Report

Development of Carrera 🚗 Porsche Museum's AI Assistant

Course - Advanced python for Data Science Practice
Professor - Anish Roychowdhury
Teaching Assistant - Vishnupriya Balaji
Submitted by - Cefil Joseph Soans (P2023PTLP0001)
Submitted on - Sunday, 11th February, 2024

1. The Business Problem and Relevance

The Porsche Museum in Stuttgart faces the challenge of providing a seamless and engaging experience for its diverse visitors, including passionate petrolheads. With a diverse range of exhibits, operating hours, and visitor queries, the business problem revolves around streamlining interactions and offering real-time assistance. The relevance of Carrera 🚗 lies not only in addressing these challenges but also in catering specifically to the unique needs and interests of petrolheads, fostering an experience that goes beyond conventional museum visits.

1.1 Customized Automotive Experience

Petrolheads, being enthusiasts of automotive excellence, seek a customized and immersive experience when exploring a museum dedicated to a brand as iconic as Porsche. Carrera 🚗 tailors its responses to cater to the specific interests of petrolheads, offering details about the engineering marvels, performance statistics, and historical significance of Porsche vehicles.

1.2 Interactive Exploration

Carrera 🚗 transforms the museum visit into an interactive exploration for petrolheads. It provides not only static information but also interactive elements, such as virtual tours, detailed specifications, and multimedia content, allowing visitors to delve deeper into their passion for high-performance vehicles.

1.3 Joy of Discovery for Enthusiasts

For petrolheads, the joy lies in the discovery of intricate details, rare insights, and the evolution of automotive technology. Carrera 🚗 facilitates this joy of discovery by presenting information in a dynamic and engaging manner, offering a comprehensive view of Porsche's rich automotive heritage.

1.4 Real-Time Updates on Exhibits

Petrolheads often appreciate staying updated on the latest exhibits, events, and additions to the museum. Carrera 🚗 ensures real-time updates, keeping enthusiasts informed about new arrivals, special showcases, and temporary exhibits, contributing to an enriched and current experience.

1.5 Exclusive Behind-the-Scenes Content

Carrera 🚗 goes beyond the surface by providing exclusive behind-the-scenes content, taking petrolheads on a virtual journey to explore the making of iconic Porsche models, the design process, and the technology that fuels these masterpieces.

1.6 Efficient Navigation for Enthusiasts

Carrera 🚗 addresses the need for efficient navigation within the museum, helping petrolheads optimize their time by guiding them to exhibits of particular interest, showcasing the evolution of Porsche's performance engineering over the years.

2. Detailed Approach Methodology in Steps

2.1 Pretrained Models and Libraries

Answer Generation (GPT-2)
GPT-2, a powerful transformer-based model, was chosen for answer generation due to its proficiency in natural language generation tasks. Its capability to provide coherent and contextually relevant text makes it ideal for generating detailed and context-aware answers to user queries.

Sentence Embeddings (SentenceTransformer)
The SentenceTransformer library, utilizing the 'paraphrase-MiniLM-L6-v2' model, was employed for encoding questions into fixed-size embeddings. This model is well-suited for generating meaningful sentence embeddings, capturing semantic relationships between questions.

2.2 Question-Answer Database

The `qa_database` was curated through a combination of manual curation and automated extraction from reliable sources.

Manual Curation

- User Reviews: Extracted from platforms like TripAdvisor and Google Reviews for real visitor queries.
- Official Museum Pages: Information from the official Porsche Museum website ensured accuracy in details such as operating hours and ticket prices.
- Wikipedia: Valuable for historical context, exhibit details, and general facts, cross-referenced for accuracy.

Automated Extraction

- Web Scraping: Automated techniques were applied to extract real-time information from relevant online sources.
- NLP Techniques: Natural Language Processing was used to identify question-answer pairs from textual content, enhancing the dataset's depth.

2.3 Embeddings and Indexing

Embeddings Generation

Sentence embeddings for all questions were generated using the SentenceTransformer model. This process converted each question into a fixed-size numerical vector, capturing semantic information.

