INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Building AI Powered Solution for Assisting Visually Impaired Individuals

Darekar Sanket Ravindra

IN9240896

About me

My name is Darekar Sanket Ravindra and I have completed my graduation in Electronics and Telecommunications Engineering. As a recent graduate in Electronics and Telecommunication Engineering, I have developed a strong foundation in technical skills and a passion for exploring the intersection of technology and data.

After my Graduation I independently tried developing my expertise into Data analytics and Data Science and Data Visualization after my graduation. I've always been fascinated with the power of Data to solve complex problems. In college, I worked on a project and used various Data analysis and Data Visualization steps that piqued my interest in Data Science. Since then, I've been working on leveraging my Data analysis and Visualization skills, and I've become even more excited about the field.

I want to engage into different real-life projects and gain hands on experience into different machine learning by leveraging my Data Science skills.

- **Linkedin Profile : Sanket Darekar**
- **Github profile :View On Github**

The Challenge: A World Built for Sight

- **Problem Statement**

For many visually impaired individuals, the world can be a confusing and frustrating place. Everyday tasks that many take for granted, such as reading a menu, crossing the street, or recognizing a friend's face, present significant challenges.

Navigating public spaces, reading visual content, and even simply understanding one's surroundings can be incredibly difficult. We aim to empower these individuals with a solution that offers real-time assistance, enhancing their autonomy and independence.

- **Use case domain : Accessibility and Assistive Technology Purpose**

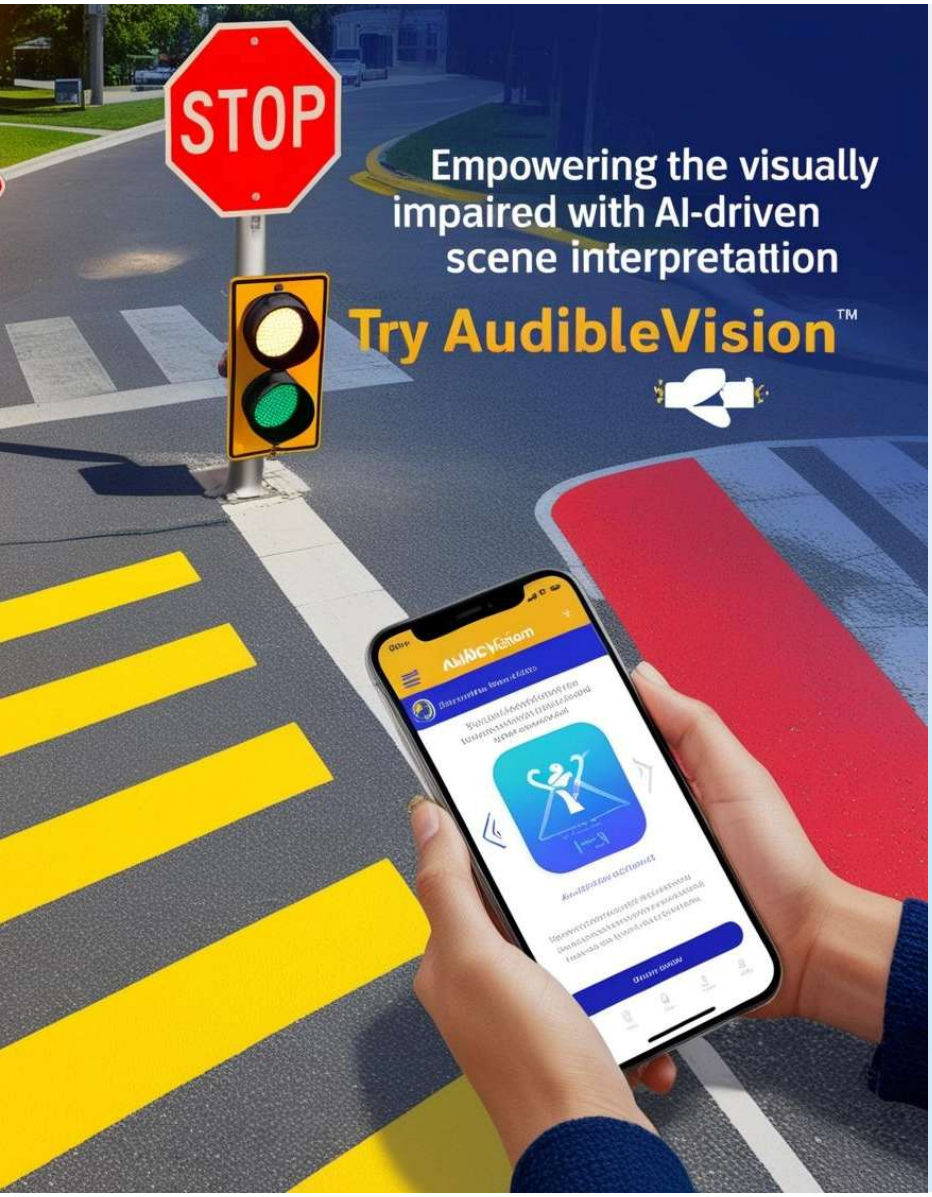
Empower visually impaired individuals to interact with their surroundings through AI..

