TravelGenie: Your Personalized AI Travel Companion

An AI-Powered RAG Application for Personalized Travel Planning

Group Members:

Full Name - NUID

Saipavan Katta - 002209016

Yashaswini Tadishetty - 002209058

Contents

1. Topic Selection

1.1 Title of the Project

TravelGenie: Your Personalized AI Travel Companion

1.2 Domain

Travel and Tourism — Personalized AI-Based Travel Planning

1.3 Project Overview

TravelGenie is an AI-driven travel planning application designed to deliver personalized travel experiences that leverage *Retrieval-Augmented Generation (RAG)* techniques and *Large Language Models (LLMs)* to provide users with highly personalized travel recommendations and itineraries.

The system architecture is designed around two core components:

a) **Knowledge Retrieval Module:**

   Utilizes a *Pinecone vector database* populated with embeddings generated from travel-related YouTube video content. This module performs semantic search to retrieve relevant information based on user queries.

b) **Real-Time Update Module:**

   Integrates a *Web Search Helper* capable of fetching the latest information directly from the internet, ensuring that users receive the most current travel insights and updates.

The retrieved data is synthesized and refined using an LLM like *GPT-4o*, which tailors the output to the user's preferences, including destination highlights, accommodation suggestions, budget recommendations, and travel tips.

The user interface is built with *Streamlit*, offering an interactive, chatbot-based environment where users can:

- Input travel-related queries.
- Receive dynamic, personalized responses.
- Modify generated travel plans based on feedback.
- Download finalized itineraries as professionally formatted PDF documents.

By integrating retrieval, generation, real-time search, and interactive customization into a single platform, TravelGenie transforms the travel planning experience into an intelligent, efficient, and enjoyable process.

2. Project Goals

The TravelGenie project is designed with a clear set of goals to ensure the successful development of an intelligent, user-friendly travel planning assistant. The objectives are divided into primary and secondary categories to distinguish between the core system requirements and additional enhancements aimed at improving the overall user experience.

2.1 Primary Objectives

- ✓ Develop a Personalized Travel Planner:
  Build an AI-based system that generates customized travel itineraries based on user preferences.
- ✓ Integrate YouTube-Derived Knowledge:
  Collect and preprocess travel information from YouTube videos to create a searchable knowledge base.
- ✓ Implement Retrieval-Augmented Generation (RAG):
  Use a Pinecone vector database to enable semantic retrieval of relevant travel data.
- ✓ Create a Dual-Helper System:
  Develop a RAG Helper for database retrieval and a Web Search Helper for fetching real-time travel updates.
- ✓ Build an Interactive Chatbot Interface:
  Use Streamlit to design a user-friendly chatbot for seamless interaction.

2.2 Secondary Objectives

- ✓ Enable Real-Time Information Updates:
  Integrate live web search to keep travel suggestions up to date.
- ✓ Deliver Personalized Responses:
  Tailor recommendations dynamically based on user inputs and preferences.
- ✓ Facilitate Travel Plan Modification:
  Allow users to modify their itineraries during the session.
- ✓ Offer Downloadable PDF Itineraries:
  Enable users to download finalized travel plans as structured PDF documents.

- ✓ Enhance User Data Management:
  Securely store login and session data using a backend database like Snowflake.

3. Project Methodology

The methodology for the TravelGenie project follows a structured approach to build an AI-based travel planning assistant.
This section outlines the major phases undertaken during the development, from data acquisition to system integration and testing.

3.1 Define the Travel Planning Scope

- The application focuses exclusively on travel planning based on user queries.
- It provides personalized itineraries, accommodation suggestions, local attractions, travel advice, safety tips, and budget estimations.
- All interactions are restricted to travel-related topics, ensuring domain-specific accuracy.

3.2 Data Collection and Preprocessing

- The data for the system is collected from travel-related YouTube videos, including destination guides and travel vlogs.
- The YouTube API is utilized to extract video titles, descriptions, and full transcripts.
- The raw data undergoes preprocessing steps to remove irrelevant content such as advertisements and casual greetings.
- Meaningful information is extracted to enhance semantic searchability.
- Processed text data is then converted into vector embeddings using OpenAI's Text Embedding Ada model.

3.3 Vector Database Setup

- The processed embeddings are stored and managed within the Pinecone vector database.
- Each embedding is associated with relevant metadata, such as video titles and sources, to maintain context.

- The database is semantically indexed to enable efficient and accurate similarity searches, facilitating fast and relevant retrieval of travel-related information during user queries.

3.4 Development of Helper System

- The system incorporates two intelligent modules to enhance information retrieval capabilities.
- The *RAG Helper* is responsible for retrieving semantically relevant content from the Pinecone database based on user queries.
- The *Web Search Helper* complements the RAG Helper by fetching real-time travel updates from external web sources using a web search API such as Tavily.
- Both helpers are integrated in a coordinated manner to synthesize responses, ensuring a balance between stored knowledge and fresh, up-to-date information.
- A query filtering mechanism is also implemented to validate user inputs and ensure only travel-related queries are processed.

3.5 Application Development

- The frontend interface is developed using Streamlit, providing users with an intuitive, chatbot-style platform for interaction.
- Through the interface, users can submit travel queries, receive dynamically generated itineraries, modify travel plans, and download finalized plans as PDFs.
- The backend architecture is built using FastAPI, which manages user request routing, helper system coordination, communication with the language model, and error handling.
- An optional authentication system is integrated to secure user logins and session data, with storage capabilities provided by databases such as Snowflake.

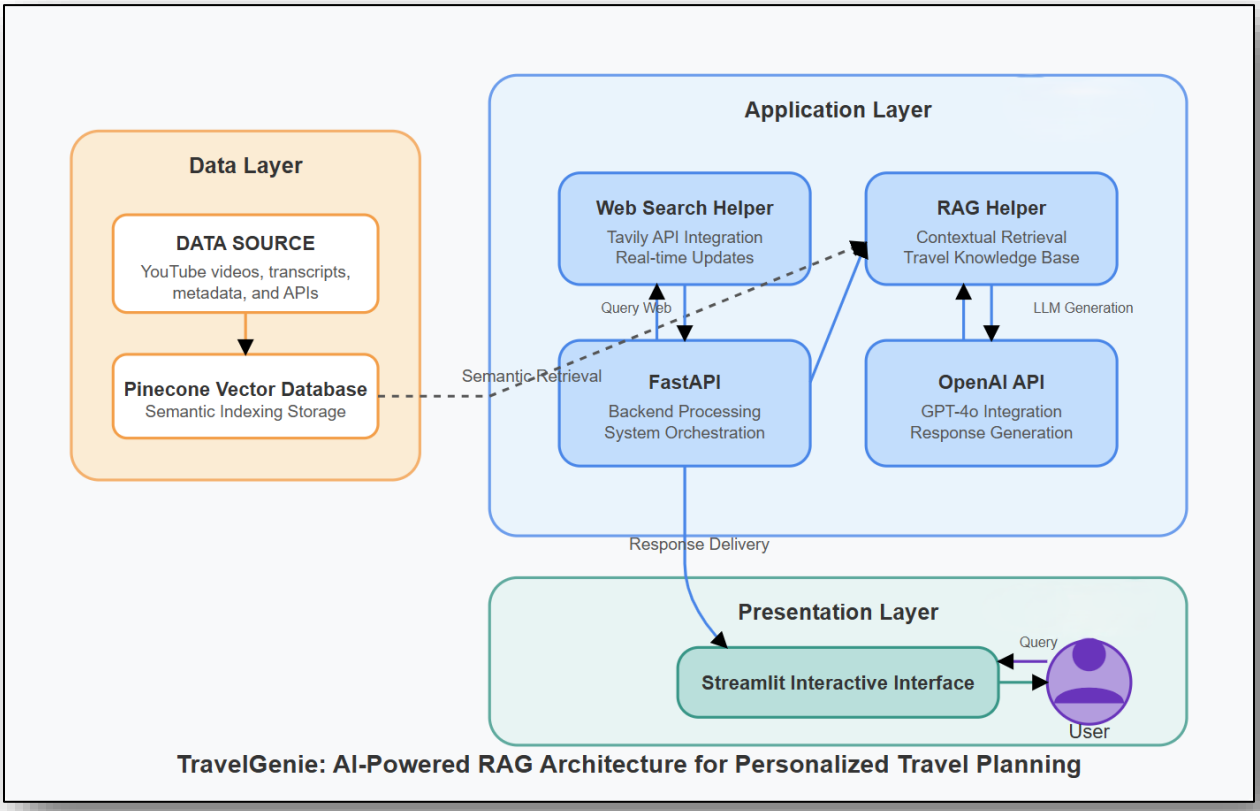
3.6 Evaluation and Testing

- The system's performance is evaluated based on several criteria.
- Accuracy testing measures how effectively the system retrieves and synthesizes relevant travel plans.
- Response time is monitored to ensure prompt feedback to user queries.

