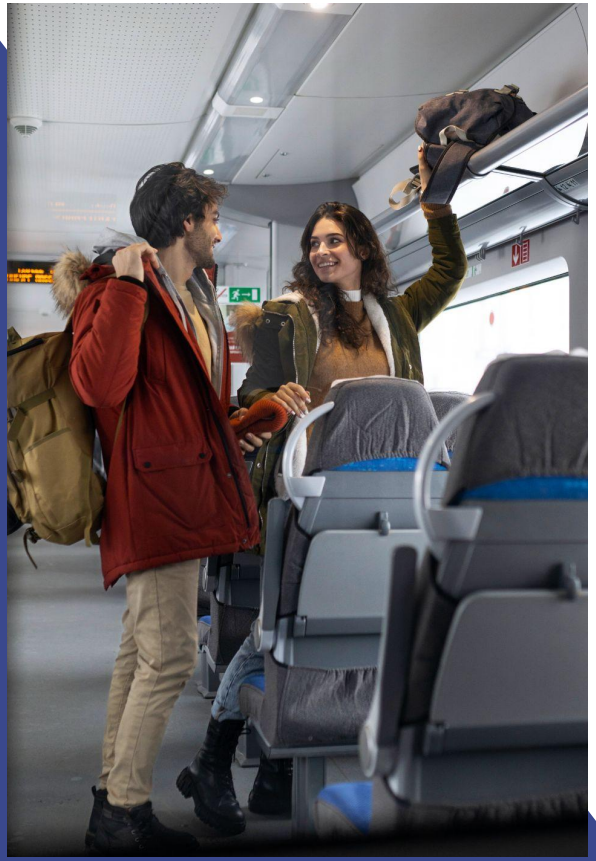
Insight Bot

An AI Assistant

Answering your questions on Trapeze Products and Solutions

Table of contents



01

Objectives of the project

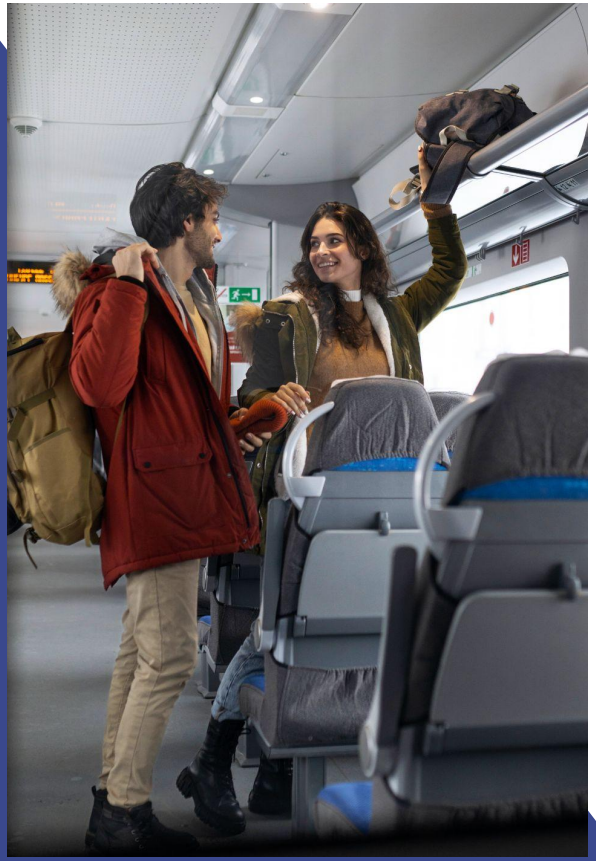
Objectives

Aim

The aim of the project is to create a valuable tool that significantly enhances how users access and interact with Trapeze Group's product information.

The goal

The goal is to implement an AI-powered assistant that streamlines the process of accessing and understanding Trapeze Group's product information, making it more efficient, user-friendly and providing a centralized, intuitive platform for users.



02

Features of the project

Features

Centralize Product Information: Develop an Retrieval Augmented Generation (RAG) system that consolidates and indexes all product and solution information offered by Trapeze Group into a single, easily accessible vector database platform.

