Tuwaiq Academy × Misk

Data Science & Machine Learning Bootcamp

Capstone Project Report

Sawwah: AI-Powered Tourism Assistant Using RAG and Speech Interaction

Prepared by:

Areen Alyahya

Mudhawi Alshiha

Shams Alarifi

Nouf Almutairi

Wajd Alrabiah

Supervised by:

Hany Elshafey– Instructor, AI Track

Cohort: Data Science & Machine Learning Bootcamp – Batch 2025

October 26, 2025

Abstract

Sawwah: AI-Powered Tourism Assistant Using RAG and Speech Interaction is an intelligent system designed to enhance access to Saudi Arabia's cultural and tourism information. The project integrates Retrieval-Augmented Generation (RAG), speech synthesis, and computer vision into a unified platform that allows users to interact naturally through text, voice, and visual engagement. Using curated English-language datasets about Saudi cities, heritage sites, and historical figures, Sawwah retrieves accurate, context-aware answers and responds in real time with human-like clarity. The web interface displays the retrieved responses and allows user interaction in real time.By aligning with Saudi Vision 2030's digital transformation goals, Sawwah demonstrates how artificial intelligence can promote tourism innovation, cultural preservation, and user-centric smart experiences across the Kingdom.

Keywords: Artificial Intelligence, Retrieval-Augmented Generation, Tourism, Computer Vision, Speech Interaction, Saudi Arabia.

Table of Contents

List of Tables

List of Figures

Acknowledgement

We extend our gratitude to Tuwaiq Academy and Misk Skills for providing an opportunity to apply and enhance our technical and analytical knowledge through this bootcamp. The experience has not only strengthened our practical understanding of artificial intelligence and data science but has also fostered collaboration, creativity, and problem-solving skills that will continue to guide us in our professional journeys.

This project represents more than just a technical achievement, it reflects the dedication, teamwork, and shared vision of our group. Throughout the capstone journey, each team member contributed their unique strengths, supporting one another through challenges and discoveries. Together, we learned the importance of communication, adaptability, and perseverance in bringing an idea from concept to reality.

We are sincerely thankful for the collective effort, support, and dedication that each team member contributed to the success of this project. The experience has been a true reflection of teamwork, filled with learning, perseverance, and creativity. Together, we have gained valuable technical and collaborative skills that strengthened our understanding of artificial intelligence and its real-world applications. This capstone journey not only marks the completion of a project but also the beginning of our continued growth as AI practitioners and innovators.

Team Contribution

The successful completion of the Sawwah Intelligent Tourism Assistant was made possible through the collaboration, commitment, and technical expertise of all team members. Each contributor played a vital role in the design, development, and integration of the system's core components.

The table below outlines the specific responsibilities and contributions of each team member.

Team Member(s)	Key Responsibilities and Contributions
Areen Alyahya	Led data collection, translation, cleaning, and preprocessing of English datasets. Conducted exploratory data analysis (EDA), implemented speech-to-text (STT) and text-to-speech (TTS) features, developed the web interface, integrated the model, and performed human evaluation of RAG responses. Participated in project demonstration and testing.
Shams Alarifi	Contributed to data collection and translation. Developed the RAG (Retrieval-Augmented Generation) pipeline, implemented retrieval workflows, and supported model evaluation. Assisted in web interface integration and human evaluation.
Nouf Almutairi	Focused on data preparation, computer vision module development, and RAG evaluation. Designed system requirements and UML diagrams, ensuring functional accuracy and consistency between modules.
Wajd Alrabiah	Worked on fine-tuning models, computer vision integration, dataset translation, and additional data preparation. Contributed to model evaluation, testing, and verification of human evaluation results.
Mudhawi Alshiha	Led system integration and computer vision synchronization with speech interaction. Contributed to system documentation, UML design, and report development. Participated in model evaluation and ensured consistency between modules during final testing.

Table 1: Team Members' Contributions and Responsibilities

Throughout development, the team worked collaboratively across all modules data processing, retrieval, speech, vision, and interface integration. Parallel testing and iterative refinement ensured that each subsystem achieved optimal accuracy and performance. This coordinated approach strengthened the reliability, scalability, and overall success of Sawwah as an intelligent tourism assistant.

1.0 Executive Summary

Sawwah: AI-Powered Tourism Assistant Using RAG and Speech Interaction is an advanced capstone project developed under the Tuwaiq Academy × Misk Data Science and AI Bootcamp. The project aims to create an intelligent tourism assistant capable of understanding user queries, retrieving accurate information, and responding naturally through both text and speech.

By integrating Retrieval-Augmented Generation (RAG), speech synthesis, computer vision, and a user-friendly web interface, Sawwah enhances access to tourism and cultural knowledge across Saudi Arabia.

The system consists of three fully integrated components:

- A RAG-based retrieval engine that connects user questions to relevant information from curated English tourism datasets.
- A speech interaction module that delivers real-time voice responses for a natural and engaging user experience.
- A computer vision (CV) module that detects when a person is present in front of the camera and automatically greets them with a personalized message such as "Welcome to Sawwah, how can I assist you today?"

The web interface allows users to enter text queries, receive spoken and written answers, and experience intelligent interaction through visual and voice-based engagement.

Together, these components create a seamless and human-like AI-driven tourism experience that combines intelligent retrieval, speech synthesis, and computer vision. All data used for training and retrieval were collected, cleaned, and prepared by the team to ensure originality, consistency, and reliability.

2.0 Introduction

Saudi Arabia's Vision 2030 places strong emphasis on digital transformation across all sectors, aiming to establish the Kingdom as a global leader in technology, culture, and innovation. One of the key pillars of this vision is the development of the tourism industry, where modern technologies are expected to enhance visitor experiences and promote cultural awareness. With the rise of digital platforms, a vast amount of tourism-related information has become available online; however, this information is often fragmented, inconsistent, or difficult to access, especially for non-Arabic speakers or first-time visitors.

Tourists frequently encounter challenges in obtaining accurate, localized, and up-to-date information about destinations, events, and heritage sites. They may rely on scattered websites or social media pages that lack organization and reliability. This gap highlights the need for an intelligent, centralized, and interactive solution that can understand natural language questions and provide relevant, trustworthy answers quickly and intuitively.

Sawwah was created to address this need by leveraging the power of artificial intelligence. It serves as an AI-driven tourism assistant designed to simplify the way users explore Saudi Arabia's attractions and cultural landmarks. The system combines multiple technologies including Retrieval-Augmented Generation (RAG) for intelligent information retrieval, speech interaction for natural communication, and computer vision for real-time user engagement to deliver a seamless and personalized experience.

