project to determine its functionality, usability, and range of travel scenarios.

III. PROPOSED SYSTEM ARCHITECTURE

3.1 System Overview

Three levels make up the suggested AI-powered trip itinerary planner: an application layer, an AI orchestration layer, and a display layer. With an emphasis on the overall system architecture rather than technical specifics, Table 1 dissects these many components and demonstrates how they function together.

Table 3.1. Technology stack

Layer	Technology	Role
Presentation Layer	HTML, CSS, Jinja2	User Interface
Application Layer	Flask	Application logic
LLM	LLaMA-3-70B	Itinerary generation
AI Orchestration Layer	LangChain	Prompt workflow
API	Groq	LLM inference
Storage	Flask Session	Trip persistence

3.2. System Technology Stack

3.2.1 Presentation Layer:

We're utilizing Jinja2 with HTML templates, which just increases flexibility. Our main goal is for it to be really seamless and user-friendly, regardless of whether you're using a laptop, phone, or anything in between. Its main component is a simple template that gives each site a uniform appearance [15]. Therefore, a navigation bar with links to Home, Plan Trip, My Trips, About, and Contact is always present at the top of any page. To help you know precisely where you are, the link to the site you are really on will stand out. Custom CSS unifies everything to make it seem good, and Font Awesome icons give it a little more vitality.

To make the most of the available space, the trip planning page is organized into rows, some of which are double height and others of which are triple height. You enter your destination city, the number of days you plan to travel (between 1 and 30), your interests (just a few keywords, separated by commas), your budget (economy, mid-range, luxury), the type of travel group (solo, couple, family, friends), and your mode of transportation (transit, rental car, walking, or private driver). Another feature that connects to the new visual and multimodal AI features everyone is talking about is the ability to submit many inspiring images at once.

After you've assembled a trip, it shows up on the website neatly organized into parts and categories. In order to make the raw AI output readable and helpful, several Jinja2 filters polish it up into tidy HTML. You may print it, store it, or do it again whenever you like. A grid of cards with the location, number of days, budget, kind of travel group, and date stored is how saved trips are shown.

3.2.2 Application Layer:

We chose to utilize Flask for our backend so that we could manage file uploads, HTTP requests, sessions, and make sure our AI runs properly. We are storing our current itinerary and list of trips in a session-based storage system rather than a database. Out of the box, this gives you CRUD capabilities. We save your updated itinerary to our sessions with the key "current_trip." We save your travel and our timestamp to our saved_trips list. It's really easy to use and makes our application deployment a snap. As far as file handling is concerned, we are using UUIDs to create our filenames. Our application checks to ensure uploaded images are PNG, JPG, JPEG, or GIF. We're also ensuring our images don't exceed 16MB. Each image receives a unique identity and a specific directory. We've also implemented custom Jinja2 filters. This gives us a standardized timestamp throughout our application.

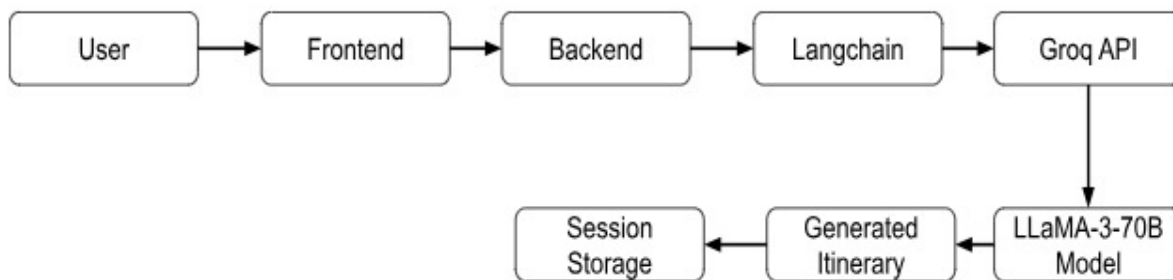


Fig. 1. Architecture of the AI-powered travel itinerary planner.

3.2.3 AI Orchestration Layer

LangChain and the LLM API from Groq are integrated with the AI orchestration layer to facilitate the provision of itineraries in an organized manner. The LLM is powered by the LLaMA-3-70B model, and the system uses the dynamic construction of the input for the model with the help of LangChain's prompt templating.