- User satisfaction is assessed through feedback collection, focusing on ease of use and quality of generated itineraries.
- Robust testing verifies the system's ability to handle a variety of travel-related queries without loss of functionality or relevance.

3.7 System Architecture

This architecture diagram for TravelGenie shows the complete system with its three main layers:



1. Data Layer

- Data Source: It collects travel-related data like YouTube video transcripts, metadata, and additional content via APIs.
- Pinecone Vector Database: The database stores this data as embeddings (semantic vectors) making it easy to quickly search and retrieve relevant travel information.

2. Application Layer

- Web Search Helper: The Web Search Helper uses Tavily API to get real-time travel updates directly from the web, keeping responses fresh and relevant.
- RAG Helper: The RAG Helper retrieves context-specific information from the Pinecone database to ensure recommendations are personalized and accurate.
- OpenAI API (GPT-4o): It processes user queries, generating human-like, detailed responses tailored to personal travel preferences.
- FastAPI: It acts as the core backend, managing all requests, interactions, and the coordination between the Web Search Helper, RAG Helper, and GPT-4o.

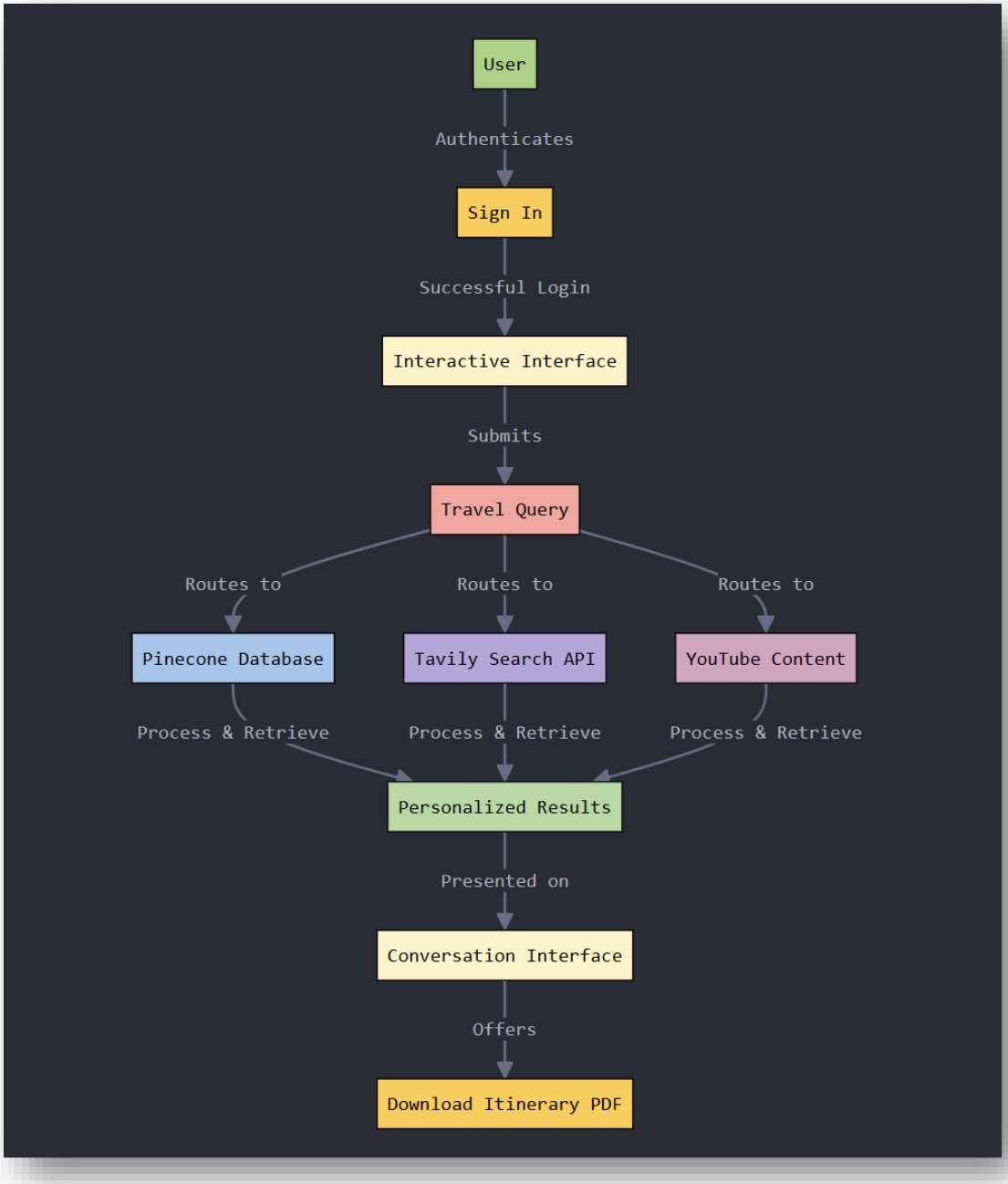
3. Presentation Layer

- Streamlit Interactive Interface: It provides a user-friendly chatbot interface, allowing easy input of travel queries and seamless interaction.
- End User: The end user directly interacts with TravelGenie through the chatbot, receiving personalized travel itineraries and recommendations.

3.8 Flowchart Illustrations

The diagrams illustrate the workflow of the TravelGenie application

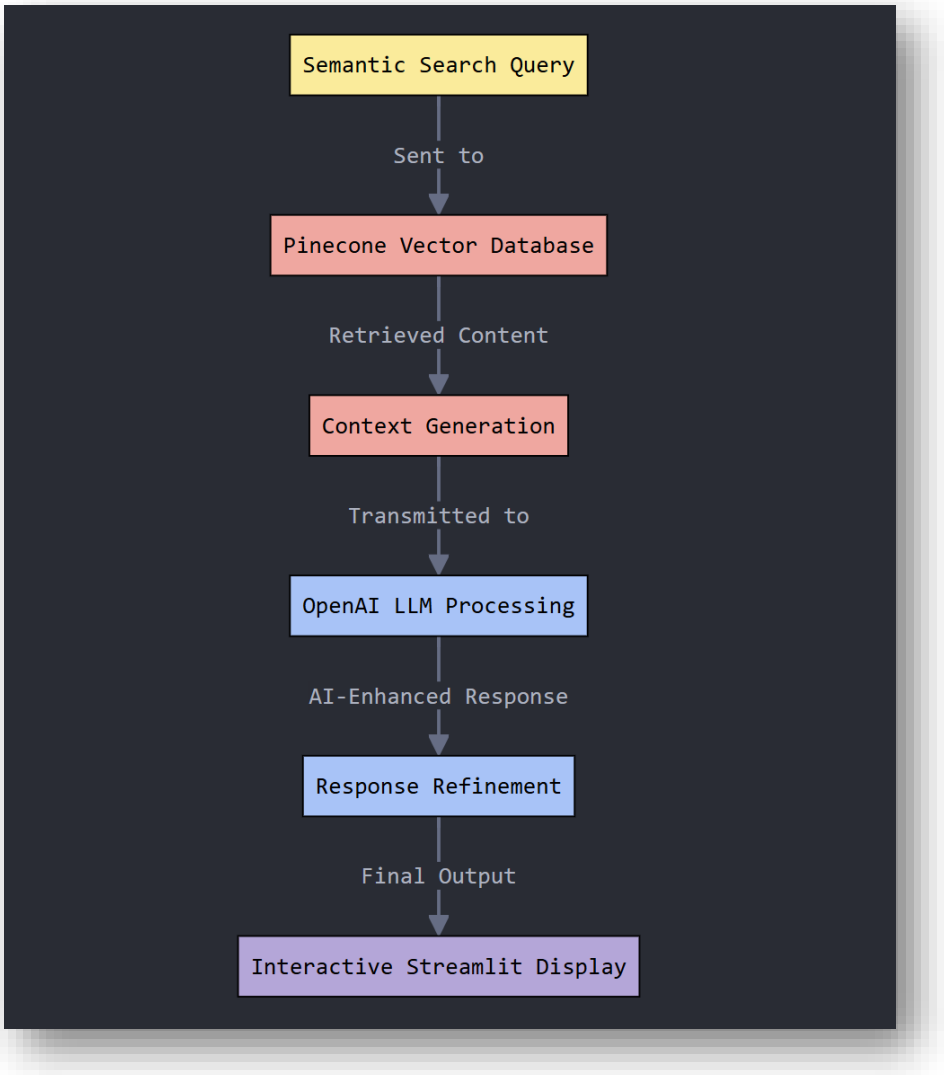
1. System Interaction Flowchart:



The above flowchart demonstrates the complete journey from user authentication to receiving and downloading personalized travel plans:

- User logs in and interacts with TravelGenie's chatbot interface.
- Inputs a travel-related query.
- The system concurrently fetches relevant data from Pinecone, Tavily Web Search API, and YouTube API.
- Combines these results to create personalized travel recommendations.
- Displays the customized results through the conversational interface.
- Provides an option to download the itinerary as a PDF document.

2. Pinecone Semantic Query Flowchart:



This flowchart specifically explains how semantic queries are handled internally:

- User submits a semantic travel query.
- Query is processed in the Pinecone vector database for semantic search.
- Relevant information retrieved from Pinecone is structured into context.
- Context is passed to OpenAI's GPT model to generate tailored responses.
- Generated responses are refined for enhanced clarity.
- Final refined responses are displayed through the Streamlit interface.

4. Tools and Technologies

The development of the TravelGenie project involves a carefully selected set of tools, technologies, and frameworks to ensure robust performance, scalability, and seamless user experience.
Each component plays a significant role in building the system's architecture, enabling efficient data retrieval, natural language generation, and real-time user interaction.

1. Generative AI Models

The project utilizes Large Language Models (LLMs) to understand user queries and generate coherent, contextually appropriate responses.
Specifically, the OpenAI GPT-3.5 Turbo model is integrated to perform both semantic response generation and embedding creation tasks.
In addition, the OpenAI Text Embedding Ada Model is employed to transform textual travel data into dense vector embeddings for storage in the vector database.

2. Databases and APIs

To support fast and scalable semantic search, the project incorporates the following database and API technologies:

- **Pinecone Vector Database:**
  The Pinecone database serves as the primary storage for all generated vector embeddings.
  It enables high-speed, approximate nearest neighbor searches, allowing the system to retrieve the most relevant travel information based on user queries.

- **YouTube API:**
  The YouTube API is used to collect structured travel-related data, including video titles, descriptions, and full transcripts.
  This dataset forms the knowledge base that populates the Pinecone vector database.
- **Web Search API (Tavily):**
  The Web Search Helper integrates with the Tavily API to fetch real-time information from external internet sources.
  This ensures that the system provides users with the latest travel updates, such as current events, safety advisories, and new travel regulations.

3. Frontend Development Tools

The frontend of the application is built using *Streamlit*, a Python-based open-source framework for creating interactive web applications.
Streamlit enables the development of a clean, intuitive, and responsive chatbot interface through which users can interact with the travel planning assistant, submit queries, modify generated plans, and download itineraries.

4. Backend Development Tools

The backend architecture is developed using *FastAPI*, a modern, high-performance Python web framework for building APIs.
FastAPI is responsible for routing user requests, managing interactions between the RAG Helper, Web Search Helper, and LLMs, and ensuring efficient communication between the frontend and backend components.
It also handles error management, session control, and PDF generation functionalities.

Optional Database for User Management

To support login and user session storage functionalities, the system optionally integrates with a *Snowflake database*.

Snowflake provides a secure and scalable platform for storing authentication data, user profiles, and session histories.

5. Introduction

The TravelGenie project addresses the challenges faced by modern travelers in accessing consolidated, reliable, and up-to-date information for trip planning. With vast amounts of travel content scattered across various digital platforms, users often experience delays and confusion in gathering useful insights. This project creates an AI-driven travel assistant that provides personalized, efficient, and responsive travel plans through a centralized, user-friendly system.

5.1 Problem Overview

The traditional method of trip planning involves visiting multiple websites, watching numerous videos, and reading articles to find destination details, accommodation options, activities, and safety tips.
The fragmented nature of this information often leads to inefficiencies and difficulties in decision-making. TravelGenie streamlines this process by integrating diverse travel-related knowledge into a single platform, allowing users to receive well-structured and easily accessible travel plans tailored to their needs.

5.2 Context in Machine Learning

The project leverages advancements in machine learning, particularly semantic search and language generation models, to deliver personalized recommendations. TravelGenie uses vector embeddings to semantically organize travel information, enabling precise matching between user queries and relevant content. By integrating large language models, the system can generate coherent, contextually appropriate responses based on retrieved data. The inclusion of real-time web search capabilities ensures that the recommendations stay updated and relevant to current travel trends and conditions.

5.3 Key Innovations Introduced

TravelGenie distinguishes itself by offering a combination of structured knowledge retrieval, real-time data updates, and user-friendly interaction.
The system supports personalized travel planning, dynamic plan modification, and offline accessibility through downloadable itineraries.
Its design ensures that users benefit from both comprehensive, curated travel knowledge and the latest available information without the need to manually search multiple platforms.

6. Relevance

The TravelGenie project holds significant relevance in the field of digital travel planning, where users often encounter challenges in obtaining trustworthy, consolidated, and personalized information. By intelligently merging pre-existing knowledge with real-time updates, the system addresses a critical gap in how travel information is accessed and utilized.

6.1 Significance of the Topic

The increasing availability of fragmented travel content across websites and digital platforms often leads to information overload for travelers.
TravelGenie offers a solution by consolidating curated and dynamic data, allowing users to receive comprehensive and reliable travel plans through a single, interactive platform.

6.2 Real-World Applications

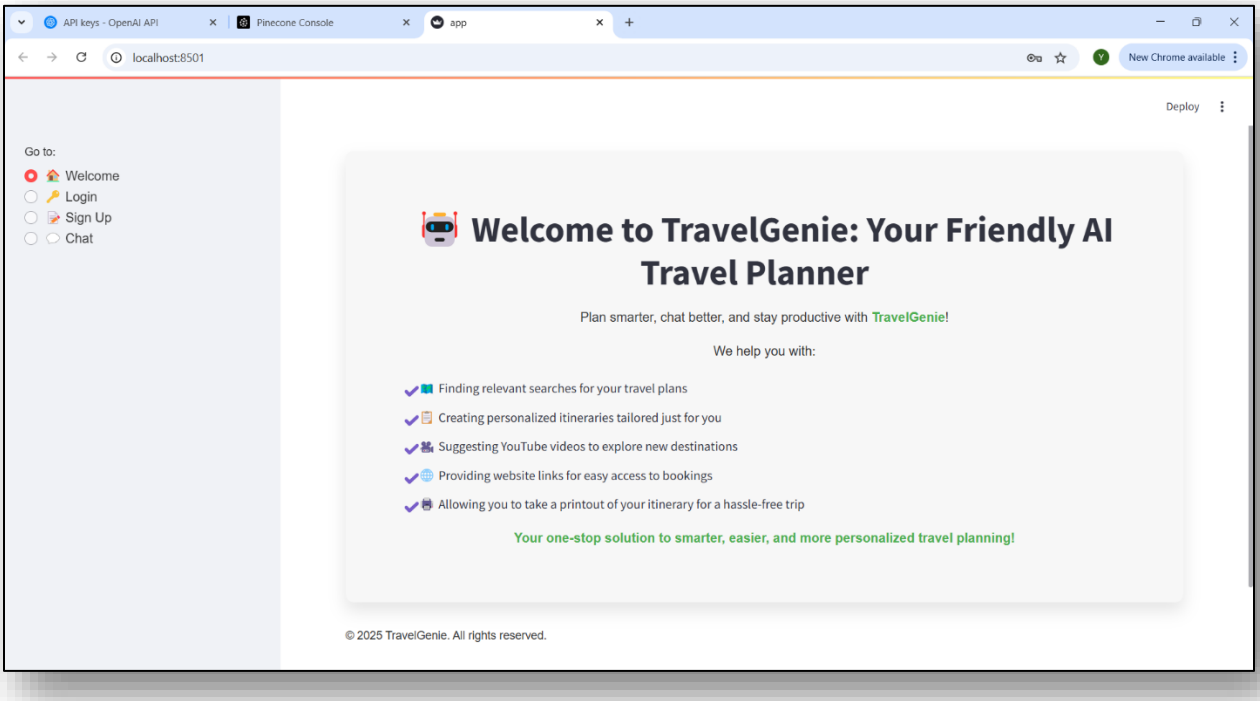
- **Travel Agencies:**
  TravelGenie can help agents quickly generate customized travel plans for clients based on their preferences and budgets.
- **Tourism Boards:**
  The platform can assist in offering interactive, AI-driven travel suggestions for promoting destinations more effectively.
- **Online Travel Platforms:**
  Travel booking websites can integrate similar AI chat assistants to provide smarter itinerary planning tools for users.
- **Customer Service Systems:**
  Travel companies can use AI chatbots like TravelGenie to answer traveler questions instantly about destinations, activities, and safety.
- **Educational Institutions:**
  Hospitality and tourism courses can use the system to teach students about AI applications in travel planning and customer engagement.