- **Objective of the Project**

1) Enhance Accessibility for the Visually Impaired
2) Enable Scene Interpretation
3) Implement Text Extraction
4) Leverage Advanced AI Model
5) Create a User-Friendly Interface

Technology Stack

- **Core Libraries and Tools:**
  - **Streamlit:** Interactive web application framework.
  - **pyttsx3:** Offline text-to-speech.
  - **pytesseract:** OCR for text extraction.
  - **Google Generative AI:** Scene analysis using Gemini-1.5-pro.
  - **LangChain Google GenAI:** Enhanced AI integration.



Our Vision: An AI-Powered Assistant

1 Real-Time Scene Understanding

Our app will analyze images captured by the user's device and provide descriptive text, providing context and information about the scene.

2 Text-to-Speech Conversion

This feature will use OCR to extract text from images, converting it to speech, making printed materials, signs, and other visual content accessible.

3 User-Friendly Interface

The application will be designed with an intuitive and accessible interface, making it easy for visually impaired individuals to utilize its features effectively.

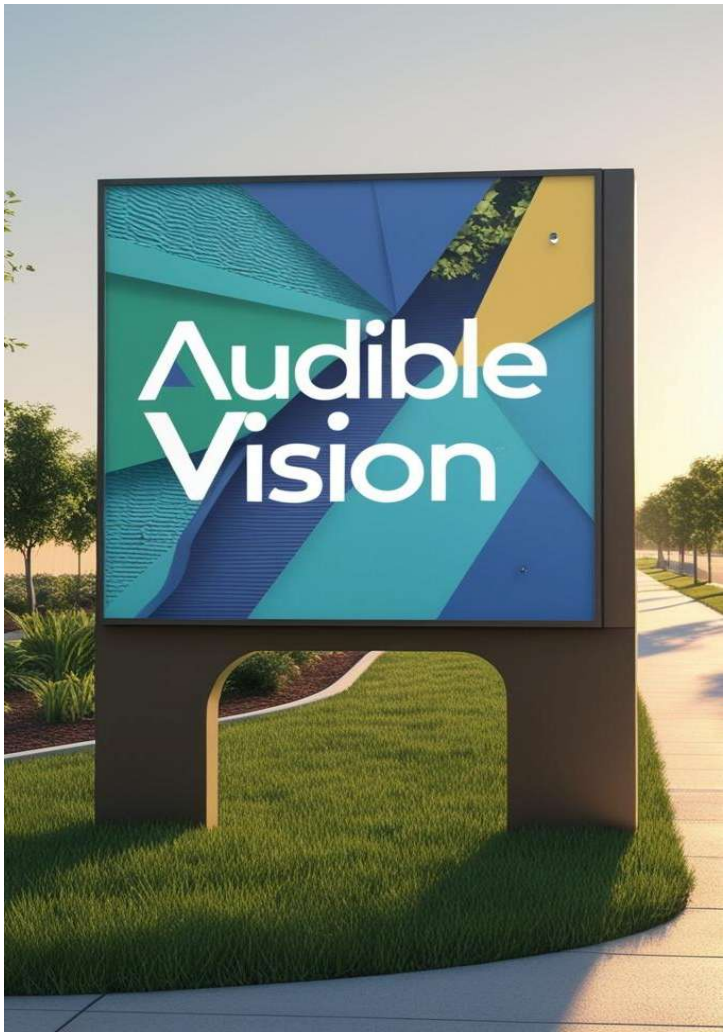
AudibleVision

Introduction

The AudibleVision project is an AI-powered application designed to assist visually impaired individuals in navigating their environment and accessing information with ease. By combining state-of-the-art technologies such as Google Generative AI (Gemini), Tesseract OCR, and text-to-speech (TTS), this project aims to bridge the accessibility gap and empower users in their daily lives.