Leverage AI Technology: Utilize advanced AI and natural language processing (NLP) capabilities to accurately interpret and deliver precise information from extensive product information sheets.

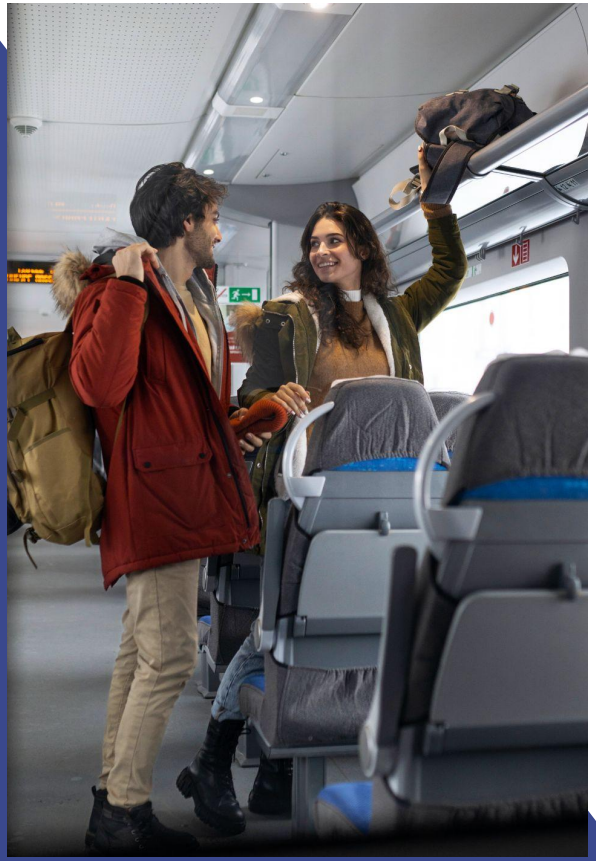
Enhance User Experience: Provide a user-friendly interface that allows interested parties to quickly and efficiently find relevant information without navigating through and downloading multiple documents.

Features

Boost Interest and Retention: Ensure that users remain engaged and interested with the product search by providing a seamless and hassle-free information retrieval experience.

Increase Efficiency: Reduce the time users spend searching for specific product details by enabling quick and accurate responses to their queries and also citing the references from the product documentation.

Knowledge Processing and Integration: A system to integrate and, process hundreds of product information sheets and manuals into an AI powered knowledge base.



03

Resources

AI Models and Tools

Llama 3 LLM Model

The Llama3 model is the core of the AI assistant, responsible for advanced natural language understanding (NLU) and generation (NLG). Llama3 generates accurate and contextually relevant responses based on Trapeze Group's extensive product information. Furthermore, it continuously learns from user interactions, improving its accuracy.

Hugging Face

The HuggingFace model is crucial for text embedding and semantic search within the project. It transforms product information and user queries into dense vector representations, enabling semantic understanding and efficient retrieval of relevant information.

Astra Vector DB

Astra DB serves as the highly scalable and resilient vector index storage and management solution. It stores embeddings generated by the HuggingFace model, providing a robust database for rapid information retrieval. It ensures high availability, handling high query loads reliably.