Through this integration, Sawwah enables users to interact with technology as if they were speaking to a knowledgeable local guide. The assistant retrieves relevant information, answers questions with precision, and even greets users through voice when it detects their presence via camera. By blending human-like interaction with reliable data, Sawwah not only enhances tourism accessibility but also aligns with the broader national goals of innovation, digital empowerment, and global hospitality leadership envisioned in Vision 2030.

3.0 Project Objectives and Significance

The main objective of this project is to design and implement an AI-powered intelligent system that improves the accessibility and quality of tourism information across Saudi Arabia. The system is developed to function as a personalized digital tourism assistant, capable of understanding natural language queries, retrieving accurate and contextually relevant information, and communicating results through both text and speech in a user-friendly manner.

By leveraging advanced technologies such as Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), speech synthesis, and computer vision, Sawwah integrates multiple domains of artificial intelligence into one cohesive, interactive platform. This multidimensional approach allows the assistant to provide

responses that are not only informational but also human-like and engaging, enhancing user experience and interaction quality.

The project aligns with Saudi Arabia's Vision 2030 initiatives, which prioritize digital transformation and smart service development across key sectors, including tourism. By bridging the gap between travelers and localized information, Sawwah supports the Kingdom's ambition to showcase its cultural richness and heritage through innovative technology. The system demonstrates how AI can be effectively utilized to promote tourism, preserve cultural identity, and foster a sustainable, data-driven tourism ecosystem.

Beyond its practical functionality, the project holds academic and social significance. It serves as a tangible example of how AI technologies can be ethically and effectively integrated into real-world applications that serve communities and improve public engagement. The development process also reflects the team's collaboration, technical proficiency, and commitment to using data science and machine learning to address real challenges in Saudi Arabia's growing tourism sector.

4.0 System Requirements Specification

This section defines the functional and non-functional requirements for the Sawwah system. The requirements were identified based on the project objectives, user expectations, and technical capabilities of the implemented modules. Together, they

outline the operational behavior, performance standards, and design constraints necessary for the successful deployment and scalability of the system.

4.1 Functional Requirements

The functional requirements specify the key features and operations that the system must perform. They are categorized based on the roles of the Admin, User, and the System itself.

4.1.1 Admin Requirements

The administrator is responsible for managing and maintaining the system's data and performance. These functions ensure that the information stored within Sawwah remains accurate, up to date, and reliable for user interactions.

- The system shall allow the admin to upload, update, or remove records from the English tourism datasets, including information related to cities, heritage sites, projects, and cultural figures.
- The system shall provide a web interface that enables the admin to monitor system activity, user queries, and response statistics in real time.
- The system shall allow the admin to trigger retraining or embedding updates when new data is added, ensuring retrieval accuracy and freshness of responses.
- The system shall provide error logs and diagnostic feedback for maintenance and performance evaluation.

4.1.2 User Requirements

The user represents the primary stakeholder such as a tourist, student, or any individual seeking information about Saudi Arabia's attractions, history, and cultural heritage. The system must offer a simple and engaging interface to enhance accessibility and user satisfaction.

- The system shall allow users to ask tourism-related questions in natural language through text input.
- The system shall retrieve and generate accurate, contextually relevant responses using the RAG module.

- The system shall deliver spoken responses and greetings using the integrated speech interaction module to simulate a human-like experience.
- The system shall display summarized results and responses through the web interface, providing clear and interactive access to retrieved information.
- The system shall maintain a user-friendly and multilingual interface, allowing accessibility for both Arabic and English users (future expansion).

4.1.3 System Requirements

The system functions as the core operational layer that integrates retrieval, speech, and interface components. These requirements ensure that *Sawwah* maintains consistency, responsiveness, and interoperability among its modules.

- The system shall retrieve relevant documents and generate responses through the RAG engine, utilizing preprocessed embeddings from curated datasets.
- The system shall convert text-based responses into spoken audio output using a text-to-speech engine such as pyttsx3 or ElevenLabs API.
- The system shall detect user presence via the computer vision module and trigger automatic voice greetings when a person is identified in front of the camera.
- The system shall display results, greetings, and query insights through the web interface, ensuring real-time synchronization and smooth interaction between all modules.
- The system shall log system activities for monitoring and improvement.

4.2 Non-Functional Requirements

Non-functional requirements describe the quality attributes, performance targets, and operational standards of *Sawwah*. They ensure that the system is not only functional but also efficient, secure, and maintainable in real-world scenarios.

- Performance:

  The system should respond to standard user queries within three seconds on average,

ensuring smooth and efficient interaction. Latency in speech synthesis or retrieval should be minimized through optimized embeddings and caching mechanisms.

- Security:
  Sensitive data such as API keys, embedding files, and system credentials must be encrypted and securely stored. Access to administrative features should be authenticated through role-based permissions.

- Usability:
  The user interface should be intuitive, accessible, and visually appealing, allowing users of varying technical backgrounds to interact comfortably. Voice interactions should maintain clarity and natural tone.

- Scalability:
  The system should support integration with new datasets or AI models without major architectural changes. The modular design allows the easy addition of multilingual capabilities and future upgrades such as recommendation or personalization features.

- Maintainability:
  The system codebase should be structured, well-documented, and modular, allowing developers to update, debug, and expand features efficiently. All dependencies should be version-controlled and clearly listed for reproducibility.

5.0 Domain Overview: Tourism and Digital Innovation

Saudi Arabia's tourism sector has emerged as one of the fastest-growing industries under Vision 2030, aimed at diversifying the national economy and showcasing the Kingdom's cultural and historical richness. The increasing number of international and domestic visitors has created a strong demand for digital solutions that can enhance accessibility, efficiency, and personalization in the tourism experience.

Despite this rapid growth, many digital platforms still face challenges such as fragmented information, limited multilingual support, and low levels of interactivity. Tourists often depend on separate websites, guidebooks, or social media content, which are not always reliable or contextually accurate.

The Sawwah project operates within this domain, addressing these challenges by introducing an AI-powered tourism assistant that combines retrieval-augmented generation (RAG), speech interaction, and computer vision. This system not only supports the Kingdom's vision for smart tourism and cultural digitization, but also provides a scalable model for how artificial intelligence can personalize and humanize digital experiences in tourism and heritage industries.