7. Application Features and Deliverables

The TravelGenie web application is user-friendly, interactive, and easy to navigate. Below is the step-by-step guide to navigating through the various functionalities clearly:

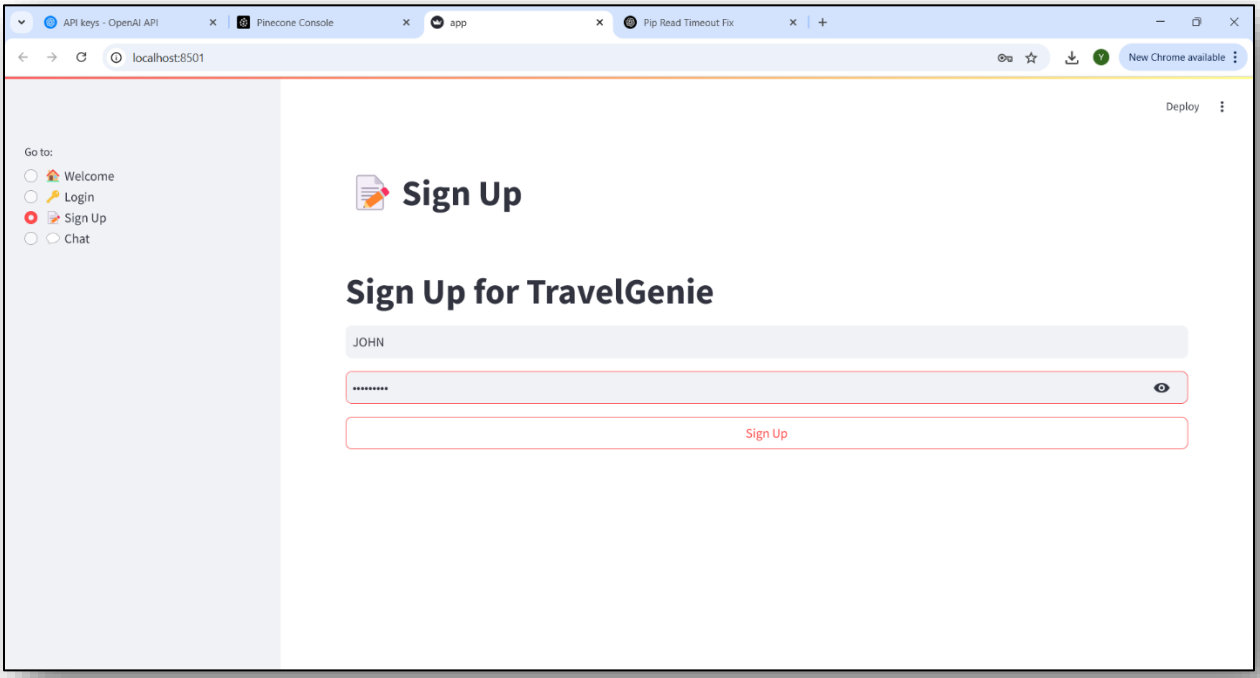
Step 1: Welcome Page

- Users first land on a friendly Welcome Page displaying TravelGenie's main functionalities.
- The sidebar clearly lists navigation options:
    - Welcome (home page overview)
    - Sign Up (create a new account)
    - Login (existing users can log in)
    - Chat (main interaction page for generating travel itineraries)
    - Logout (log out from the account securely)



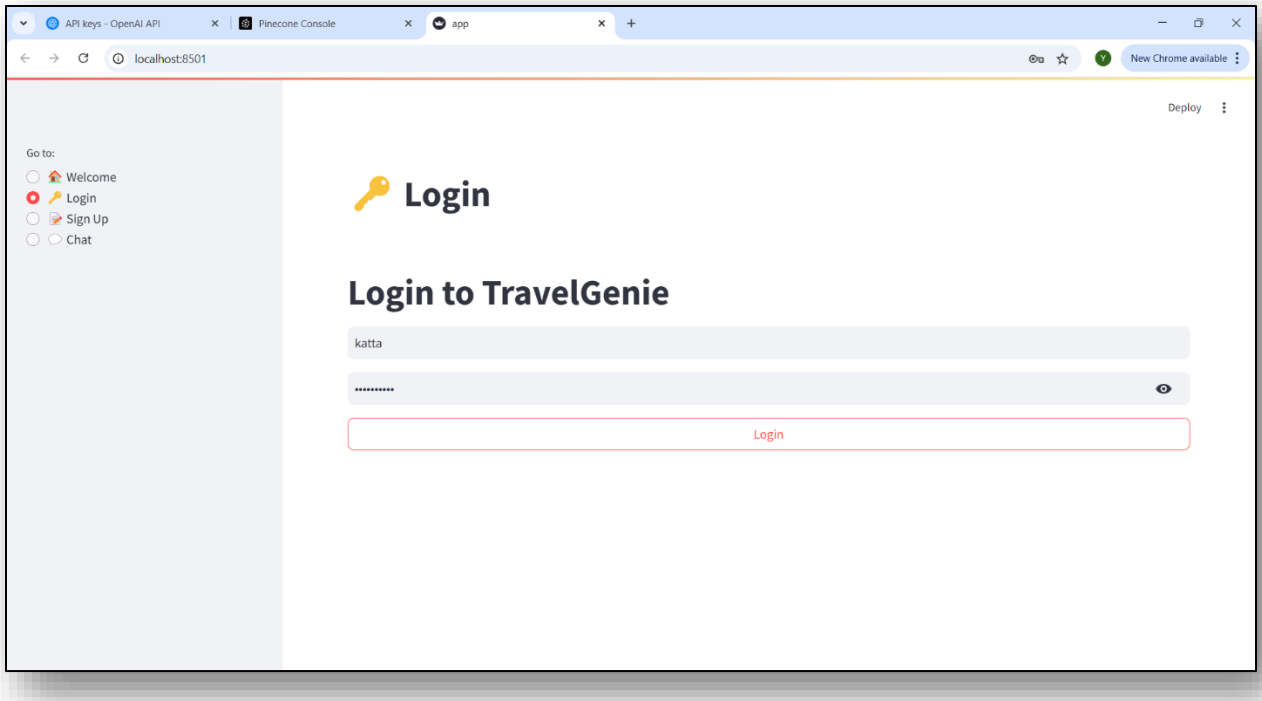
Step 2: Signup Page

- New users navigate to the Sign Up page.
- Users provide necessary details like username and password to create their account.
- After successful signup, users can proceed directly to login.