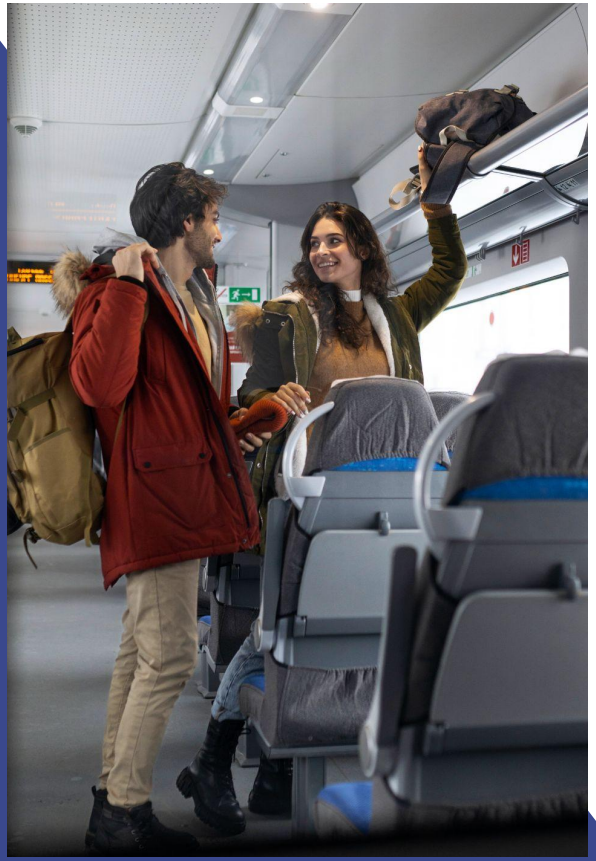
AI Models and Tools

Lang Chain

LangChain serves as the backbone for integrating and orchestrating various AI components within the project. It acts as a middleware that ensures seamless interoperability between the Llama3 model, Groq hardware, Hugging Face model, and Astra DB. LangChain manages the workflow of the AI assistant, enabling smooth data flow and efficient processing.

Groq LPU

Groq provides the necessary hardware acceleration specifically designed for AI and machine learning applications. By significantly boosting computational speed, Groq enables the AI assistant to process large volumes of data swiftly. Its scalability supports the increasing number of user queries and larger datasets without compromising performance.



04

Budget

Budget

The project uses completely Open source models and Free tier tools. There is no cost associated with the project if it used at its current capacity.

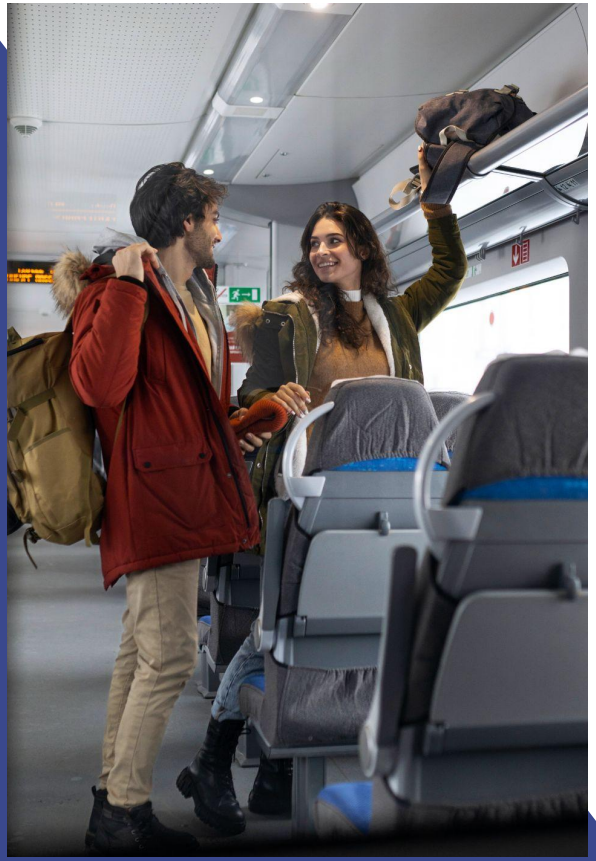
Use of Open Source Models and Tools

- **Cost-Effective**: Open source models and tools reduce development costs significantly as they are freely available and do not require expensive licensing fees.
- **Community Support**: A large community of developers contributes to and supports these tools, providing a wealth of knowledge, resources, and troubleshooting assistance.
- **Flexibility and Customization**: Open source models can be tailored to meet specific project requirements, allowing for greater flexibility and customization.
- **Continuous Improvement**: These tools are constantly updated and improved by the community, ensuring that the project benefits from the latest advancements and bug fixes.
- **Transparency**: Open source code provides full visibility into the workings of the models, allowing for better understanding, auditing, and optimization.