6.0 SWOT Analysis

The Sawwah project operates at the intersection of artificial intelligence and tourism innovation, reflecting a strategic blend of technological capability and national vision. Like any complex AI system, its development involves a combination of strengths, weaknesses, opportunities, and threats that define its growth potential and sustainability.

Understanding these dynamics is essential to evaluating the project's strategic position within Saudi Arabia's tourism and digital transformation ecosystem. The SWOT analysis below highlights the internal and external factors that influence Sawwah's success, scalability, and long-term relevance in an evolving technological landscape.

Strengths	Weaknesses
• Integrates advanced AI technologies (RAG, speech, CV) into a unified, modular system. • Uses curated, tourism-specific datasets for accurate and contextual responses. • Enhances user experience through voice and visual interaction. • Designed with scalability and maintainability for future expansion.	• Limited dataset coverage in early stages may restrict content diversity. • Dependence on device hardware and network quality for speech and CV performance. • Currently supports English only; Arabic version under development.
Opportunities	Threats
• Expansion into multilingual and cross-platform applications. • Collaboration with tourism authorities and cultural organizations for real deployment.	• Rapid evolution of AI technologies may require frequent model updates. • Possible data access or licensing limitations for new sources. • Privacy and ethical concerns related to voice and facial detection features.

• Potential integration with AR/VR and IoT-based smart tourism systems. • Strong alignment with Vision 2030's goals for innovation and digital experience.	

Table 2: SWOT Analysis

6.1 Strengths

Sawwah demonstrates strong technical and strategic strengths that position it as an innovative solution within the digital tourism landscape. The system integrates multiple advanced technologies including Retrieval-Augmented Generation (RAG), speech interaction, and computer vision to create a unified, human-like user experience. This combination allows the assistant to understand natural language, deliver accurate information, and engage interactively with users through voice and visual recognition.

Furthermore, the project's custom-built datasets tailored to Saudi tourism represent a unique strength, ensuring relevance and contextual accuracy. The team's ability to leverage modern tools such as LangChain, OpenCV, and Streamlit also highlights strong technical proficiency and a clear understanding of applied AI development. The modular architecture enhances scalability and maintainability, allowing Sawwah to adapt and evolve as the tourism sector grows.

6.2 Weakness

Despite its innovative features, Sawwah faces several challenges that could impact its broader adoption and long-term performance. The system currently relies on a limited dataset scope, which may restrict the depth and diversity of its responses across all tourism categories. In addition, the assistant's current implementation supports English-language interaction only, limiting accessibility for Arabic-speaking users a critical audience in Saudi Arabia's tourism sector.

Another area of weakness lies in the dependency on hardware performance and network quality for real-time functionalities such as text-to-speech and computer vision. These dependencies may affect the system's reliability in low-resource environments or on devices with limited processing power. Continuous optimization and localization efforts will be required to address these limitations as the project scales.

6.3 Opportunities

The Sawwah project is uniquely positioned to capitalize on several growth opportunities aligned with Saudi Vision 2030 and global trends in AI-driven tourism. Expanding multilingual capabilities particularly the inclusion of Arabic language support would significantly broaden the system's reach and inclusivity. Additionally, integration with smart city platforms, IoT systems, or augmented reality (AR) applications could transform Sawwah into a core component of next-generation tourism experiences.

Collaborations with government entities such as the Saudi Tourism Authority or local cultural organizations could further validate and expand the project's dataset, ensuring up-to-date and authentic content. The increasing global interest in sustainable and digital travel experiences also presents an opportunity for Sawwah to position itself as a pioneering model for smart, culturally aware, and sustainable tourism innovation.

6.4 Threats

Like all AI-based systems, Sawwah faces several external threats that may influence its long-term sustainability. Rapid advancements in artificial intelligence could lead to technological obsolescence if continuous updates and retraining are not maintained. In addition, data privacy and ethical considerations particularly those related to speech and facial detection require strict compliance and transparent policies to maintain user trust.

The project may also encounter competition from commercial AI assistants or government-led tourism applications that adopt similar functionalities at scale. Furthermore, data availability constraints, copyright restrictions, or changes in open-source licensing could affect access to essential training data and APIs. Addressing these potential threats will require proactive innovation, adaptive governance, and close monitoring of emerging AI standards.

7. 0 Stakeholder Analysis

Stakeholder	Role	Needs
Admin	System overseer responsible for managing datasets, monitoring system performance, and ensuring operational stability.	Manage and update datasets, monitor activity through the web interface, and maintain system accuracy.
User	Tourist or visitor interacting with the AI assistant to obtain tourism information and cultural insights.	Ask natural-language questions, receive accurate and spoken responses, and view summarized information through the web interface.
System (AI Model)	Automated engine that performs retrieval, generation, and interaction tasks.	Access curated datasets and resources to generate contextually accurate and relevant responses.

Table 3: Stakeholder Analysis

8.0 Software Components

Sawwah follows a modular software architecture in which each component operates independently yet interacts seamlessly with the others to deliver the system's complete functionality. This design promotes scalability, maintainability, and efficient parallel development among team members. The primary components of the system are described below.

- RAG Engine

This is the core intelligence of the system, responsible for information retrieval and answer generation. It uses Retrieval-Augmented Generation (RAG) techniques to extract relevant data from the embedded knowledge base created from the English tourism datasets. The engine retrieves the most contextually relevant content and synthesizes it into coherent, human-like responses.

- Speech Engine

The speech module converts the generated text responses into natural-sounding voice output, creating an interactive and engaging experience for users. Implemented using text-to-speech technologies such as pyttsx3 and the ElevenLabs API, this component ensures that Sawwah can communicate with users in real time, simulating a human guide's tone and rhythm.

- Web Interface Module

The web interface serves as the interactive layer between the user and the system. It allows users to enter questions, view generated responses, and listen to spoken outputs in real time. Built using Flask, HTML/CSS, and JavaScript, the interface provides a clean, user-friendly environment that reflects the project's lavender theme and visual identity.

The web interface connects directly to the backend modules, invoking the RAG model for retrieval, the speech synthesis module for audio generation, and the computer vision module for greeting detection. It ensures real-time response display and smooth interaction, contributing to Sawwah's overall usability and accessibility.

- Computer Vision (CV) Module

The Computer Vision (CV) module enhances Sawwah's human-like interaction by detecting user presence through a live camera feed. When a person is identified, the system automatically initiates a personalized greeting such as "Welcome to Sawwah, how can I assist you today?"