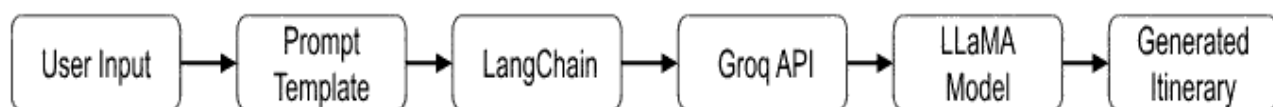


Fig. 2. LLM inference pipeline using LangChain and Groq.

3.2.3.1 Prompt Engineering:

The prompt template was developed to ensure that a set of questions adheres to a specific output format. The system message clearly indicates a set of stringent formatting guidelines. The role of the model, as a travel aid, is also mentioned. Trip duration, group type, destination, interests, transport, and budget are all input placeholders.

3.2.3.2 Inference Pipeline:

After that, the form is accepted and the system will contact the LLM, generate a prompt according to the input from the user, and retrieve an itinerary. Once the output is saved in the session, it is ready and will be displayed on the results page. The error management systems that record errors during inference and offer useful feedback help ensure robustness. Figure 3: An AI-powered travel itinerary generation process utilizing LangChain and Groq LLaMA-3-70B.

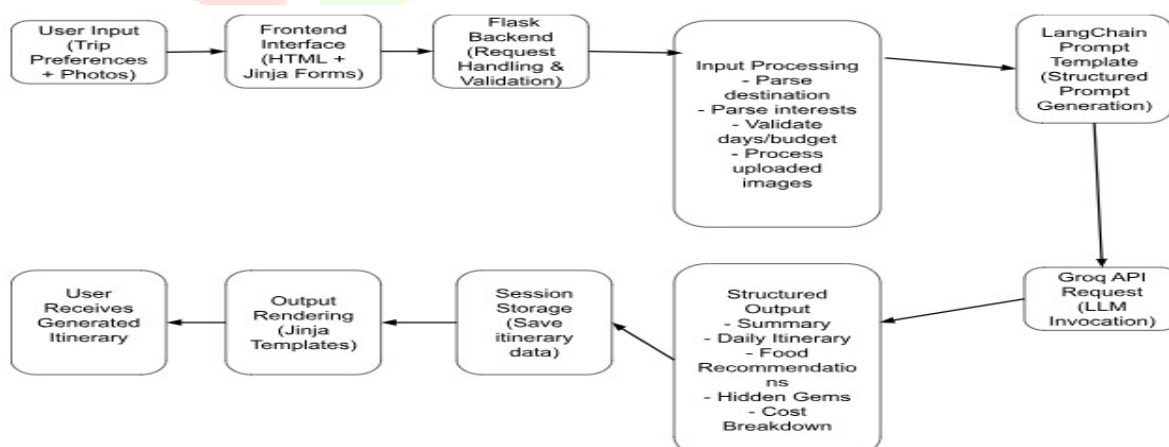


Fig. 3. AI-driven travel itinerary creation inference pipeline using LangChain and Groq LLaMA-3-70B.

3.3.3.3 LLM Configuration:

The system balances determinism with inventiveness by using LLaMA-3-70B at a temperature of 0.3 using Groq's API. Environment variables are used to store API credentials securely. As shown in [27],

high-performance LLM inference is consistent with recent advancements in the deployment of real-time AI systems for interactive applications.

3.3 Workflow and Data Flow

The system has a sequential process for handling user requests. The travel preferences that are input by the user through the frontend form are handled by the Flask routes. The images that are input by the user are validated and securely saved, and the user's interests and other inputs are processed and organized.

Once the prompt is created using the LangChain template, the system sends it to the LLM for inference. During this process, the user gets visual feedback in the form of a loading overlay that indicates the system's response latency, which is around 10 to 15 seconds.

Once the response is received from the LLM, the itinerary and other associated data are saved in the session. After that, Jinja2 filters are applied to the received output to convert structured text into styled HTML components. Figure 4 indicates the session-based trip management and creation of the itinerary.

