

AI-Powered Vision Assistance **VisionPal**

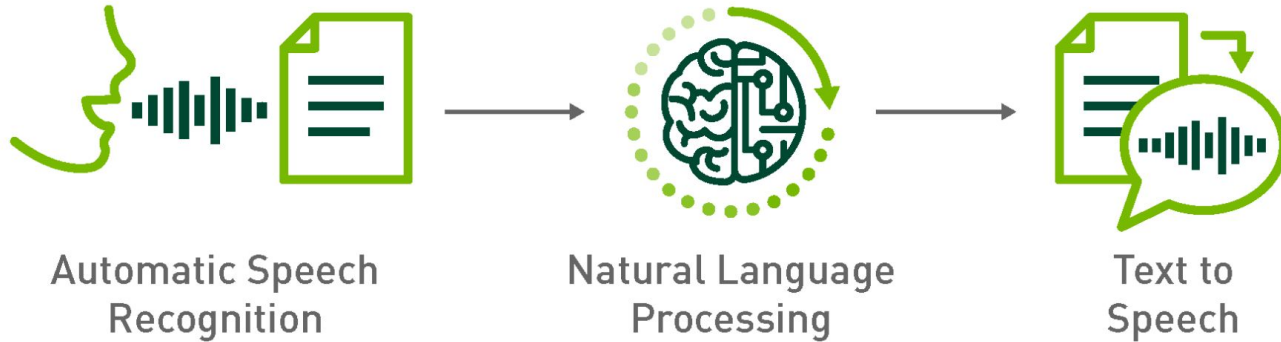
Presented By: Mariam Ashraf Mohamed

Systems & Biomedical Engineering Department,
Faculty of Engineering Cairo University