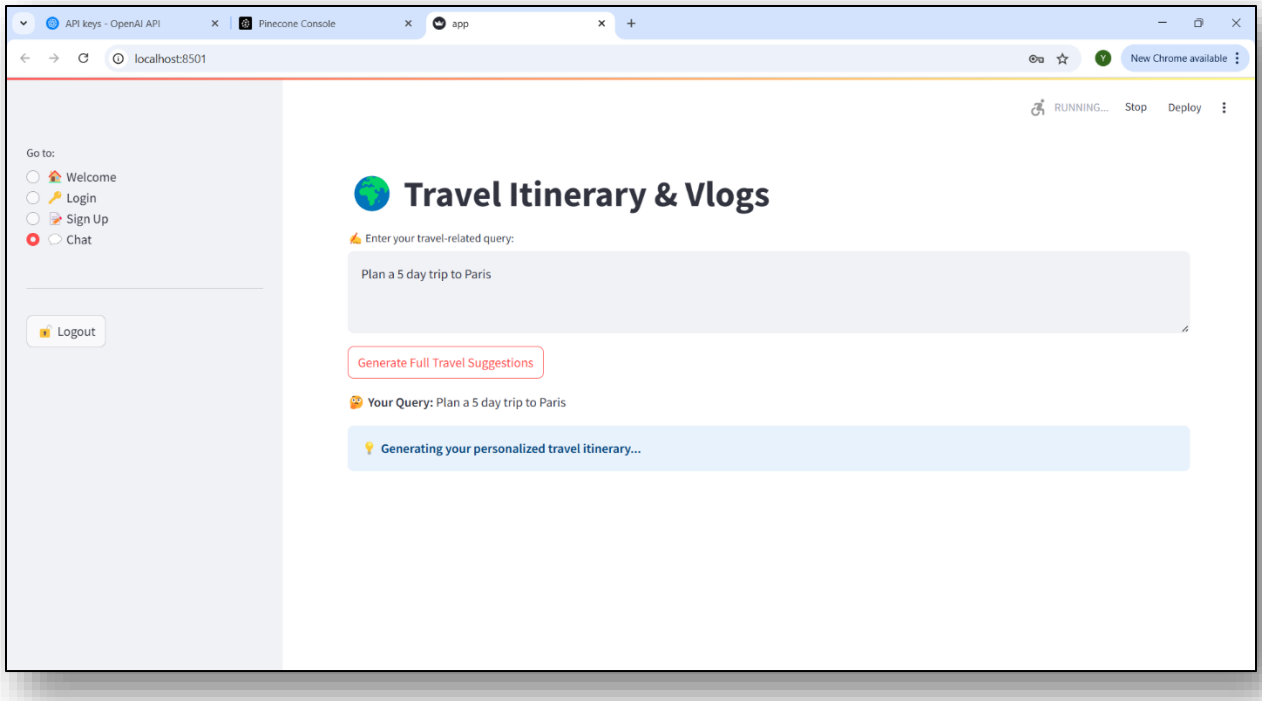
Step 3: Login Page

- Existing users access their accounts through the Login page.
- Users enter the previously registered username and password.
- Successful login redirects the user to the main interactive Chat page.



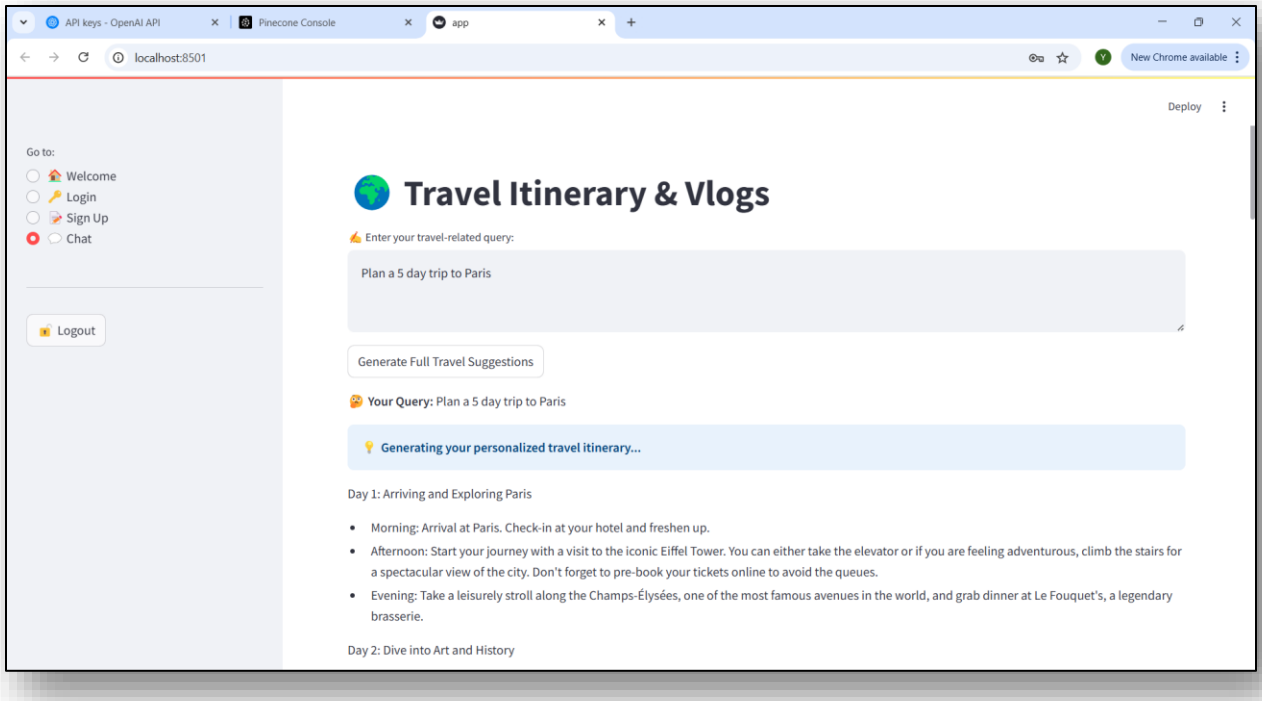
Step 4: Chat Page

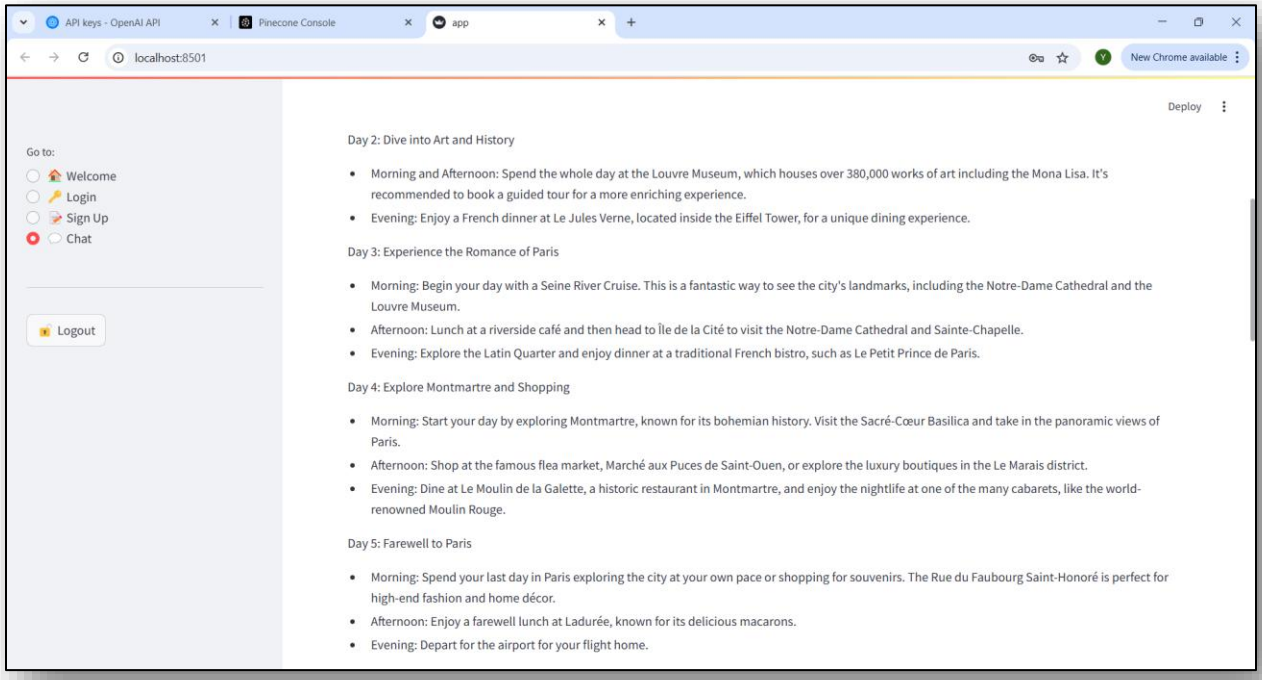
- Users enter their travel-related queries (for example: "Plan a 5-day trip to Paris") in the input box.
- Clicking the "Generate Full Travel Suggestions" button initiates itinerary generation.