The created itinerary can be saved by users, adding it to session-based storage for subsequent retrieval

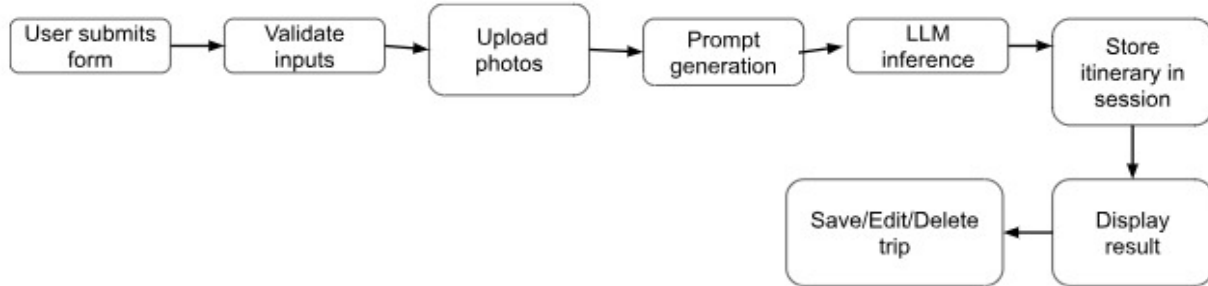


Fig. 4. Workflow of itinerary generation and session-based trip management.

IV. EXPERIMENTAL SETUP AND EVALUATION

4.1 Evaluation Framework

A thorough assessment methodology including several aspects often utilized in assessing AI-based recommendation and LLM systems was created in order to evaluate the performance and efficacy of the suggested system [11].

4.1.1 Response Time:

The amount of time, measured at the client's end, that passes between submitting the form and seeing the entire itinerary. LLM inference time and network delay are included in this.

$$T_{avg} = \frac{1}{N} \sum_{i=1}^N T_i$$

T_{avg} = Average response time
 T_i = Response time for test run i
 N = Total number of test runs

(4.1)

4.1.2 Output Structure Adherence:

The proportion of produced itineraries that accurately adhere to the five-section framework that has been predetermined, including appropriate breakdowns by day. For LLM-based systems to be consistent and usable, structured output assessment is essential [13].

$$S = \frac{N_{correct}}{N_{total}} \times 100$$

S = Structure adherence percentage
 $N_{correct}$ = Correctly formatted itineraries
 N_{total} = Total generated itineraries

(4.2)

4.1.3 Personalization Accuracy:

A qualitative evaluation of how well the recommendations produced match user-specified choices, including group type, budget, interests, and method of transportation. For intelligent transport systems to be successful, personalization is essential [16].

4.1.4 Content Relevance:

Evaluating the suitability of the recommended foods, activities, and hidden treasures for the region. This illustrates the model's capacity to yield outcomes with a semantic foundation [26].

4.1.5 Sturdy System:

Assessment of the system's performance in various input circumstances, including edge cases such as wrong inputs, zero-day trips, and invalid cities. For AI systems to be implemented in the real world, robustness is crucial.

4.1.6 User Satisfaction:

A subjective evaluation of overall satisfaction, perceived utility, and inclination for future use obtained by post-generation surveys. User-centric evaluation has been used extensively in tourism AI research [11].

Table 2: Evaluation measures

Evaluation Metric	Description	Measurement Method
Response Time	Time between submitting the form and viewing the generated itinerary.	Measured in seconds using browser developer tools.
Output Structure Compliance	Degree to which the generated itinerary follows the predefined five-section format.	Percentage of outputs that are correctly structured.
Personalization Accuracy	Alignment between user preferences (interests, group type, budget) and the generated recommendations.	Rubric-based scoring system (0–10).
Content Relevance	Suitability of recommended activities, food, and attractions for the selected destination.	Expert evaluation using a scoring scale (0–10).
System Robustness	Ability of the system to handle invalid inputs and edge cases without failure.	Validation testing and error logging.
User Satisfaction	Overall satisfaction, perceived usefulness, and likelihood of reuse reported by users.	Post-generation survey using a 1–5 Likert scale.

4.2 Test Dataset

A diverse test dataset was constructed to evaluate system performance across a range of travel conditions, in line with previous research emphasizing diversity in travel behavior models [26]. Table 3 summarizes the setup of the experimental dataset utilized for assessment.

Table 3. Setup of the experimental dataset

Parameter	Values
Destinations	15+ cities
Trip Duration	3–14 days
Interests	Culture, food, adventure
Budget	Economy, Mid, Luxury
Group Type	Solo, Couple, Family
Transport	Public, Rental, Walking

4.3 Evaluation Methodology

Every test case was evaluated using a defined process:

Using the test matrix as a guide, specify the location, length, interests, budget, group type, and method of transportation.