Operating Cost

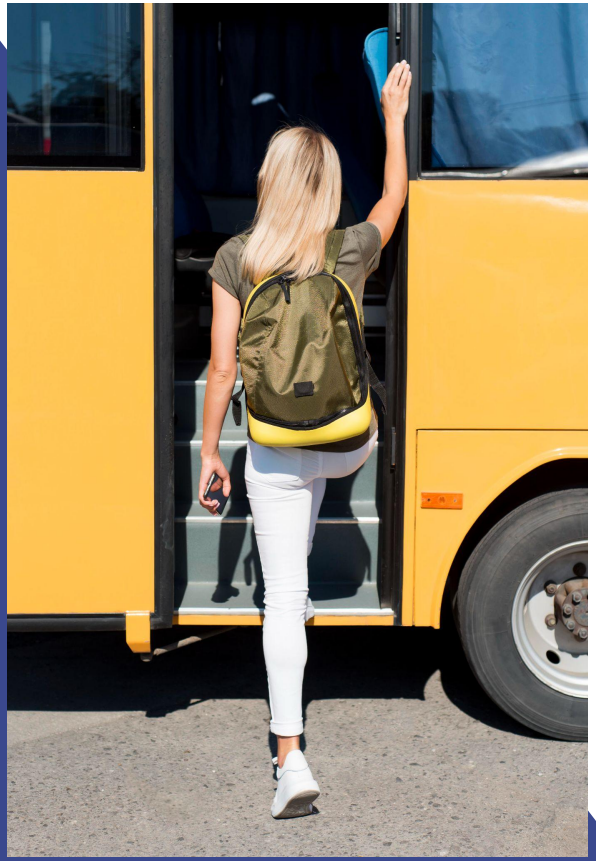
Factors affecting cost

- **Infrastructure:** Costs associated with cloud services or on-premises hardware to run the AI models and manage data storage.
- **Compute Resources:** Expenses related to hardware acceleration (e.g., Groq hardware) in paid tiers, which can enhance the performance and scalability of the AI assistant.
- **Data Storage:** Costs for storing embeddings data, particularly if going for higher tier using scalable and resilient solutions of Astra DB.
- **Development and Maintenance:** Budget for developers, data scientists, and ongoing maintenance to ensure the AI assistant continues to function optimally.
- **Licensing and Subscriptions:** While open-source tools are free, there may be associated costs for enterprise-level support, advanced features, or additional services.



05

Project



Project Details

Code: https://github.com/RamyaAnand27/PRODUCT_INFBOT

Technology Stack: Python, StreamLit
APIs used: Lang chain, Groq, Hugging Face, Astra DB, FAISS
Development Environment: VS Code, Miniconda3, Bash

Code / Resource	Description
/Data	Contains Product Information pfds downloaded from the Trapeze group public website https://www.trapezegroup.com/solutions
requirements.txt	Project dependencies and AI libraries
.env	Environment variables where app secrets are stored
ProductInfoAssistant_VectorDb.py	The specific index within Astra DB where data is stored and managed

Prerequisites / Configuration

- Create a .env (environment file) in the project folder
- Add below secret names to the file
- Create an account in GROQ and ASTRA DB website (links given under references)
- Create API KEY on the Groq account, copy it to the .env file
- Create an Index in the Astra Db, Get Application Token, endpoint information, add to .env

Secret name	Product	
GROQ_API_KEY	GROQ	The API key used to authenticate and access Groq's API services
ASTRA_DB_API_ENDPOINT	ASTRA	The endpoint URL for accessing Astra DB's API
ASTRA_DB_APPLICATION_TOKEN	ASTRA	The token used for authenticating API requests to Astra DB
ASTRA_DB_COLLECTION	ASTRA	The specific index within Astra DB where data is stored and managed