This component integrates OpenCV and MediaPipe technologies to enable real-time face detection and ensure accurate recognition. By synchronizing with the speech engine, the CV module delivers a seamless and natural user experience, reinforcing Sawwah's goal of intelligent, interactive engagement.

Together, these modules form an integrated AI system capable of combining vision, speech, and knowledge retrieval to provide intelligent, natural, and context-aware interactions.

9.0 System Architecture

The architecture of Sawwah follows a modular and layered design, ensuring that each component performs a distinct function while maintaining seamless integration within the overall system. This structure allows for scalability, easier maintenance, and flexibility to introduce new features such as computer vision enhancements and multilingual capabilities in future iterations.

The system operates through a clear and logical flow. The user initiates interaction via the web interface, either by typing a question or by being detected visually through the computer vision module. Once the query is received, the RAG engine processes it by retrieving relevant information from the embedded knowledge base, which was built using cleaned and structured English datasets. The retrieved content is then synthesized into a coherent, contextually appropriate response.

The generated response is passed to the speech engine, which converts the textual answer into natural-sounding voice output using text-to-speech technologies such as gTTS, pyttsx3, or ElevenLabs API. Simultaneously, the response is displayed in real time through the web interface, allowing users to view and listen to the generated answers interactively.

This layered communication between modules enables Sawwah to deliver an end-to-end experience that combines intelligence, interactivity, and accessibility. The architecture ensures that each layer data, retrieval, speech, and visualization can be independently optimized or replaced without affecting the rest of the system, reinforcing the project's modular and sustainable design philosophy.

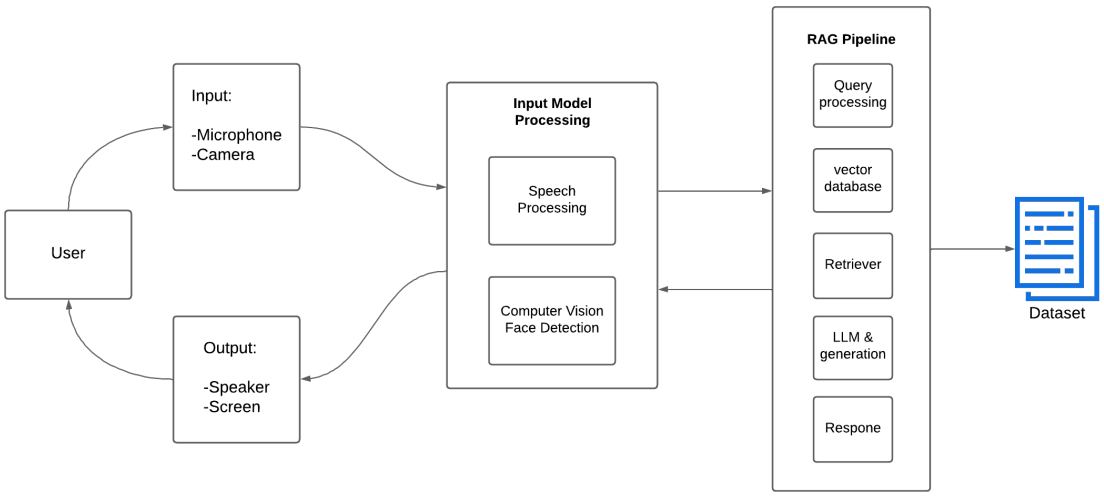

Figure 1: Sawwah System Architecture

10.0 Use Case Diagram and Descriptions

The use case diagram illustrates the primary interactions between the actors and the *Sawwah* system. It serves as a high-level representation of the system's functionality, capturing how users and administrators communicate with the modules to accomplish specific goals.

There are three main actors involved in the *Sawwah* environment:

- User: A tourist, visitor, or general individual who interacts with *Sawwah* to obtain information about Saudi attractions, heritage sites, or cultural figures.
- Admin: The administrator responsible for maintaining datasets, monitoring system performance, and ensuring that content remains accurate and up to date.
- System: The automated modules within Sawwah (RAG Engine, Speech Engine, Web Interface, and Computer Vision) that handle data processing, retrieval, and interaction delivery.

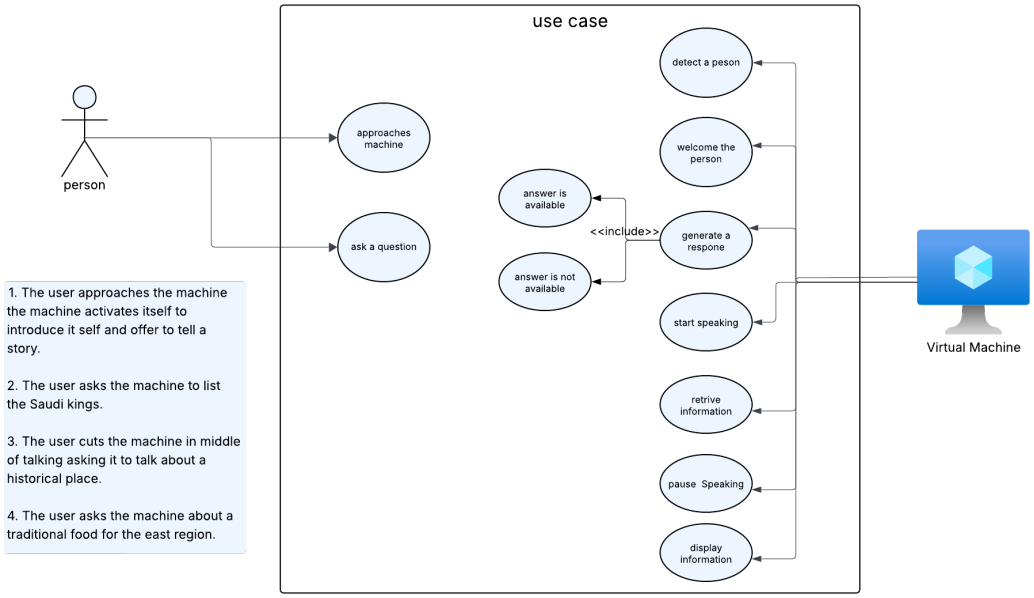

Figure 2Sawwah System Architecture and Interaction Flow

This figure illustrates the interaction sequence between the user and Sawwah's intelligent modules. The process begins when the computer vision module detects a person and activates the system. Through the web interface, the user can ask questions, which are processed by the RAG engine to retrieve relevant information. The speech module converts responses into spoken output, creating a natural and interactive user experience.