Step 5: Personalized Travel Itinerary

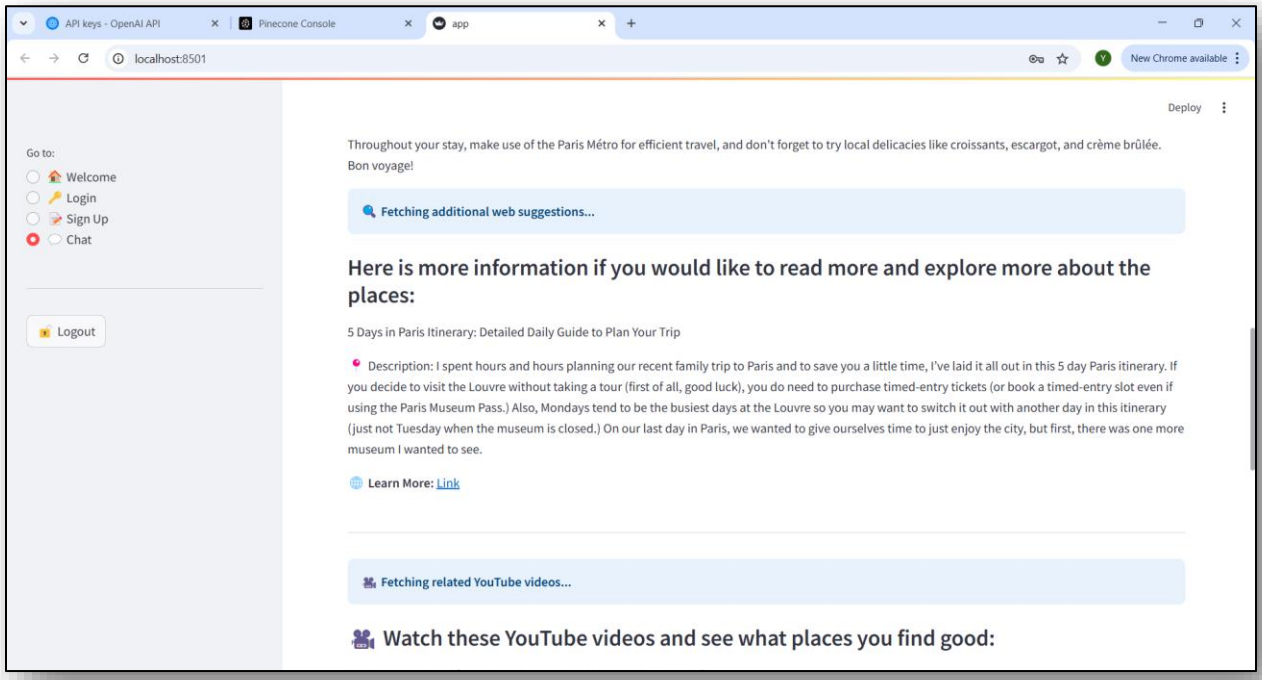
- TravelGenie provides a detailed, personalized itinerary based on the user's input.
- Recommendations include day-wise activities, sightseeing spots, and dining suggestions.

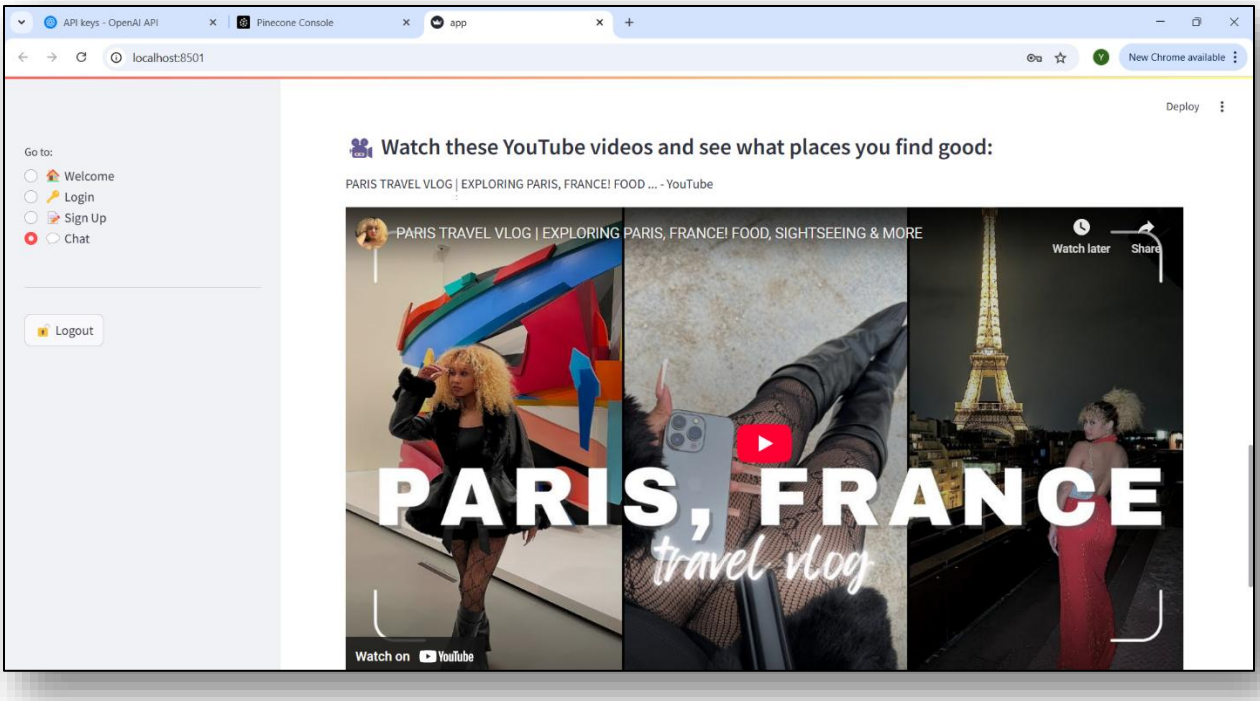

API keys - OpenAI API
Pinecone Console
app
localhost:8501
New Chrome available
Deploy
Go to:
Welcome
Login
Sign Up
Chat
Logout
Travel Itinerary & Vlogs
Enter your travel-related query:
Plan a 5 day trip to Paris
Generate Full Travel Suggestions
Your Query: Plan a 5 day trip to Paris
Generating your personalized travel itinerary...
Day 1: Arriving and Exploring Paris
Morning: Arrival at Paris. Check-in at your hotel and freshen up.
Afternoon: Start your journey with a visit to the iconic Eiffel Tower. You can either take the elevator or if you are feeling adventurous, climb the stairs for a spectacular view of the city. Don't forget to pre-book your tickets online to avoid the queues.
Evening: Take a leisurely stroll along the Champs-Élysées, one of the most famous avenues in the world, and grab dinner at Le Fouquet's, a legendary brasserie.
Day 2: Dive into Art and History