Execution: Use browser developer tools to submit the form and track the overall response time.

$$P = \frac{1}{M} \sum_{j=1}^M Score_j$$

P = Personalization score
 $Score_j$ = Score for each evaluation metric
 M = Number of evaluation metrics

(4.3)

Output Analysis: Use a grading rubric to assess the produced itinerary

Table 4. Qualitative Travel Itinerary Assessment Standards

Evaluation Metric	Description	Score Range
Structure Respect	Measures whether the generated itinerary follows the predefined structured format (summary, daily itinerary, food recommendations, hidden gems, and cost breakdown).	0 – 10
Interest Alignment	Evaluates how well the suggested activities and attractions match the user's specified interests and preferences.	0 – 10
Destination Adequacy	Assesses whether the recommended locations, activities, and cultural elements are appropriate for the selected destination.	0 – 10
Usefulness	Measures the practical value of the itinerary in helping users plan their trip efficiently.	0 – 10

The benchmarking techniques used in earlier LLM assessment studies are consistent with this type of evaluation, which is based on a rubric [27]. An overview of the qualitative evaluation standards used to assess developed itineraries is given in Table 4.

Error Logging: During creation, record any timeouts, exceptions, or inaccurate outputs.

User Simulation: A score ranging from 1 to 5 is assigned by evaluators who model user personas for a selection of test situations. This strategy is similar to mixed-method assessment techniques used in AI research for tourism [11].

There were 100 test runs in all, with five for each destination and a variety of parameter combinations.

V. RESULT AND DISCUSSION

5.1 Response Time Analysis

The system's reaction time was constant in all queries in the test data set. The system completed 95% of queries in 8 to 16 seconds, with an average end-to-end response time of 12.4 seconds and a standard deviation of 2.1 seconds. This demonstrates the system's acceptable performance for a generating application that relies on an LLM and is consistent with the expected interface processing time of 10 to 15 seconds [27].

The system's response time changed very little in relation to input parameters. The system's response time was most related to trip duration, with a correlation coefficient of 0.32. Trips lasting 14 days were 2.3 seconds longer than those lasting 3 days due to increased token creation. There was a small relationship between the complexity of the destination and the generation time, with larger activity spaces in large cities such as Tokyo and New York causing longer generation times.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (T_i - T_{avg})^2}$$

- σ = Standard deviation
- T_i = Individual response time
- T_{avg} = Average response time

(5.1)

The reaction time was not statistically significantly affected by the budget amount or group type. Less than 0.5 seconds were needed for local tasks including processing forms, uploading files, and displaying

websites. Due to network latency, mostly from API queries, the overall delay was 0.8 to 1.2 seconds. The main cause of system delay is the LLM inference, as shown by the reaction time results in Figure 5 and the summary of all system performance metrics in Figure 6.