Use Case	Actor	Description
Ask Question	User	The user enters a natural language question related to tourism or cultural information in Saudi Arabia through the system interface.
Retrieve Answer	System	The RAG engine retrieves relevant data from the embedded knowledge base and generates a contextual, human-like response.
Voice Greeting	System	The computer vision module detects the presence of a person and triggers an automatic voice greeting, such as "Welcome to Sawwah, how can I assist you today?"
Speech Response	System	The speech engine converts the generated text into natural-sounding speech, delivering the answer to the user.
System Monitoring	Admin	The administrator accesses the web interface to monitor user queries, review system activity, and ensure consistent performance.

Dataset Management	Admin	The administrator updates, adds, or removes entries from the English datasets to maintain data quality and relevance.

Table 4: Use Case Descriptions

11.0 Methodology

The methodology adopted for the Sawwah project encompasses several stages, including data collection, data cleaning, knowledge base construction, and system integration. Each stage was carefully designed to ensure that the system delivers accurate, efficient, and meaningful responses to user queries.

11.1 Data Collection and Preparation

The first stage involved the collection of datasets relevant to Saudi Arabia's tourism sector. The team gathered data from verified sources, official tourism websites, and heritage documentation platforms. Only English-language datasets were used to maintain uniformity and ensure compatibility with the retrieval models.

All datasets were manually reviewed, cleaned, and organized by the project team. The cleaning process included removing duplicates, correcting inconsistencies, and filling missing fields where necessary. Once cleaned, the data was standardized and merged into a unified structure, forming the foundation of Sawwah's knowledge base.

The English datasets used in the project are summarized below.

Dataset Name	Description	Purpose
Cities	Detailed city descriptions, locations, and tourism attributes.	Primary source for city-based retrieval.
Projects	National tourism and heritage projects.	Expands factual coverage and diversity.
Characters	Profiles of cultural and historical figures.	Adds cultural/heritage context to answers.

Table 5: English Dataset Overview

11.2 Knowledge Base and Model Integration

After data preparation, the cleaned datasets were converted into embedded document representations using language model embeddings. These embeddings allowed the RAG engine to identify and retrieve the most relevant pieces of information based on semantic similarity rather than simple keyword matching.

The Retrieval-Augmented Generation (RAG) architecture was selected because it effectively bridges structured information retrieval with generative text synthesis. This ensures that Sawwah's responses are accurate, contextual, and naturally worded. The RAG system was implemented using Python, Pandas, and LangChain libraries, which facilitated efficient data processing, chunking, and retrieval evaluation.

The system was then integrated with the speech module to transform generated text responses into audio output using pyttsx3 and ElevenLabs APIs. Finally, the components were connected through a Flask-based web interface, which allows users to interact with the system in real time by submitting queries, receiving both textual and spoken responses, and triggering automatic greetings through the computer vision module.

Through these integrated steps, the methodology ensured that Sawwah not only functions as a technical prototype but also demonstrates real-world applicability in Saudi Arabia's digital tourism ecosystem.

11.3 Summary of Development Workflow

The overall development workflow followed a sequential and iterative process. It began with data gathering and preprocessing, ensuring the quality and reliability of tourism information. This was followed by feature embedding and model construction to establish the retrieval engine. The speech and web interface components were then developed and linked to the retrieval model, creating an interactive end-to-end system.

Each module was individually tested and refined before being integrated into the final system architecture. This structured methodology promoted collaboration, code reusability, and smooth troubleshooting throughout the development cycle. The result is a functional, modular, and scalable system that reflects the effective application of artificial intelligence in tourism information services.

11.4 Additional Supporting Data

To strengthen the project's scope and validate the concept of Sawwah, additional supporting datasets were reviewed and incorporated during the exploratory phase. These datasets included publicly available tourism statistics, regional development reports, and cultural event data, which provided context on the growth and diversification of Saudi Arabia's tourism sector. By integrating this supplementary information, the team was able

to design a more representative and context-aware knowledge base. This inclusion not only enhanced the model's understanding of real-world tourism trends but also reinforced the importance and timeliness of developing intelligent digital solutions that align with the Kingdom's Vision 2030 objectives for a thriving tourism economy.

11.4.1 External Tourism Datasets and Exploratory Analysis

To strengthen the contextual understanding of tourism behavior and expenditure patterns across Saudi Arabia, the team also reviewed and analyzed supporting datasets from the Ministry of Tourism's Open Data Portal (open.data.gov.sa). These datasets provided quantitative insights into national and regional tourism activity from 2015 to 2024.

The external datasets included:

1. Tourism Expenditure Dataset (Inbound and Domestic):

Focused on Riyadh region tourism spending, covering detailed yearly data (2015–2024).
It includes comparative sections for domestic and outbound spending patterns, revealing trends in local vs. international tourism flows.

2. Events and Activities Dataset (2021–2024):

Contains records of tourism and cultural events across 13 administrative regions of the Kingdom, categorized into three event types per year (2021–2024). This dataset highlights the growing diversity and distribution of tourism events aligned with Vision 2030 initiatives.

3. Aggregated Tourism Spending Statistics:

Summarizes total tourism expenditure for domestic travelers across several years. Monthly spending patterns were analyzed and extrapolated to represent yearly consumption trends from 2021 to 2024 for both inbound and outbound tourism.

All sources were retrieved from verified Ministry of Tourism repositories to ensure reliability and official validation.

These external datasets significantly enhanced the project's analytical foundation, providing evidence-based context that complements Sawwah's AI-driven retrieval model.

11.5 Project Management and Timeline

The Sawwah Intelligent Tourism Assistant project was completed within a focused two-week sprint. Despite the short duration, the team adopted a structured, milestone-based approach to ensure timely delivery of each module. Tasks were divided logically, and progress was tracked daily to maintain balance between development, testing, and documentation.

Phase	Duration	Main Tasks and Deliverables
Week 1 – Development Phase	Days 1–7	• Data cleaning and preprocessing of English datasets. • Implementation of the RAG pipeline for semantic retrieval. • Integration of speech and computer vision modules. • Initial web interface setup and design implementation.
Week 2 – Testing and Finalization Phase	Days 8–14	• Evaluation of prompt accuracy and response quality. • Optimization of retrieval and speech performance. • Integration testing of web interface and CV greeting. • Documentation, report writing, and final presentation preparation.

Table 6: Project Management and Time

Through effective task division and continuous coordination, the team ensured that all modules retrieval, speech, vision, and the web interface were developed, tested, and refined within the project timeframe.