API keys - OpenAI API
Pinecone Console
app
localhost:8501
New Chrome available
Deploy
Go to:
Welcome
Login
Sign Up
Chat
Logout
Day 2: Dive into Art and History
Morning and Afternoon: Spend the whole day at the Louvre Museum, which houses over 380,000 works of art including the Mona Lisa. It's recommended to book a guided tour for a more enriching experience.
Evening: Enjoy a French dinner at Le Jules Verne, located inside the Eiffel Tower, for a unique dining experience.
Day 3: Experience the Romance of Paris
Morning: Begin your day with a Seine River Cruise. This is a fantastic way to see the city's landmarks, including the Notre-Dame Cathedral and the Louvre Museum.
Afternoon: Lunch at a riverside café and then head to Île de la Cité to visit the Notre-Dame Cathedral and Sainte-Chapelle.
Evening: Explore the Latin Quarter and enjoy dinner at a traditional French bistro, such as Le Petit Prince de Paris.
Day 4: Explore Montmartre and Shopping
Morning: Start your day by exploring Montmartre, known for its bohemian history. Visit the Sacré-Cœur Basilica and take in the panoramic views of Paris.
Afternoon: Shop at the famous flea market, Marché aux Puces de Saint-Ouen, or explore the luxury boutiques in the Le Marais district.
Evening: Dine at Le Moulin de la Galette, a historic restaurant in Montmartre, and enjoy the nightlife at one of the many cabarets, like the world-renowned Moulin Rouge.
Day 5: Farewell to Paris
Morning: Spend your last day in Paris exploring the city at your own pace or shopping for souvenirs. The Rue du Faubourg Saint-Honoré is perfect for high-end fashion and home décor.
Afternoon: Enjoy a farewell lunch at Ladurée, known for its delicious macarons.
Evening: Depart for the airport for your flight home.

Step 6: Relevant Additional Content

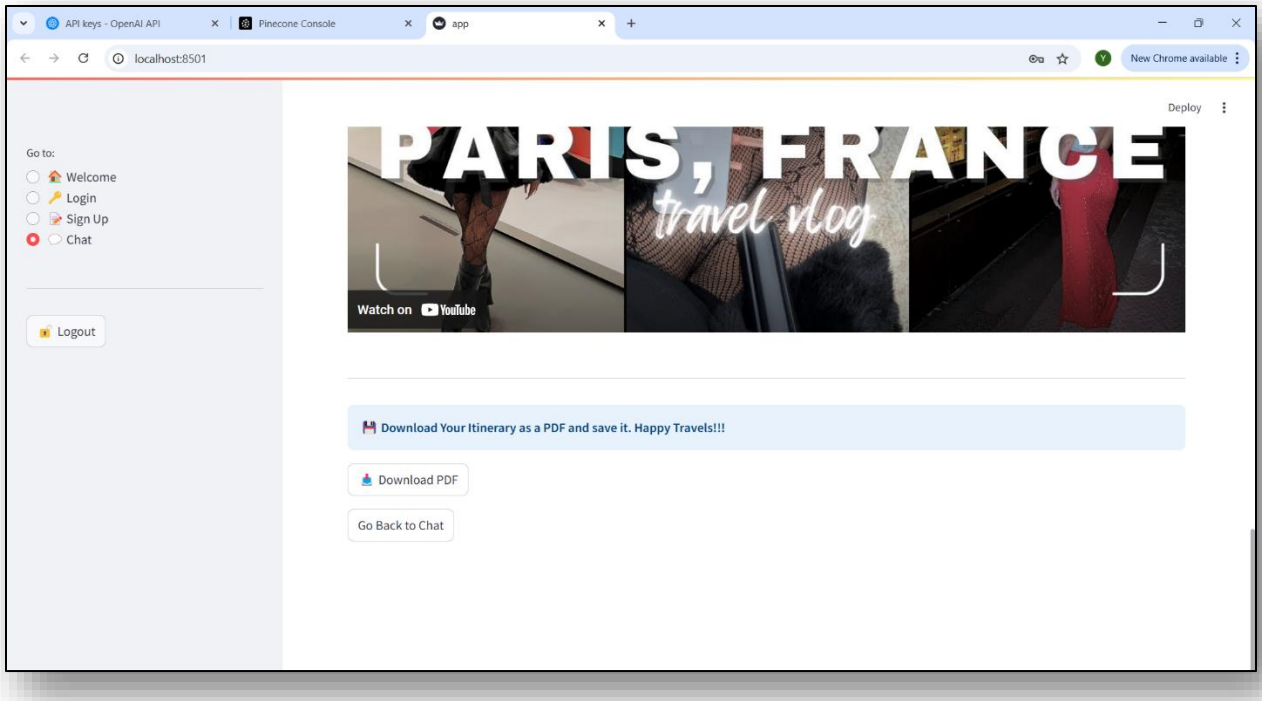
- The application also fetches and displays additional relevant information:
  - Detailed textual descriptions from web sources to help users explore more about the chosen destination.
  - Embedded YouTube videos (such as travel vlogs and destination guides) to visually enhance user exploration.