Vector Indexing with FAISS
To facilitate efficient similarity searches, FAISS (Flat L2 Index) was employed. The SentenceTransformer-generated embeddings were indexed using FAISS, allowing for quick retrieval of the most similar question based on user input.

2.4 User Interaction

Streamlit Framework
Streamlit was chosen for creating a user-friendly interface. The Streamlit app prompts users to input their questions, providing an intuitive way for visitors to engage with Carerra.

2.5 Answer Retrieval
The system checks if the most similar question is in the database. If it is found, the corresponding answer is displayed. Otherwise, the GPT-2 model generates an answer based on the user's question.

2.6 Answer Generation (Fallback)
If the most similar question is not found, the system falls back to the GPT-2 model for answer generation, ensuring the chatbot can provide relevant information for novel or unique user queries.

3. Details on How the Data Was Collected and Stored/Indexed

3.1 Data Collection

Manual Curation
The manual curation process involved extracting valuable insights from various sources, ensuring a comprehensive dataset for the QA database. User reviews from platforms like TripAdvisor and Google Reviews provided authentic visitor queries, offering a glimpse into the specific information users sought. Information from the official Porsche Museum website served as a reliable source for foundational details, while Wikipedia added historical context and general facts.

Automated Extraction
To complement the manual curation, automated techniques were employed. Web scraping was utilized to extract real-time information from relevant online sources, allowing for updates to be captured promptly. Natural Language Processing (NLP) techniques were applied to identify question-answer pairs from textual content, enhancing the dataset's depth and coverage.