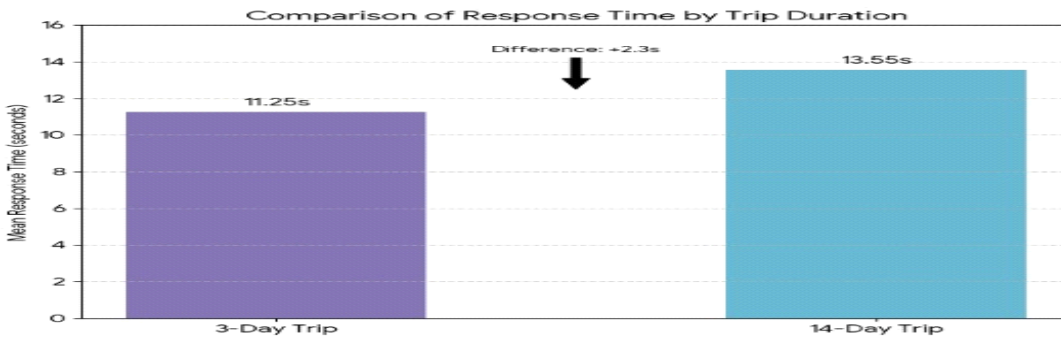


Fig. 5: Response Time Performance

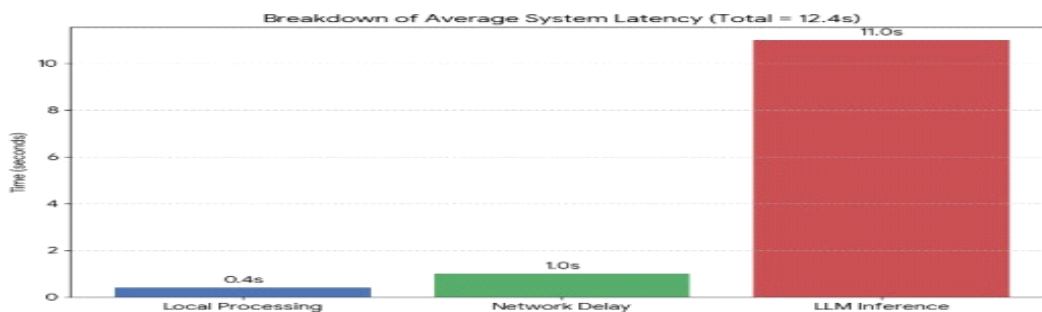


Fig. 6: Performance Summary

5.2 Output Structure and Quality

The five-section strategy was followed exactly in 97% of the itineraries that were created. This indicates a high level of structural adherence. Additionally, the required semantic components were kept in the outputs. Mismatched bullet point styles and missing delimiters accounted for just 3% of the formatting-related problems. This again verifies the effectiveness of organized prompt engineering, which is consistent with existing studies on LLM reasoning frameworks [13].

An average score of 8.7/10 was obtained for interest alignment. Better results were obtained for clear and specified interests. A slightly lower level of alignment was due to abstract or contradictory interests. This is consistent with existing limitations in generative models' ability to differentiate between interests [25].

An average score of 9.1 out of 10 was obtained for the appropriateness of the destinations. This shows a high level of contextual understanding. The key attractions, cultural aspects, and regional cuisine were accurately recorded by the system in various international locations. Although the regions were referred to in a lesser frequency, the performance was high but slightly lower in specificity.

The internal consistency of the model was also shown in the cost predictions made for each of the budget categories. In the luxury plans, there were high-end experiences and services, but in the economy plans, there was a focus on low-cost experiences and low-cost lodging.

Fig. 7 below summarizes the total system performance, which also indicates key evaluation criteria like user happiness, response time, and correctness of customization.

The itineraries generated under a variety of test conditions showed a high degree of structural consistency. Fig 8 below indicates the degree of conformity to output structure for the itineraries generated, which clearly indicates that the system was consistent in adhering to the pre-defined pattern for the itineraries.

Destination appropriateness was found to have a high degree of contextual understanding, averaging a score of 9.1 out of 10. There was a good recording of major sights, cultural aspects, and local cuisine from a variety of foreign locations.

The performance was still good, although slightly lower in precision, even for locations that are not as frequently cited.