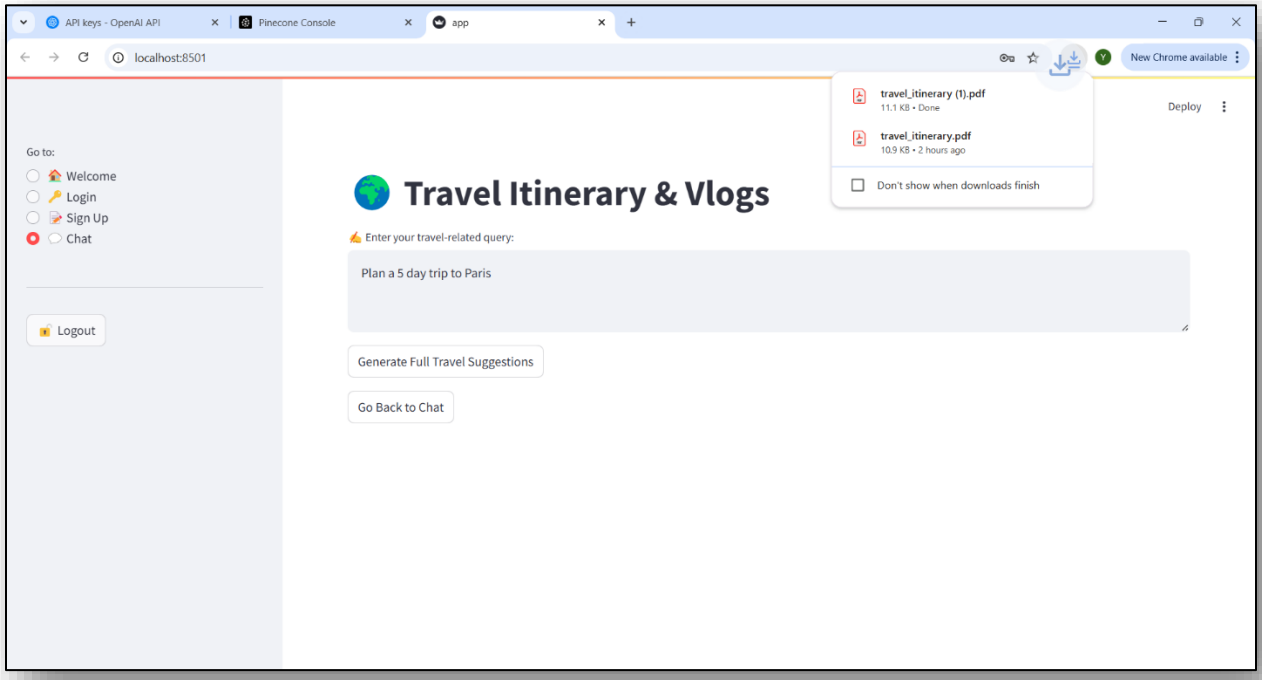




Step 7: Download Itinerary as PDF

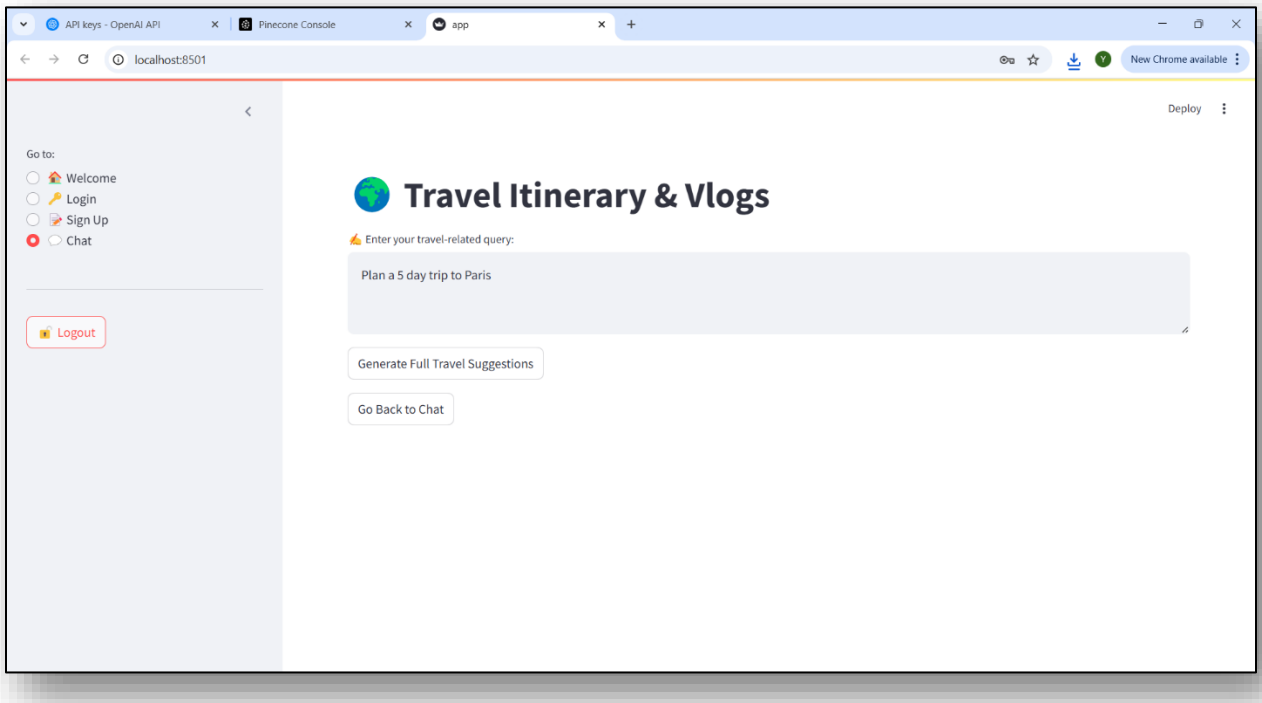
- After itinerary generation, users have the option to download their travel plans.
- Clicking on the "Download PDF" button allows users to save a professionally formatted PDF itinerary for offline access.

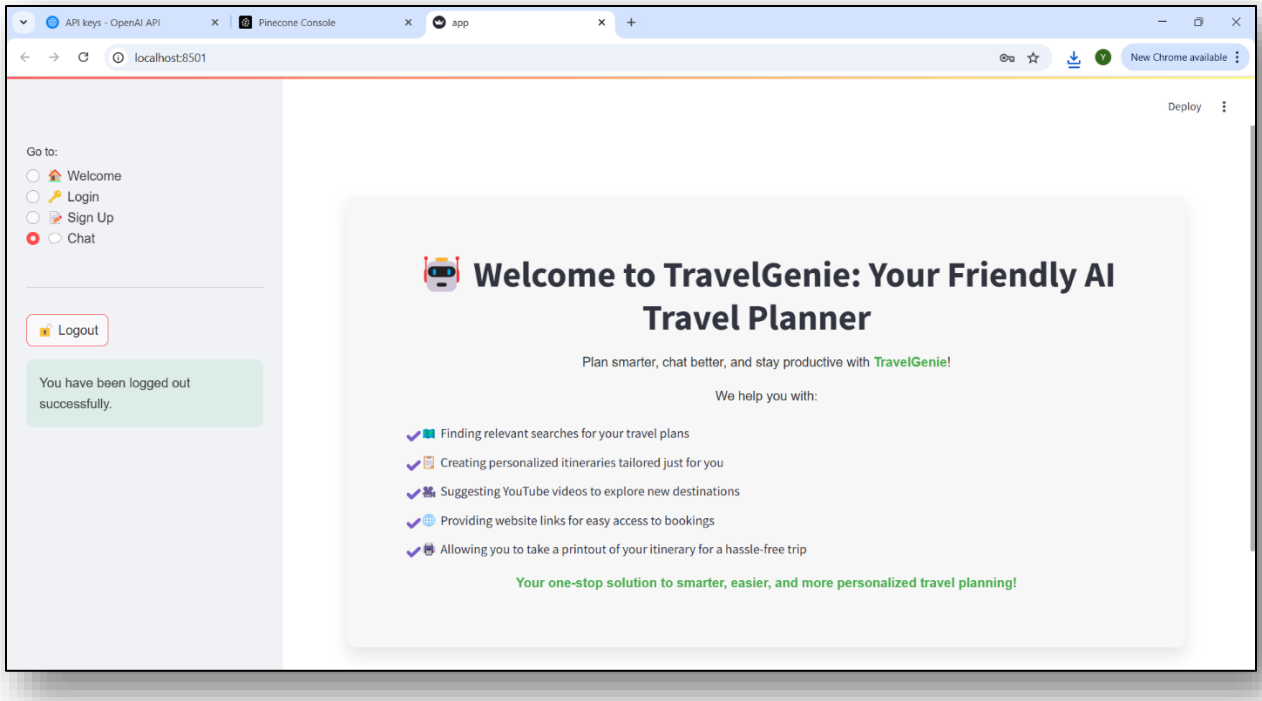

API keys - OpenAI API
Pinecone Console
app
localhost:8501
New Chrome available
Deploy
Go to:
Welcome
Login
Sign Up
Chat
Logout
PARIS, FRANCE
travel vlog
Watch on YouTube
Download Your Itinerary as a PDF and save it. Happy Travels!!!
Download PDF
Go Back to Chat


API keys - OpenAI API
Pinecone Console
app
localhost:8501
New Chrome available
travel_itinerary (1).pdf
11.1 KB • Done
travel_itinerary.pdf
10.9 KB • 2 hours ago
Don't show when downloads finish
Deploy
Go to:
Welcome
Login
Sign Up
Chat
Logout
Travel Itinerary & Vlogs
Enter your travel-related query:
Plan a 5 day trip to Paris
Generate Full Travel Suggestions
Go Back to Chat

Step 8: Logout

- Users securely log out using the "Logout" option from the sidebar.
- Logging out returns the user to the Welcome page, confirming successful logout.





8. Evaluation Metrics

TravelGenie is evaluated based on how well it helps users plan trips quickly, easily, and accurately.

The following points highlight the main areas checked during testing:

i. Accuracy of Responses:
Checks if the system gives correct and relevant travel suggestions based on what users ask.

ii. Speed of Results:
Looks at how fast the system can generate a full travel plan after getting a query.

iii. Quality of Travel Plans:
Reviews if the travel plans are complete, easy to understand, and match what users prefer.

iv. Ease of Using the Chatbot:
Finds out if users can interact smoothly with the chatbot without confusion or difficulties.

v. PDF Itinerary Download:
Makes sure users can download their final travel plans as a PDF file without any problems.

9. Conclusion

The TravelGenie project successfully creates an intelligent travel planning assistant by combining Retrieval-Augmented Generation (RAG), large language models, and real-time web search integration.The platform helps users receive personalized, relevant, and updated travel plans through a simple chatbot interface without the need to manually search multiple websites.

The system focuses on providing a seamless experience where users can generate travel itineraries, adjust their plans, and download final schedules in a clean and organized format. By consolidating information retrieval, personalization, and real-time updates, TravelGenie offers a faster, smarter, and more user-friendly solution to modern travel planning challenges.

The project shows the real-world potential of AI-driven solutions in improving user experiences in the tourism sector, highlighting how technology can simplify complex decision-making processes and offer greater convenience to travelers.

10. References

1) OpenAI API – https://openai.com
2) Pinecone Vector Database – https://www.pinecone.io
3) Streamlit – https://streamlit.io
4) FastAPI – https://fastapi.tiangolo.com
5) YouTube Data API – https://developers.google.com/youtube/v3
6) Tavily Web Search API – https://www.tavily.com