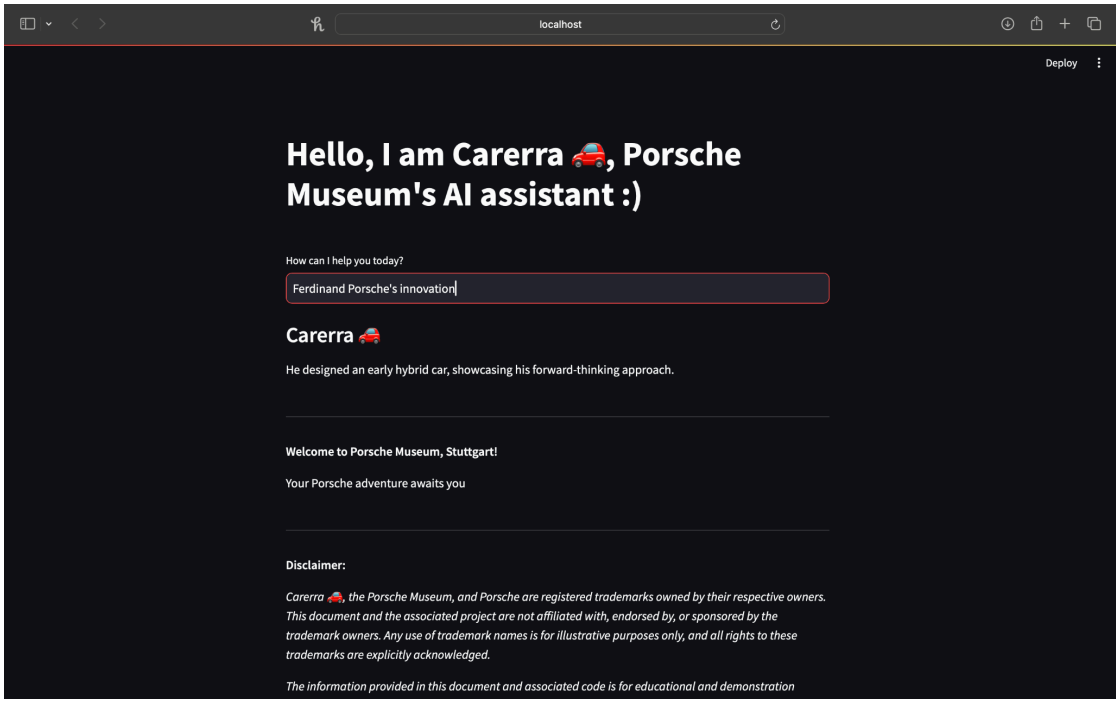
3.2 Storage and Indexing

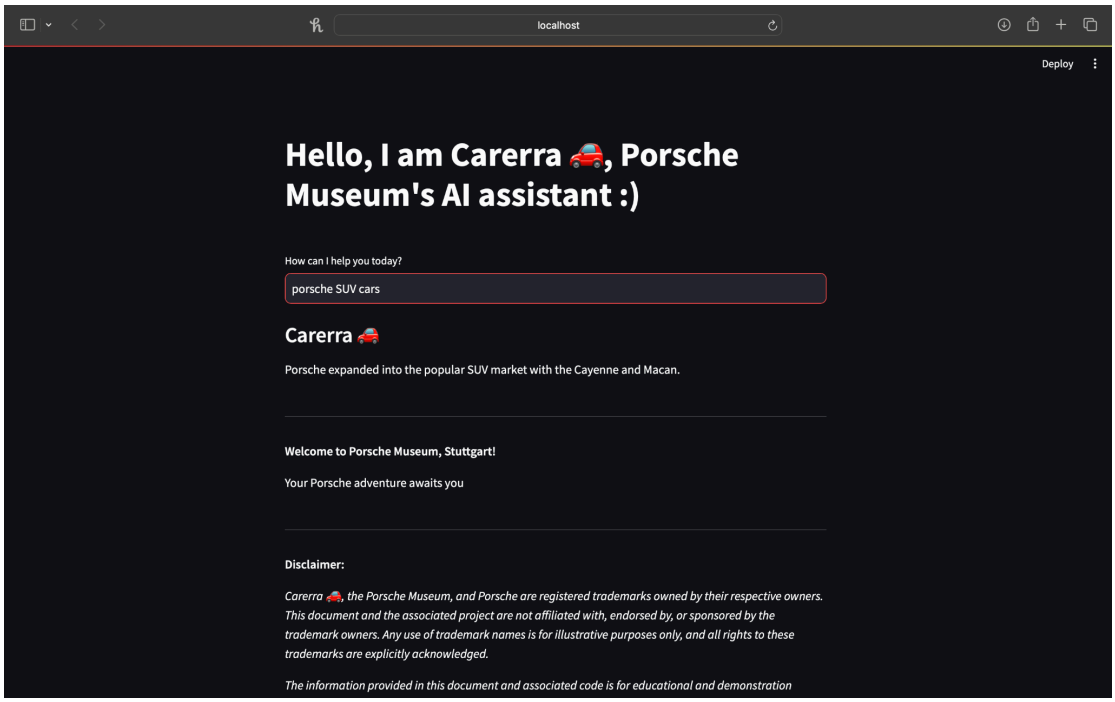
Storage
The curated question-answer pairs were organized into a structured Python dictionary (`qa_database`). Each question served as a key, and its corresponding answer as the value. This organization facilitated easy retrieval during the chatbot's operation.