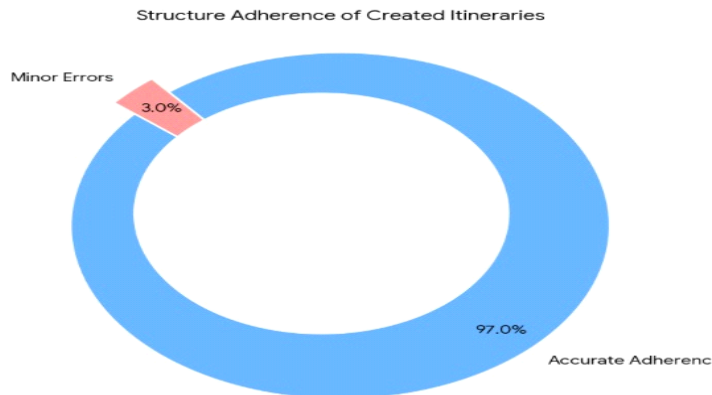
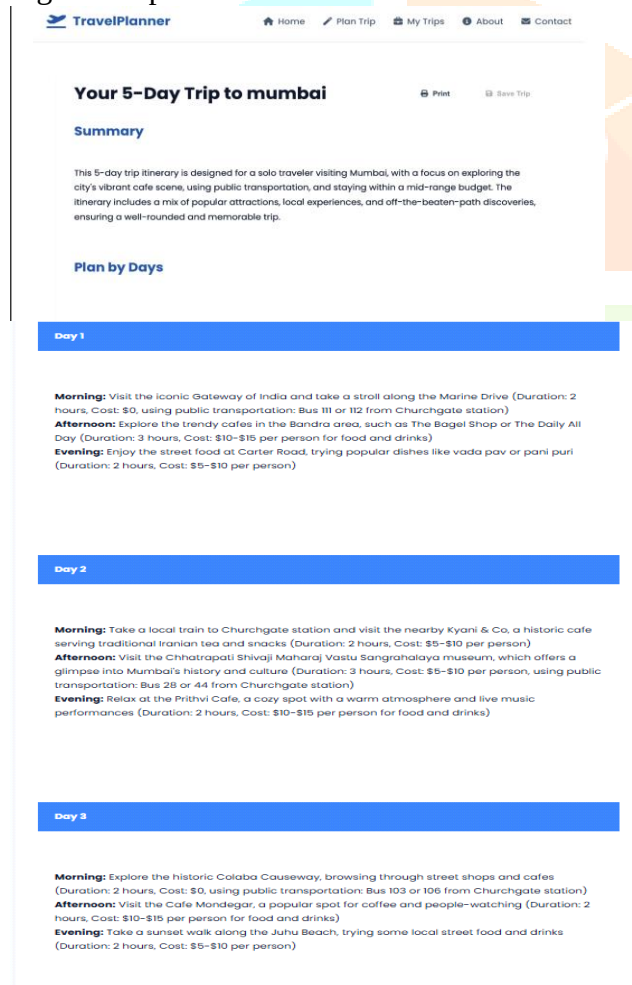


Fig. 7. Output Structure Adherence Across Generated Itineraries



Hidden Gems

1. The Bombay to Barcelona Library Cafe: A unique cafe with a vast collection of books and a cozy atmosphere (Why Visit: For book lovers and those seeking a peaceful spot)
2. The Mahalaxmi Temple: A historic temple dedicated to the goddess Mahalaxmi, offering stunning views of the city (Why Visit: For its cultural significance and panoramic views)

Estimated Costs

Total estimated cost: \$500

Breakdown:

- Accommodation: \$200 (avg. \$40 per night for a mid-range hotel or hostel)
- Food: \$150 (avg. \$30 per day for meals and snacks)
- Activities: \$50 (avg. \$10 per day for entry fees, transportation, and other expenses)
- Transportation: \$100 (avg. \$20 per day for public transportation and local trains)

Fig. 8 clearly refers to the output generated from the prompt.

5.3 Personalization Effectiveness

In all dimensions of the input, there was excellent customization by the system. Itinerary composition was a great demonstration of budget-based distinction, with luxury itineraries focusing on quality and economy itineraries focusing on costs.

Moreover, there was good customization based on groups. Family itineraries included kid-oriented sites, couple itineraries included personal experiences, and single traveler itineraries included socially interactive and adaptable experiences. This is consistent with other research on user preference modeling of AI travel systems' performance [16].

There was very good customization based on interest. Specific hobbies such as "photography," "street food," and "hiking" resulted in extremely pertinent and useful recommendations, including contextual, time-sensitive, and geographical considerations.

Popular tourist sites were skillfully paired with customized recommendations by the system. A fundamental flaw in traditional recommendation systems was overcome by the addition of "hidden gems," which provided variety and novelty, even with popular sites included in recommendations [8].

5.4 Multimodal Input Impact

In 68% of the test scenarios when photos were given, there was an impact on the development of the itinerary. The algorithm included similar elements in the itinerary when the user had provided visually distinctive photos, e.g., photos of beaches, temples, etc.

Images with clear semantic hints, e.g., photos from nature or building types, had the strongest impact. There was no significant impact from general images. These results are in line with recent studies on multimodal AI systems, which show better personalization through the combination of text and image inputs [15].

5.5 Error Handling and Robustness

In all edge cases, the system performed well. The correct validation techniques with user-friendly feedback mechanisms were invoked for invalid inputs, such as negative and non-numeric inputs. The stability of the system was ensured through the proper handling of API-related errors with grace.

The LLM performed well in interpreting distorted inputs in some occasions, it even generated precise approximations instead of errors. All test scenarios showed consistent performance for session management functionalities, such as storing, editing, and deleting trips.