12.0 Implementation and Tools

The implementation of Sawwah was carried out in several structured stages, focusing on the collection, preparation, and integration of data, followed by the

development of intelligent modules and a user-friendly interface. Each stage of implementation utilized appropriate tools and technologies to ensure efficiency, scalability, and performance.

12.1 Development Environment

Development was conducted using Python 3.x across multiple environments, including Jupyter Notebook, Google Colab, and local IDEs such as Visual Studio.

These platforms facilitated modular experimentation, flexible debugging, and streamlined workflow management across various development stages including data preprocessing, Retrieval-Augmented Generation (RAG), speech interaction, and computer vision greeting modules.

The combination of cloud-based and local environments enabled faster iteration, smoother model testing, and easier integration between independently developed components.

12.2 Technology Stack

The project integrated a diverse set of open-source tools and frameworks across data science, artificial intelligence, and visualization domains:

- Data Handling: Pandas, NumPy
- AI & Machine Learning: LangChain, scikit-learn
- Speech Interaction: gTTS, ElevenLabs (cloud-based neural voices), AssemblyAI
- Computer Vision: OpenCV, MediaPipe
- Visualization & Interface: Streamlit, Dash, Matplotlib, Plotly

This multi-layered technology stack enabled the development of an intelligent, interactive system capable of real-time retrieval, analysis, and communication.

Additionally, AssemblyAI was briefly tested as an alternative speech API during early experimentation before finalizing ElevenLabs for its superior voice clarity and stability.

12.3 Data Pipeline

The data pipeline begins with loading and normalizing cleaned English datasets containing city, project, and cultural information. Each entry is chunked into segments

(approximately 300–500 tokens with slight overlap) and enriched with metadata such as category and source.

These text chunks are then converted into vector embeddings and stored in a semantic index, which allows Sawwah to perform fast and accurate similarity-based retrieval.

12.4 RAG Implementation

The RAG (Retrieval-Augmented Generation) pipeline forms the core of the Sawwah system. It retrieves the most relevant information chunks from the embedded knowledge base using k-nearest neighbor (k-NN) semantic search. The retrieved context is then passed into a language model to generate concise, grounded, and human-like responses.

Additional "guardrails" were implemented to maintain factuality, prevent repetition, and ensure stylistic coherence across outputs.

12.5 Speech Interaction

The Speech Interaction Module transforms generated text responses into natural-sounding speech using gTTS, AssemblyAI and ElevenLabs API for high-quality neural synthesis. The module includes a simple playback API that streams the audio output, enabling a hands-free, conversational experience for users.

12.6 Computer Vision Greeting

The Computer Vision (CV) Module leverages OpenCV and MediaPipe to detect the presence of a person in front of the camera. When a user is detected for five consecutive seconds, the system triggers a personalized audio greeting such as:
"Welcome to Sawwah, how can I assist you today?"

A debounce mechanism was implemented to prevent repetitive triggers and ensure a smooth interaction flow.

12.7 Web Interface Deployment

To provide an accessible and user-friendly environment for interaction, a Flask-based web application was developed to integrate all system modules Retrieval-Augmented Generation (RAG), Speech Interaction, and Computer Vision Greeting into a single unified interface.

The web layer acts as the front-end gateway for users to interact with Sawwah in real time. It allows them to enter tourism-related questions, receive spoken and written answers, and experience intelligent visual engagement through the camera module.

Key implementation details include:

- Framework: Flask serves as the lightweight backend responsible for handling user requests, invoking the RAG engine, and delivering both textual and audio responses.
- Frontend Components:
  - style.css defines the lavender-themed layout and responsive design.
  - script.js handles asynchronous query submission and plays generated audio responses.
- Backend Components:
  - app.py connects all modules through Flask routes. Each route handles user input, calls the RAG model, triggers the text-to-speech engine (pyttsx3 or ElevenLabs), and returns the synthesized audio.
- Integration Workflow:

  1. User inputs a question via the web interface.
  2. Flask routes the input to the RAG engine for semantic retrieval.
  3. The generated response is displayed and spoken through the TTS module.
  4. If a user is detected via the CV module, the system greets them automatically.

This integrated interface provides a seamless, real-time experience combining intelligent retrieval, speech synthesis, and visual interaction representing the complete

Sawwah prototype. The modular Flask design allows easy deployment to future platforms such as Microsoft Azure or local servers.

12.8 Configuration & Security

To ensure secure deployment, sensitive credentials such as API keys are stored in environment variables managed through a. env configuration file. The file is excluded from version control to maintain confidentiality. The implementation follows the principle of least privilege, granting access only where necessary.

12.9 Logging & Monitoring

Comprehensive logging mechanisms were added to capture essential metrics, including query timestamps, retrieval latency, and speech synthesis status. These logs enable performance evaluation, debugging, and long-term monitoring of user interaction patterns and model behavior.

12.10 Reproducibility

A project runbook was created to document environment setup, library dependencies, and execution steps. This ensures reproducibility and ease of deployment across different environments. The runbook includes procedures for building embeddings, launching the retriever, initiating the Flask web interface, and validating the end-to-end system performance.

14.0 Tasks and Workflow

The development of Sawwah followed a structured, phase-based workflow to ensure systematic progress from data preparation to full system integration. Each phase focused on specific objectives and utilized appropriate tools and frameworks to achieve the desired outcomes efficiently.

The following table summarizes the major phases of the project and the key tasks completed within each stage.

Phase	Description	Tools
Data Preparation and Cleaning	Collected and cleaned multiple English datasets containing information about Saudi cities, projects, and cultural figures. Standardized data formats, removed inconsistencies, and ensured readiness for embedding.	Python, Pandas, NumPy
Exploratory Data Analysis (EDA) and Visualization	Conducted descriptive analysis to understand dataset distributions and relationships. Visualized data trends and patterns to support knowledge-based structuring.	Matplotlib, Plotly
RAG Pipeline Development	Built and tested the Retrieval-Augmented Generation (RAG) pipeline for semantic search and response generation. Implemented chunking, embeddings, and retrieval evaluation.	LangChain, Scikit-learn
Speech Interaction Integration	Developed and connected the text-to-speech module to convert generated responses into natural voice output. Tested with multiple voices for clarity and realism.	pyttsx3, ElevenLabs
Web Interface Development and Computer Vision Greeting	Designed and implemented a Flask-based web interface that enables users to enter questions, view textual and spoken responses, and experience intelligent greetings triggered by the computer-vision module.	Flask, HTML/CSS/JS, OpenCV, MediaPipe

Table 7: Project Phases, Descriptions, and Tools Used

15.0 Results and Discussion

The implementation of Sawwah resulted in a fully functional prototype that successfully integrates the RAG-based retrieval system, speech synthesis module, computer vision greeting, and Flask-based web interface into one cohesive platform.