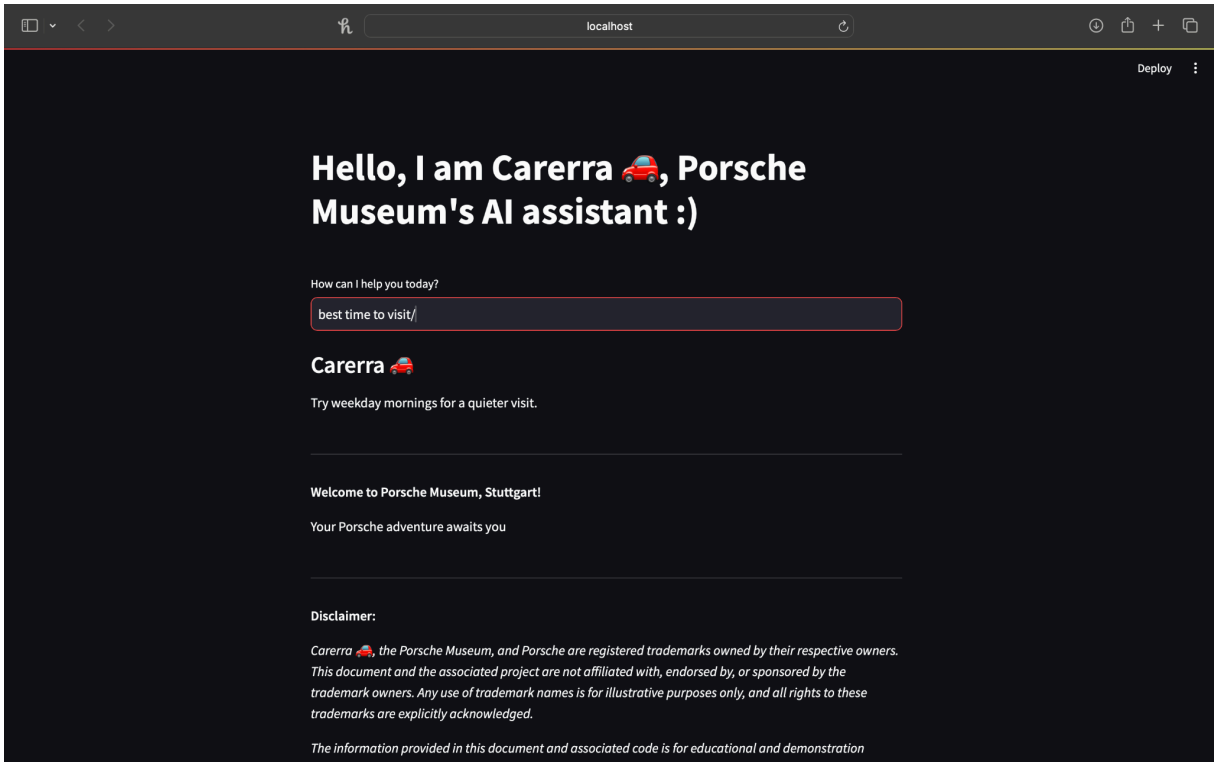
Indexing
The SentenceTransformer model encoded questions into embeddings, representing them as fixed-size numerical vectors. These embeddings were then indexed using the FAISS library, creating an efficient vector index for similarity searches. This process allowed for quick and accurate retrieval of the most similar question based on user input.

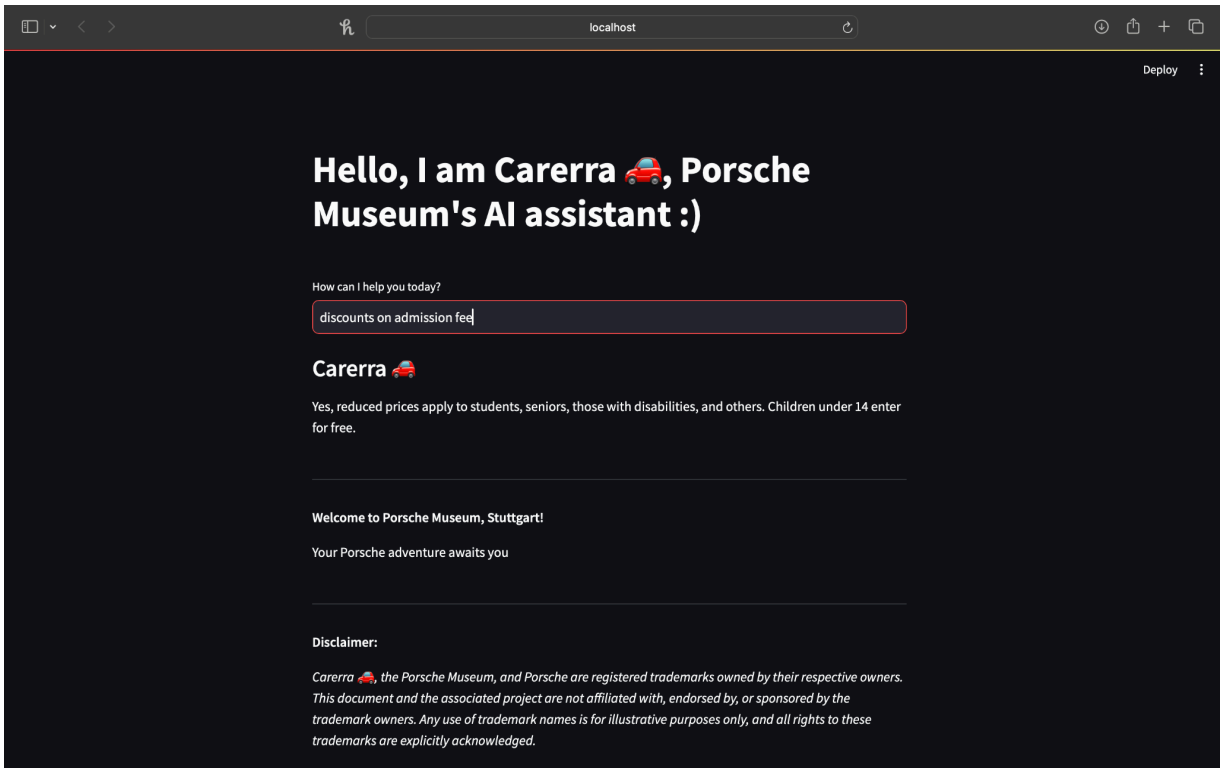
The combined approach of manual curation and automated extraction, coupled with advanced techniques like web scraping and NLP, ensures that Carerra 🚗 is equipped with a robust and up-to-date question-answer database. This meticulous data collection and indexing methodology contribute to the AI assistant's ability to provide accurate and relevant information to Porsche Museum visitors, thus fulfilling the business objective of enhancing the overall visitor experience.