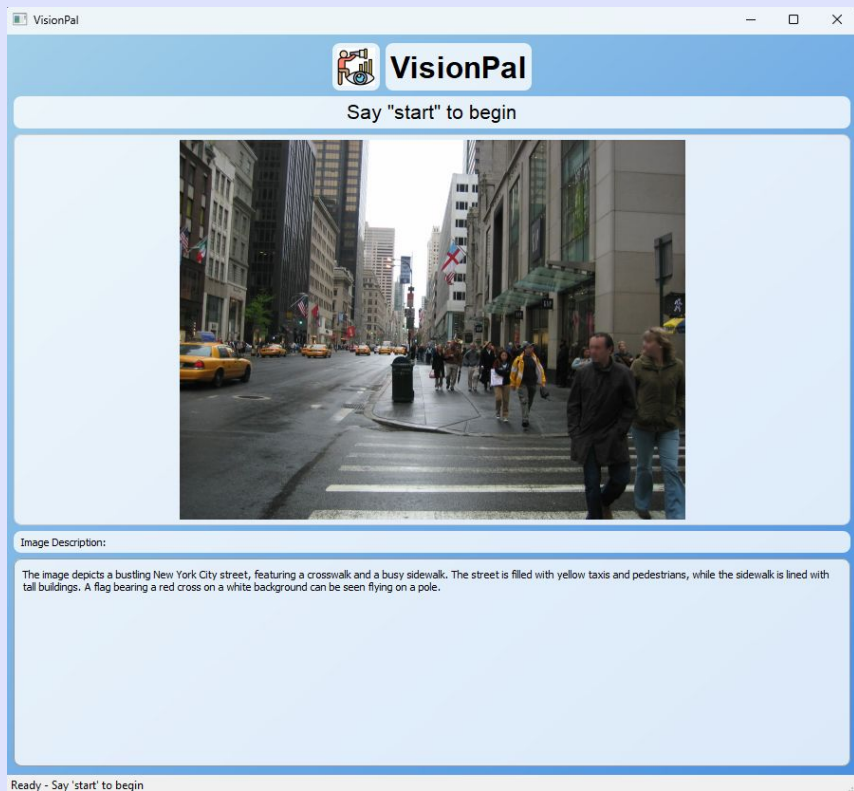
Introduction

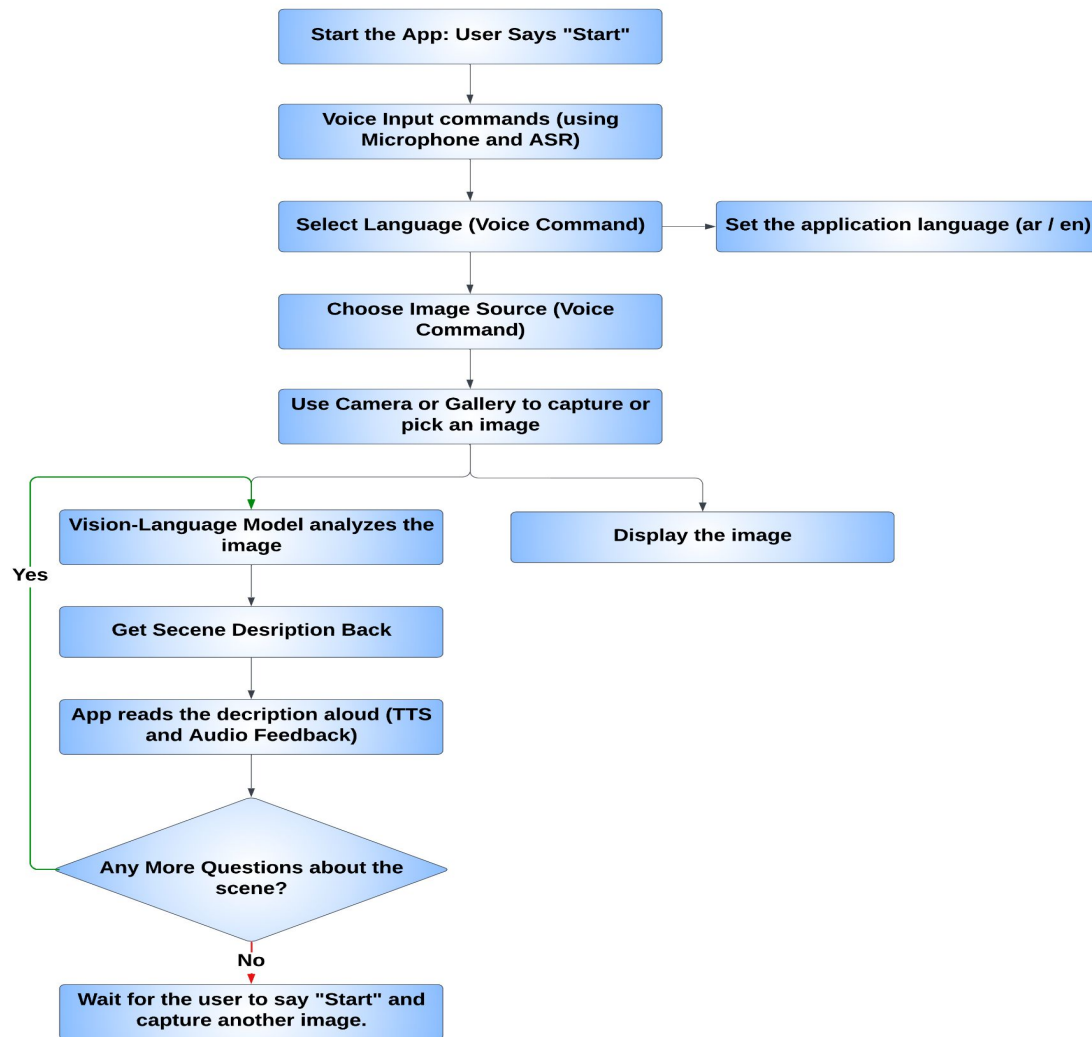
- VisionPal is an AI-powered desktop app.
- It captures images from the webcam and describes them using speech.
- **Purpose: To assist blind and visually impaired users in understanding their surroundings.**

Application Workflow (Speech to Speech)



VisionPal Desktop Application





Current Challenges

● Audio Playback Issues

- Pygame and gTTS audio have conflicts so can't be stopped once started.
- It requires internet access.

● Voice Recognition Problems

- Misinterpretation of voice commands.
- Not reliable in noisy environments.
- It's very very **Unreliable** to make the application works only based on voice commands.
- Higher and more accurate ASR models like Whisper need more complex hardware requirements.

● Slow Performance

- Delay between capturing the image and receiving a description.

Current Challenges

● Language

- Works well with english but doesn't always perform well with the arabic language

● Mobile application

- Couldn't send the image to the online server from the mobile application due to incomprehensible Post failed requests

Future Work



Enhance Voice Interaction

- Support a reliable Arabic language source and clearer command options.
- Add accessibility shortcuts for blind users based on voice commands.



Mobile Version

- Develop an Android version with similar capabilities.
- Deploy Streamlit application and add the other features into it.



SS Model

- Use direct **SPEECH TO SPEECH** models like OpenAi.

Thank You