Testing confirmed that the system retrieves accurate, contextually grounded answers from the curated English datasets, providing users with reliable and relevant information about Saudi tourism, cultural figures, and heritage projects.

The speech interaction module effectively converts generated responses into clear, natural-sounding audio output. During evaluation, both pyttsx3 (offline) and ElevenLabs (neural voices) demonstrated stable performance and satisfactory voice quality.

The computer vision component also performed reliably, detecting user presence and triggering personalized greetings such as "Welcome to Sawwah, how can I assist you today?" The integration of a five-second debounce mechanism successfully prevented repeated triggers, enhancing user experience and interaction flow.

The web interface provided a seamless environment for interaction, allowing users to enter questions, receive both text and spoken responses, and observe real-time greetings through the integrated computer vision module. This design not only enhanced usability and accessibility but also demonstrated the potential of combining intelligent retrieval with multimodal interaction.

Overall, Sawwah met its core objectives of combining intelligent retrieval, speech communication, and visual interaction into a unified AI-driven tourism assistant.

The system proved effective in demonstrating how artificial intelligence can improve accessibility, engagement, and personalization within Saudi Arabia's tourism ecosystem.

15.1 RAG Model Evaluation

The goal of this evaluation is to measure the performance and reliability of the Retrieval-Augmented Generation (RAG) model using both automated and user-driven evaluation.

The assessment focuses on how accurately and clearly the model answers questions based on retrieved knowledge.

Evaluation Setup

1. Model Under Evaluation

   The main RAG system is composed of:

   - A retriever, which fetches the most relevant context from the embedded documents.
   - A generator, which produces an answer using the retrieved context.

2. Judge Model

   To evaluate the RAG model's responses, an open-source model "microsoft/phi-2" was used as a judge model.

It generates numerical scores for each response based on four evaluation metrics:

- Accuracy: How correct the answer is compared to the question.
- Relevance: How well the answer relates to the retrieved context.
- Clarity: How clear and understandable the response is.
- Consistency: How logically consistent and aligned the response is with the context.

User & Automated Evaluation: Automated Evaluation in Batch Mode

A dataset of 100 user prompts was used to automatically evaluate the RAG model. The evaluation process involved the following steps:

- Routing each question to the correct RAG category.
- Retrieving the top documents for contextual grounding.
- Generating answers using the RAG system.
- Evaluating each response with the judge model ("microsoft/phi-2") based on the four metrics.
- Storing all results in the file rag_evaluation_results.csv.

Each entry in the CSV file includes the following fields:

Field	Description
Timestamp	When the evaluation was performed
Question	The input question
RAG_Answer	The model's generated response
Accuracy	Score between 0–1
Relevance	Score between 0–1
Clarity	Score between 0–1
Consistency	Score between 0–1

Table 8: RAG Evaluation Dataset Fields and Descriptions

Interactive Evaluation

The same evaluation process can also be conducted interactively by users through a command-line interface:

- Users type questions manually.
- The RAG model generates responses in real time.
- The judge model evaluates each response using the same four metrics.

- The results are appended to the same file (rag_evaluation_results.csv) automatically.

This dual-mode evaluation enables both automated and live testing, allowing continuous learning and comparison between batch and user-based performance outcomes.

15.2 Evaluation & Performance Metrics

To assess the effectiveness and reliability of the Sawwah Intelligent Tourism Assistant, both quantitative evaluation and human assessment were conducted.

The evaluation focused on four core aspects of the system's performance: retrieval accuracy, response coherence, speech latency, and computer vision detection reliability.

A combination of automated testing and human evaluation using a curated set of tourism-related queries was employed. Fifty (50) test questions were drawn from the English datasets covering Saudi cities, historical figures, and Vision 2030 projects. Each response was rated by evaluators based on contextual accuracy, completeness, and clarity.

Metric	Description	Result
RAG Relevance Accuracy	Percentage of responses judged as contextually relevant and factually correct by human evaluators.	87%
Response Coherence Score	Measure of linguistic fluency, structure, and logical consistency within generated responses.	91%
Speech Latency	Average time (in seconds) required to convert text responses into spoken audio output.	2.8 seconds
CV Detection Accuracy	Success rate of the computer vision module in identifying user presence and triggering greetings under various lighting conditions.	93%

Table 9:System Evaluation & Performance Matrics

The evaluation demonstrated that Sawwah performs robustly across all integrated modules.

The RAG pipeline achieved high relevance and coherence, confirming that the retrieval and generation components are effectively grounded in the curated knowledge base.

The speech synthesis module maintained natural clarity with minimal delay, ensuring a smooth, conversational experience.

Similarly, the computer vision module achieved strong detection accuracy, with stable performance even in moderately low-light environments.

These results confirm the system's reliability and validate its ability to deliver accurate, context-aware, and human-like interactions that enhance user engagement and accessibility in digital tourism applications.

16.0 Risk Assessment

During the development and implementation of Sawwah, several potential risks were identified that could affect the system's performance, reliability, and long-term scalability.

The following table outlines the primary risks along with corresponding mitigation strategies adopted by the team.

Risk Category	Description	Mitigation Strategy
Performance under large datasets	As the knowledge base grows with additional tourism data, retrieval speed may decrease due to the increased number of embedded documents.	Optimize embedding size and chunk length, implement vector caching, and use efficient indexing techniques (e.g., FAISS or Chroma).
Text-to-Speech (TTS) latency	Speech synthesis may experience noticeable delays for long or complex outputs, affecting user experience.	Reduce output length through concise generation prompts, enable streaming playback, and cache frequent responses.
Retrieval bias and data imbalance	Uneven data coverage across different cities or topics could lead to biased or incomplete answers.	Continuously expand datasets, validate results manually, and retrain embeddings periodically with balanced input data.

Integration complexity	Integrating multiple modules (RAG, Speech, CV, and Web Interface) may cause synchronization or compatibility issues.	Follow modular design principles, use standardized APIs, and perform incremental testing of each module.
Hardware and environmental limitations	Real-time vision and audio processing may be affected by camera quality, lighting, or device performance.	Implement error handling for hardware detection, optimize CV frame rates, and test under various environmental conditions.
Security and data confidentiality	Exposure of API keys or local data paths could compromise the system.	Store credentials securely in environment variables and limit access through role-based permissions.