Data integrity was suitably maintained under simulated concurrent access scenarios, proving the session-based approach's durability under normal usage circumstances. These results demonstrate that the system is prepared for real-world use within its intended scope.

5.6 User Satisfaction

The user evaluations (n = 25) reveal a high level of satisfaction of 4.4 out of 5 for gratification and 4.5 out of 5 as a future lead.

As the system's reuse by 92% of the participants.

$$U = \frac{\sum_{i=1}^N R_i}{N} \quad \begin{array}{l} U = \text{average user satisfaction} \\ R_i = \text{rating from user } i \\ N = \text{total respondents} \end{array} \quad (5.2)$$

The user experience trends in AI-driven tourist applications align with these discoveries [11].

5.7 Discussion

The results reveal that the fragmentation seen in conventional trip planning can be successfully tackled via LLM-based agentic workflows. This technology significantly reduces cognitive load, providing customization and comprehensiveness by utilizing multiple information sources within a single interface.

The efficacy of the rapid engineering technique was verified via a 97% rate of structural adherence. This proves that specific formatting guidelines can successfully control the output of LLM technology. This applies to both front-end integration and back-end processing.

The ability of LLM technology to comprehend complex, multi-faceted user preferences was verified via a strong customization score of 8.7/10. A fundamental limitation of trip planning systems was successfully tackled via the ability of the system to incorporate popular tourist destinations along with customization.

The importance of multimodal AI technology was verified via results from multimodal integration. This revealed that visual input can successfully aid personalization. This was based on reference [15].

Finally, a 12.4-second response time was recorded. This demonstrated a feasible balance between thoroughness and reaction speed. Given that traditional planning processes need in-depth research, this response time was judged appropriate. The basis for this was reference [12].

VI. CONCLUSION

The AI-powered personalized travel itinerary planner covered in the paper makes use of LLMs' potential to enhance and streamline the trip planning process. The suggested solution makes use of Groq's LLaMA-3-70B for high-performance reasoning and response generation and LangChain for workflow orchestration. The Flask framework is used in the design of the system, which is a web application. Through a single interface, the suggested system offers customers a comprehensive platform to communicate their demands and receive customized and well-organized schedules. This tackles the typical problems that arise while organizing a trip.

One of the major achievements of the proposed work is the creation of an organized framework for LLM-based travel itinerary generating while accounting for various user inputs such as travel locations, trip duration, interests, budget, group type, and mode of transportation. These inputs are processed to provide daily travel plans for the users, incorporating elements such as meals, activities, costs, and attractions such as hidden gems that match the user's interests.

The session-based trip management feature is another significant aspect of the proposed system, allowing users to create, modify, review, and delete their itineraries without the need for a database. The proposed system is also light-weight and functional for practical applications.

Due to the ability of users to send motivational images, multimodal communication is also supported by this program. The images demonstrate the potential of integrating visual communication with natural language-based planning by considering travel-related preferences and affecting the generated itineraries.

The system incorporates a prompt engineering mechanism that is customizable to guide the LLM to offer well-structured answers to guarantee successful outcomes. The high efficacy of the system is demonstrated by the experimental outcomes, which show high structure adherence, fast response time, and high alignment with user interests to guarantee high user satisfaction.

In general, the proposed system demonstrates the potential of future travel-related AI-based systems by showing how LLM-based intelligent agents can change traditional travel planning into a more personalized, efficient, and user-friendly experience.