User Interface

Product Information Assistant

Enter your questions on the Trapeze Products and Solutions

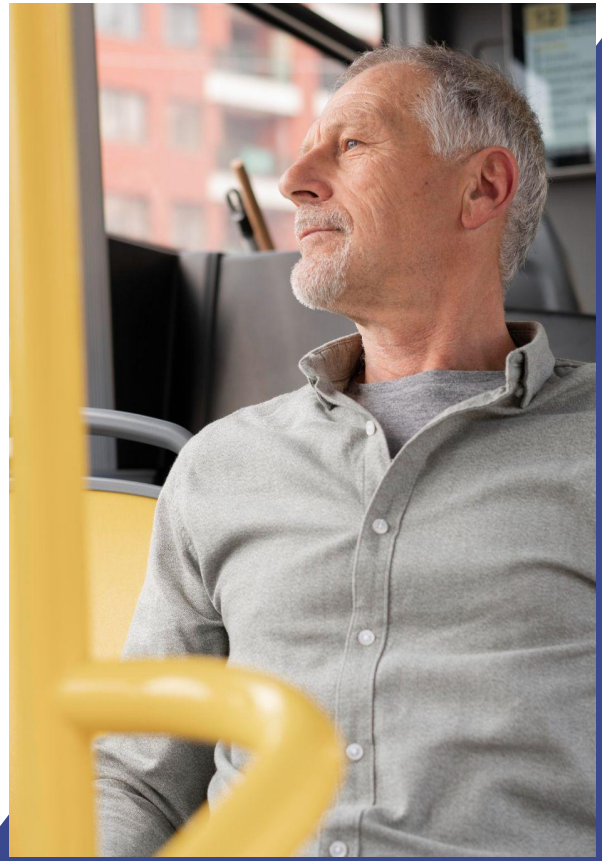
What are the features of One Mobility?

Based on the provided context, the features of OneMobility are:

1. Reduced No-Shows and Call Volume: allows passengers to make flexible trip arrangements, confirming, editing, or canceling bookings with ease.
2. Improved On-Time Performance: sends passengers native push notifications of their approaching ride, reducing vehicle wait times, and providing real-time information.
3. Increased Customer Engagement: provides a user-friendly service with a clean, modern look, allowing passengers to book and pay anytime, anywhere.
4. Greater Customer Convenience: enables passengers to find the information they're looking for, book and pay, and access recurring trips with one-click booking.
5. Enhanced Rider Experience: empowers riders to make their own decisions, providing feedback (e.g., rating drivers), and enabling agencies to assess performance and customer satisfaction.
6. Increased Safety and Peace of Mind: minimizes the need for physical interactions, provides real-time information, and informs passengers of vehicle and driver details.

OneMobility also includes Trapeze Workforce Management (OPS) for agencies to identify root causes, track key workforce metrics, and make predictions to minimize the impacts of workforce shortages and absences.

