VisuoSpeech Intelligence (VSI)

A Real-Time Multimodal AI System

FIVEMINDS, ONE VISION

- MOHAMED TAHER
- MOHAMED GOMAA
- MARK EDWAR
- EYAD WAEL
- ABDELRAHMAN KHALED

Project Goal:

- To create an intelligent real-time system that:
- Detects and tracks multiple people via camera.
- Recognizes what each person says in Arabic.
- Generates instant informative responses from Wikipedia.
- Provides **visual analytics** showing how many sentences each person (ID) has spoken, **sorted from most to least active**.
- Displays the process through an interactive dashboard.

Technologies Used:

- Built with the following Python libraries:
- tensorflow
- opencv-python
- opencv-contrib-python
- mtcnn
- SpeechRecognition
- PyAudio
- wikipedia
- pandas
- streamlit
- numpy

Demo Summary:

- Start VSI → Camera detects people.
- Speak → VSI recognizes speech.
- VSI fetches instant Wikipedia reply.
- Dashboard updates live with info & analysis.

Applications:

- AI education & teaching demos.
- Smart classrooms.
- Interactive exhibits.
- Research on multimodal human-AI communication.