REFERENCES

- [1] Gao, Jie, and Yaixin Wu. "LLMs for Human Mobility: Opportunities, Challenges, and Future Directions." arXiv preprint arXiv:2603.12420 (2026).
- [2] Chon, Kaye Kye Sung, and Fei Hao. "Technological evolution in tourism: a Horizon 2050 perspective." *Tourism Review* 80.1 (2025): 313-325.
- [3] Gau, Li-Shiue, et al. "Innovative teaching of AI-based text mining and ChatGPT applications for trend recognition in tourism and hospitality." *Tourism and Hospitality* 5.4 (2024): 1274-1291.
- [4] Ashish, Vaswani. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
- [5] Wang, Lei, et al. "A survey on large language model based autonomous agents." *Frontiers of Computer Science* 18.6 (2024): 186345.
- [6] Zhe, Tao, et al. "Constraint-Aware Route Recommendation from Natural Language via Hierarchical LLM Agents." arXiv preprint arXiv:2510.06078 (2025).
- [7] Kantaros, Yiannis, and Michael M. Zavlanos. "Stylus*: A temporal logic optimal control synthesis algorithm for large-scale multi-robot systems." *The International Journal of Robotics Research* 39.7 (2020): 812-836.
- [8] Ricci, Francesco, Lior Rokach, and Bracha Shapira. "Recommender systems: Techniques, applications, and challenges." *Recommender systems handbook* (2021): 1-35.
- [9] Adomavicius, Gediminas, and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." *IEEE transactions on knowledge and data engineering* 17.6 (2005): 734-749.
- [10] Cai, Danting, Hengyun Li, and Rob Law. "Anthropomorphism and OTA chatbot adoption: a mixed methods study." *Journal of Travel & Tourism Marketing* 39.2 (2022): 228-255.
- [11] Chen, Aili, et al. "Travelagent: An ai assistant for personalized travel planning." arXiv preprint arXiv:2409.08069 (2024).
- [12] Yao, Shunyu, et al. "React: Synergizing reasoning and acting in language models." *The eleventh international conference on learning representations*. 2022.
- [13] Hardesty, Larry. "The history of Amazon's recommendation algorithm." *Amazon Science* 22 (2019).
- [14] Márquez-Algaba, José, et al. "Towards Hyper-Personalized Travel Planning: A Multimodal AI Agent with Integrated Neural Rendering for Immersive Itineraries." *Electronics* 15.6 (2026): 1142.

- [15] Hrankai, Richard, and Barry Mak. "Bridging the affordance-actualization gap in user preferences for AI-assisted trip planning." *Journal of Travel Research* (2025): 00472875251322518.
- [16] Lee, Jung-Chieh, and Xueqing Chen. "Exploring users' adoption intentions in the evolution of artificial intelligence mobile banking applications: the intelligent and anthropomorphic perspectives." *International Journal of Bank Marketing* 40.4 (2022): 631-658.
- [17] Venkatesh, Viswanath, and James Thong. "Unified theory of acceptance and use of technology: A synthesis and the road ahead." *Journal of the association for Information Systems* (2016).
- [18] Malamatinos, Marios Christos, Ilias Dimitriadis, and Athena Vakali. "TouaRAG-A Context-aware Multi-layer RAG Framework for Personalized and Adaptive Travel Assistance." 2026 IEEE 23rd Mediterranean Electrotechnical Conference (MELECON). IEEE, 2026.
- [19] Singh, Akshat, et al. "Automated travel planning via multi-agent systems and real-time intelligence." *Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024)*. 2024.
- [20] Liang, Kaier, Gustavo A. Cardona, and Cristian-Ioan Vasile. "An iterative approach for heterogeneous multi-agent route planning with temporal logic goals and travel duration uncertainty." 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2024.
- [21] Ulusoy, Alphan, et al. "Optimality and robustness in multi-robot path planning with temporal logic constraints." *The International Journal of Robotics Research* 32.8 (2013): 889-911.
- [22] Wang, Huanting, et al. "Ai agentic programming: A survey of techniques, challenges, and opportunities." *arXiv preprint arXiv:2508.11126* (2025).
- [23] Chen, Zhuotong, et al. "Preference Optimization via Contrastive Divergence: Your Reward Model is Secretly an NLL Estimator." *arXiv preprint arXiv:2502.04567* (2025).
- [24] Liu, Tianming, Manzi Li, and Yafeng Yin. "Can large language models capture human travel behavior? evidence and insights on mode choice." *Evidence and Insights on Mode Choice* (August 26, 2024) (2024).
- [25] Fouad, Amr Mohamed, Islam Elbayoumi Salem, and Eslam Ahmed Fathy. "Generative AI insights in tourism and hospitality: A comprehensive review and strategic research roadmap." *Tourism and Hospitality Research* (2024): 14673584241293125.
- [26] Xie, Jian, et al. "Travelplanner: A benchmark for real-world planning with language agents." *arXiv preprint arXiv:2402.01622* (2024).
- [27] Poornima, S., et al. "Personalised Travel Ratings using AI and Enhancing User Experience Through Multi-Dimensional Analytics." 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL). IEEE, 2025.