The primary purpose of the AudibleVision project is:

1) To assist visually impaired individuals by offering tools that enable them to interpret visual content, access textual information, and receive auditory feedback.
2) To enhance accessibility and independence by providing real-time support for various scenarios, from reading documents to understanding the environment.



AudibleVision

Features

1.Scene Analysis

1. AI-powered interpretation of images to describe surroundings, objects, and their context.
2. Assists users in identifying objects, understanding actions, and receiving safety suggestions.

2.Text Extraction

1. Utilizes Optical Character Recognition (OCR) technology to detect and extract text from uploaded images.
2. Enables users to access written content from physical documents, signs, and other visual materials.

3.Text-to-Speech Conversion

1. Converts extracted text and scene descriptions into speech for real-time auditory assistance.
2. Ensures that users can listen to the content without the need for additional support.

Real-Time Scene Understanding: Bringing the World into Focus

- **Image Capture**

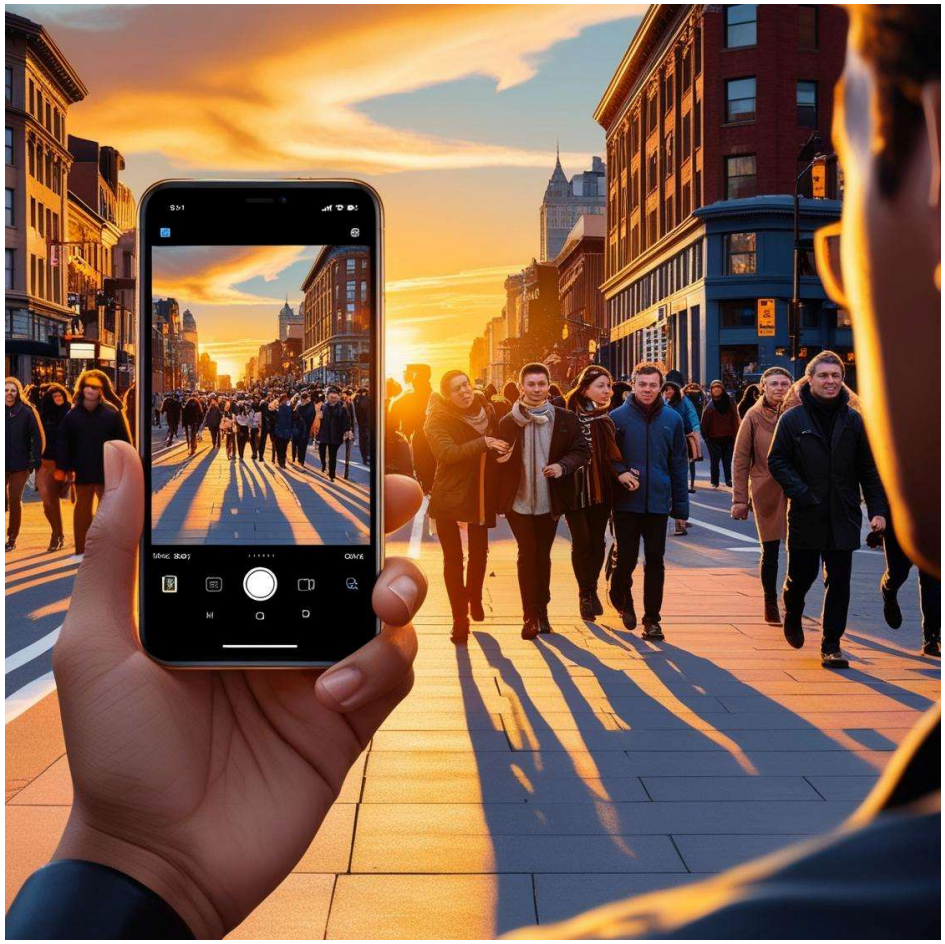
The user captures an image of their surroundings using their phone's camera.

- **AI Analysis**

The AI model analyzes the image, identifying key objects, locations, and relationships between them.

- **Descriptive Text**

The AI generates a detailed, natural language description of the scene, providing the user with a clear understanding of their environment.