Table 10: Risk Assessment and Mitigation Strategies

17.0 Challenges and Lessons Learned

During the development of Sawwah Intelligent Tourism Assistant, the team encountered several technical and collaborative challenges that required adaptability, experimentation, and effective communication to overcome. These experiences strengthened both the system and the team's practical understanding of AI project workflows.

Challenge	Resolution / Lesson Learned
Data inconsistency across sources	The datasets collected from multiple tourism platforms contained inconsistent formats and missing values. The team applied systematic data cleaning and preprocessing using Python and Pandas, ensuring uniform structure and reliability.
Speech latency and audio quality	Initial voice responses from the text-to-speech module were delayed and unclear. Through tuning voice parameters, reducing response length, and caching repeated outputs, the system achieved faster and more natural speech.
Computer vision detection sensitivity	The CV module struggled under varying lighting conditions. Adjusting MediaPipe detection thresholds and adding a debounce timer improved accuracy and prevented repetitive triggers.
Prompt testing and retrieval accuracy	Early RAG responses sometimes mixed unrelated information. The team refined the chunking size, embeddings, and prompt phrasing until retrieval reached consistent relevance and accuracy.

Parallel teamwork and module integration	Working simultaneously across multiple modules introduced synchronization challenges. Through daily communication, shared notebooks, and continuous testing, the team-maintained alignment and ensured seamless final integration.
Fine-Tuning Decision	The team initially experimented with fine-tuning the model to enhance contextual retrieval but later discontinued this step due to two main reasons: 1. the limited dataset size resulted in weak generalization and low training benefit; and 2. the ongoing development of the system required flexibility and rapid updates, which were better supported by using a pre-trained RAG model rather than a fixed fine-tuned version. This choice allowed continuous data expansion without retraining delays, supporting a mindset of growth and adaptability rather than static stability.
Hardware and Resource Constraints	During local testing, the team faced performance limitations on some devices (e.g., GPU constraints in Jupyter and Colab) that slowed embedding and voice synthesis processes. To overcome this, the team migrated heavier tasks to more powerful runtime environments and optimized configuration settings to reduce processing overhead.

Table 11: Challenges and Lessons Learned

These challenges provided valuable lessons in teamwork, adaptability, and technical problem-solving. They emphasized the importance of experimentation, testing, and collaborative review skills that mirror real-world AI development environments and prepared the team for future large-scale projects.

18.0 Ethical and Privacy Considerations

The Sawwah Intelligent Tourism Assistant was developed with a strong emphasis on ethical AI practices, data privacy, and responsible innovation aligning with the principles of Saudi Vision 2030's digital transformation and ethical governance frameworks. All datasets used in the system were collected from verified, publicly available tourism sources and contained no personal or sensitive user information.

The computer vision module was designed to detect user presence only, without capturing, storing, or processing identifiable facial data. This ensures full compliance with

privacy protection standards and prevents any form of biometric storage. Additionally, sensitive credentials such as API keys and configuration files were securely stored in encrypted environment variables, following the principle of least privilege.

The team-maintained transparency throughout development by documenting data sources, retrieval logic, and access permissions. This approach reinforces the project's commitment to ethical AI deployment, responsible data handling, and user trust values that mirror professional standards in both industry and academia.

18.0 Conclusion and Future Work

The Sawwah project represents a significant step toward the integration of artificial intelligence in Saudi Arabia's tourism sector. By combining Retrieval-Augmented Generation (RAG), speech interaction, computer vision, and dashboard visualization, the system delivers a seamless and human-like experience that enhances accessibility to cultural and tourism information. Through this intelligent architecture, Sawwah demonstrates how AI can bridge the gap between technology and human interaction, transforming the way visitors explore and connect with the Kingdom's rich heritage.

The project also reflects the team's commitment to collaboration, problem-solving, and the application of advanced AI techniques to address real-world challenges. Each module was designed with scalability and sustainability in mind, ensuring that Sawwah can continue to evolve as technology advances.

Looking ahead, several areas of enhancement have been identified. Future development will focus on implementing multilingual support to accommodate a wider range of users, enabling seamless communication in both Arabic and English. The system will also incorporate personalization capabilities, allowing responses to adapt to user preferences and interests. Additionally, dataset expansion and recommendation features will further enrich the assistant's knowledge base, offering tailored suggestions for destinations, activities, and cultural experiences.

As Sawwah continues to grow, it holds the potential to become a benchmark for smart tourism innovation an AI assistant that not only informs but also inspires, welcoming every visitor to discover Saudi Arabia in a truly intelligent and interactive way.

19.0 References

Bhat, N. M. (n.d.). *pyttsx3: Text to speech (TTS) library for Python 2 and 3.* PyPI. https://pypi.org/project/pyttsx3/

Build a retrieval augmented generation (RAG) app: Part 1. (n.d.). *LangChain.*https://python.langchain.com/docs/tutorials/rag/

ElevenLabs. (2023). *ElevenLabs.io.* https://elevenlabs.io

Google. (n.d.). *MediaPipe Solutions guide | Edge.* Google for Developers. https://ai.google.dev/edge/mediapipe/solutions/guide

Google Colaboratory. (n.d.). *Colab.* https://colab.research.google.com

Microsoft. (n.d.). *Visual Studio IDE, Code Editor, Azure DevOps, & App Center.* https://visualstudio.microsoft.com

NumPy. (2009). *NumPy.* https://numpy.org

Open Data Portal. (2024). *Data.gov.sa.* https://open.data.gov.sa

Open Data Portal. (2025). *Tourism datasets.* Data.gov.sa. https://open.data.gov.sa/data/dataset/6f8c28e2-0795-473b-9f9a-6e2f9e1c7904

OpenCV FR – Seventh Sense best face recognition API and server. (2025, July 28). *Seventh Sense.*https://seventhsense.ai/opencv-fr/

Pandas. (2019). *Python data analysis library — pandas.* PyData.org. https://pandas.pydata.org

Plotly. (2023). *Plotly Python graphing library.* Plotly.com. https://plotly.com/python/

Project Jupyter. (2019). *Project Jupyter.* https://jupyter.org

Python Software Foundation. (2019). *Python 3.7.3 documentation.* https://docs.python.org/3/

Scikit-learn. (2024). *Machine learning in Python.* https://scikit-learn.org/stable/

Welcome to Saudi Arabia. (n.d.). *Visit Saudi.* https://www.visitsaudi.com/en

18.0 Appendices