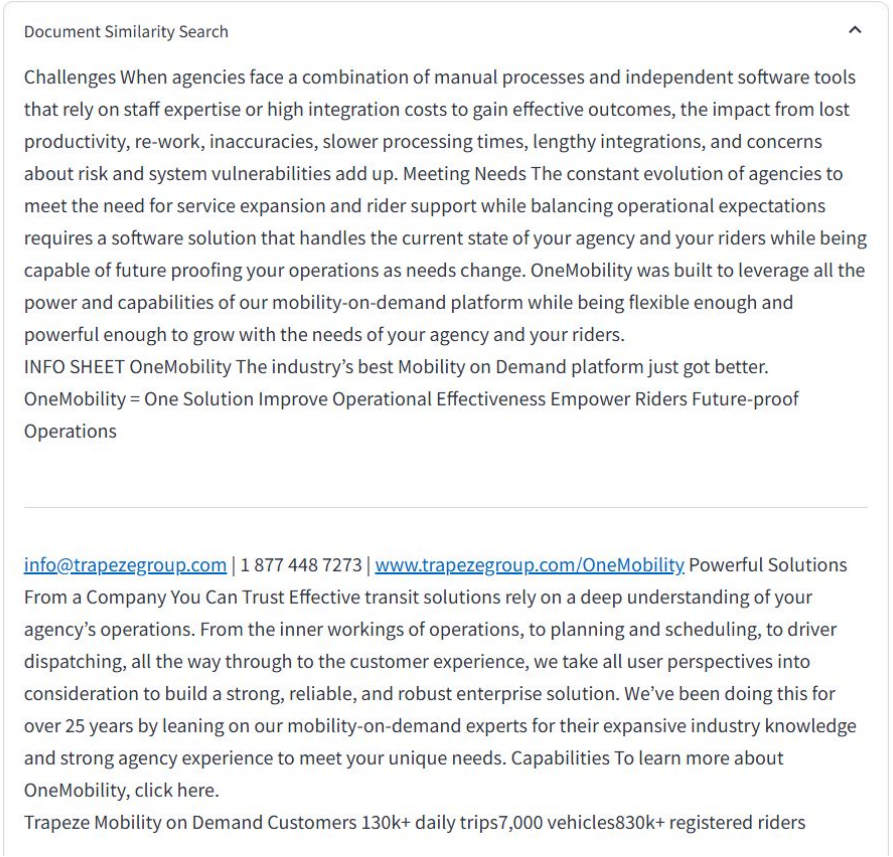
Document Similarity Search



Product data

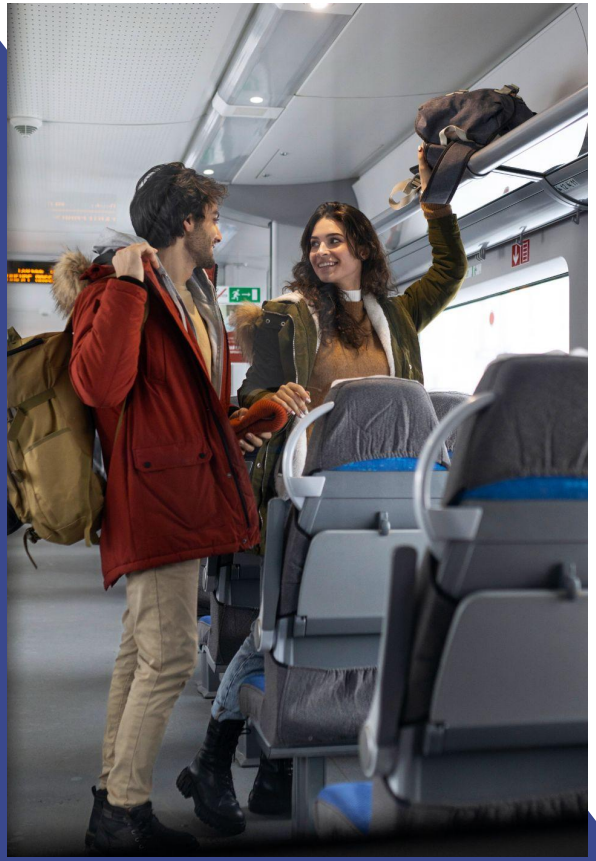
Benefits of RAG

The project offers visual organization, enabling clear presentation of information in a structured format. They enhance the data comparison against document and improve overall clarity



Sample prompts

1. What is PASS? What services does it offers?
2. How improved is PASS over the legacy application?
3. Which product should I go for, for scheduling trips?
4. Name some product that offers a cloud-based solution for paratransit?
5. Which product offers a cloud-based solution for paratransit services like booking and scheduling of trips?



06

Solutions

Solutions - other suggestions using AI

Virtual AI Assistant

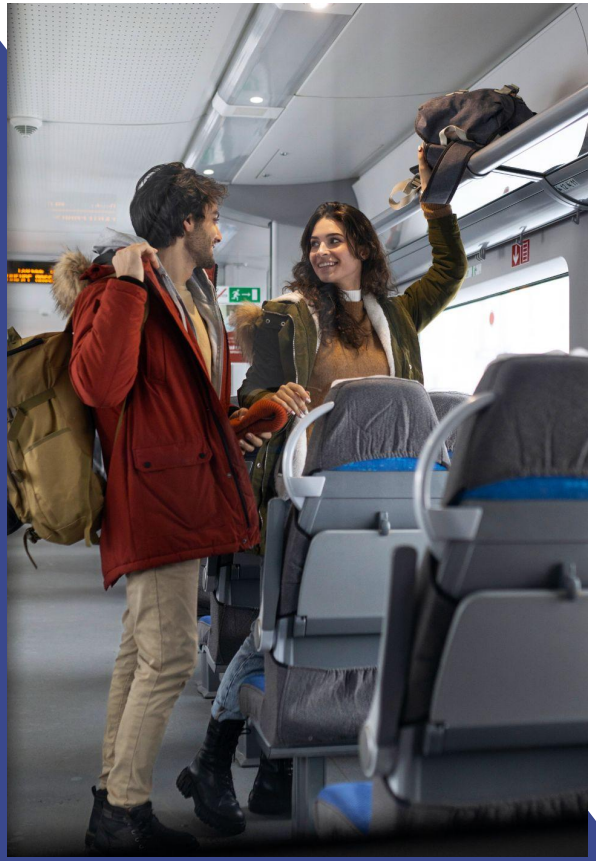
The app could allow users to request rides, provide feedback on their experience, and communicate any special needs or assistance requirements to operators, improving the overall quality of service. The app could recommend the best para-transit options for each user, considering factors such as wheelchair accessibility, assistance requirements, and preferred pick-up times.