Text-to-Speech Conversion: Breaking Down Barriers to Information

- **Image Upload**

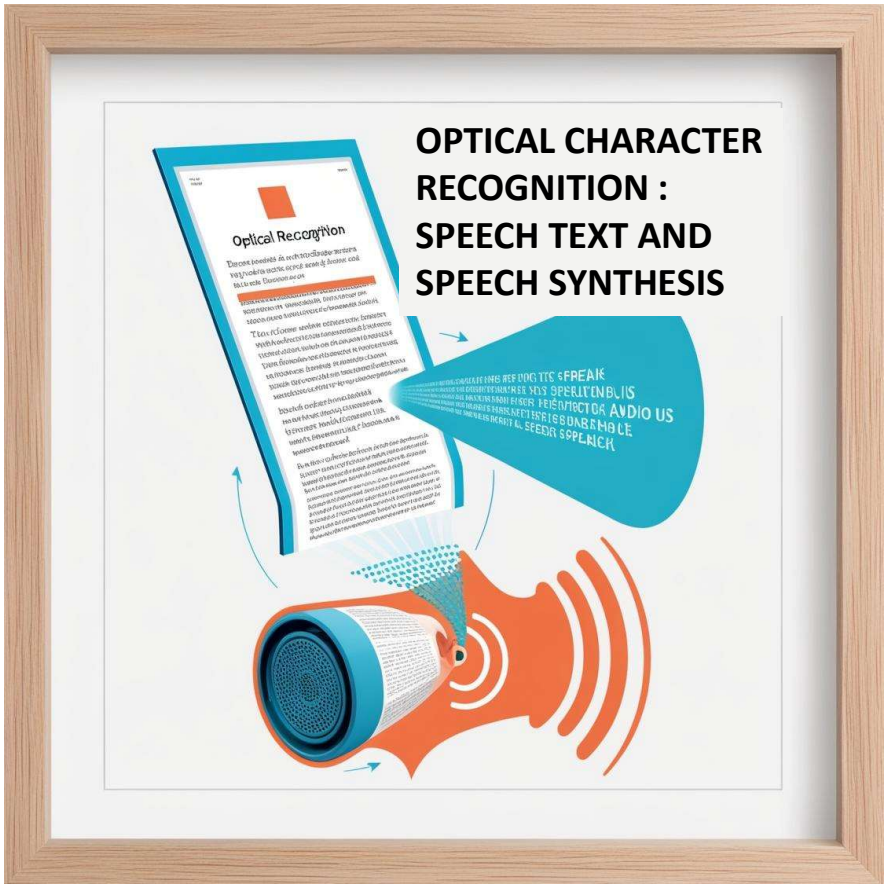
The user uploads an image containing text, such as a document, a sign, or a menu.

- **OCR**

Optical Character Recognition (OCR) technology extracts the text from the image, converting it into a digital format.

- **Speech Synthesis**

The extracted text is then converted into audible speech, allowing the user to access the information quickly and easily.



Experience

Challenges Faced

1.Model Integration Issues:

Integrating AI tools like Google Gemini, Tesseract OCR, and TTS into a seamless workflow was challenging due to compatibility and dependency concerns.

2.Real-time Processing:

Ensuring minimal latency for real-time scene analysis and text-to-speech conversion required optimization.

3.User Accessibility Testing:

Designing a user-friendly interface for visually impaired users and gathering feedback posed unique usability challenges.

Learning and Impact

- **Technical Expertise**:
  - Gained hands-on experience in integrating AI-powered APIs for accessibility solutions and optimizing OCR pipelines.
- **Empathy Development**:
  - Improved understanding of the needs of visually impaired individuals, enabling the design of more inclusive applications.
- **Social Impact**:
  - The project emphasized the role of AI in creating equitable opportunities and empowering underserved communities.

Conclusion

•Successfully developed an **AI-powered** tool integrating **scene analysis**, **OCR**, and **text-to-speech** to assist visually impaired individuals.

•Overcame technical challenges in achieving real-time processing and accessibility optimization through iterative development and testing.

•Gained hands-on experience in **AI model deployment**, **user-centric design.**

•Demonstrated the potential of AI in creating inclusive solutions, **positively impacting the lives of visually impaired users.**

•Identified future enhancements, such as **multilingual support** and **hardware-independent deployment**, for broader accessibility and scalability.

THANK YOU