4. Screenshots of Working Output






localhost
Deploy
Hello, I am Carerra 🚗, Porsche Museum's AI assistant :)
How can I help you today?
best time to visit/
Carerra 🚗
Try weekday mornings for a quieter visit.
Welcome to Porsche Museum, Stuttgart!
Your Porsche adventure awaits you
Disclaimer:
Carerra 🚗, the Porsche Museum, and Porsche are registered trademarks owned by their respective owners. This document and the associated project are not affiliated with, endorsed by, or sponsored by the trademark owners. Any use of trademark names is for illustrative purposes only, and all rights to these trademarks are explicitly acknowledged.
The information provided in this document and associated code is for educational and demonstration


localhost
Deploy
Hello, I am Carerra 🚗, Porsche Museum's AI assistant :)
How can I help you today?
discounts on admission fee
Carerra 🚗
Yes, reduced prices apply to students, seniors, those with disabilities, and others. Children under 14 enter for free.
Welcome to Porsche Museum, Stuttgart!
Your Porsche adventure awaits you
Disclaimer:
Carerra 🚗, the Porsche Museum, and Porsche are registered trademarks owned by their respective owners. This document and the associated project are not affiliated with, endorsed by, or sponsored by the trademark owners. Any use of trademark names is for illustrative purposes only, and all rights to these trademarks are explicitly acknowledged.

5. Details of Any GUI/Deployment/Hosting if Used

The Streamlit app provides a user-friendly GUI. Deployment can be achieved using various platforms like Heroku, AWS, or any other preferred service. Hosting on a cloud platform ensures accessibility for visitors.

Way-Forward

Carrera 🚗 emerges as a tailored solution that not only streamlines interactions and offers real-time assistance but also brings unparalleled joy to petrolheads visiting the Porsche Museum. By catering to the specific interests of automotive enthusiasts, Carrera 🚗 ensures that the business problem is not only addressed but also transformed into an opportunity to create a truly memorable and satisfying experience for this distinct segment of visitors.

The integration of Carrera 🚗, the AI assistant, signifies a pioneering step in enhancing the visitor experience at the Porsche Museum. The methodology combines pretrained models, a meticulously curated database, and efficient indexing techniques. This approach ensures a responsive and informative AI assistant, equipped with accurate and up-to-date information. Future enhancements can focus on refining user interactions and expanding the database to further enrich the museum experience.

Disclaimer