Predictive Maintenance

The app could analyze vehicle data, such as engine performance, mileage, and maintenance history, to generate predictive maintenance schedules and alerts for operators, ensuring the reliability and safety of the para-transit fleet. The app could suggest alternative routes with optimal pick-up and drop-off points, minimizing travel time and maximizing the efficiency of the para-transit service.

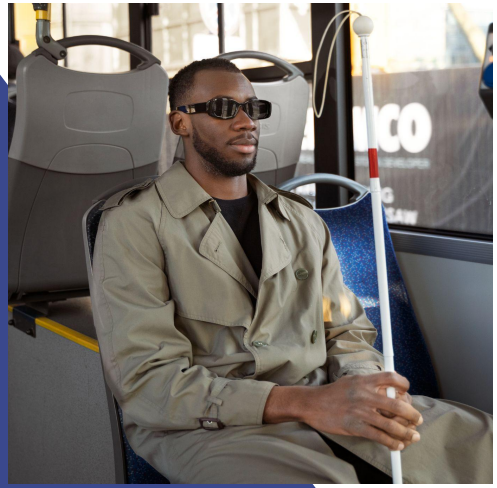
Security screening

By harnessing the capabilities of Generative AI tools, the app can analyze video feeds from surveillance cameras and integrate passenger data to identify potential security threats and verify passenger identities in real-time. Generative AI algorithms analyze the video streams to detect suspicious behaviors, unauthorized access, and potential security threats.



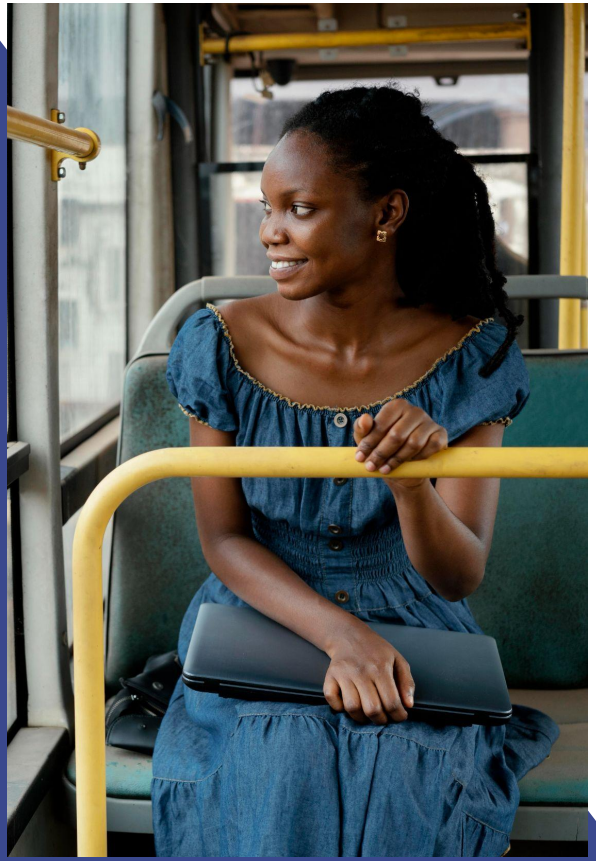
07

Conclusion



Dedication to Trapeze

This project was developed as a tribute to the Trapeze Group, reflecting my appreciation for the rewarding journey and enriching work experience I have had with the company.



Thanks!

Do you have any questions?

connect.ramyaanand@outlook.com
+1 905 514 9649
https://www.linkedin.com/in/link-ramya-anand/

References

https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/
https://www.langchain.com/
https://groq.com/
https://huggingface.co/
https://www.datastax.com/products/datastax-astra
https://streamlit.io/