# AI - Vision Assist

# Building AI Powered Solution for Assisting Visually Impaired Individuals

By jawahar jovah

This project aims to leverage Generative AI to assist visually impaired individuals in perceiving

and interacting with their surroundings.

Visually impaired individuals often face challenges in understanding their environment, reading

visual content, and performing tasks that rely on sight.

An AI-powered solution for assisting visually impaired individuals can revolutionize their daily lives by addressing key challenges like navigation, object recognition, and interaction with digital content. Leveraging technologies such as computer vision, natural language processing, and advanced sensors, such solutions can enable tasks like reading text through OCR, identifying objects and faces, and guiding users in unfamiliar environments with GPS and real-time obstacle detection. For instance, a mobile app or wearable device could use a camera to interpret the surroundings and deliver audio feedback to the user, providing a seamless and intuitive experience. With a focus on affordability, accessibility, and privacy, these solutions can bridge critical gaps and foster independence, ensuring a positive societal impact.

There is a need for an intelligent, adaptable, and user-friendly solution that provides:

● Real-time scene understanding.

● Text-to-speech conversion for reading visual content.

● Object and obstacle detection for safe navigation.

● Personalized assistance for daily tasks and chats

Use cases:

1. Real-Time Scene Understanding

Generate descriptive textual output that interprets the content of the uploaded image,

enabling users to understand the scene effectively.

2. Text-to-Speech Conversion for Visual Content

Extract text from the uploaded image using OCR techniques and convert it into audible

speech for seamless content accessibility.

3. Object and Obstacle Detection for Safe Navigation

Identify objects or obstacles within the image and highlight them, offering insights to

enhance user safety and situation awareness.

4. Personalized Assistance for Daily Tasks

Provide task-specific guidance based on the uploaded image, such as recognizing

items, reading labels, or providing context-specific information.

Model implemented: LangChain, Google Generative AI

# LangChain:

LangChain is a framework designed for building applications powered by large language models (LLMs). It specializes in integrating LLMs with other systems like databases, APIs, or custom logic to enable complex workflows. Key features include:

Prompt Management:

Creates structured prompts for LLMs, improving consistency and reliability.

Memory:

Enables context retention for conversational applications.

Integration:

Connects LLMs with external tools such as search engines, APIs, or custom databases.

Chain Building:

Orchestrates multiple LLM interactions for solving complex problems, such as answering multi-step queries.

# Google Generative AI

Google Generative AI refers to Google's suite of tools and models focused on content generation. These are part of the Vertex AI ecosystem and include cutting-edge LLMs capable of generating text, summarizing information, and even creating code. Highlights include:

Palm 2 Models:

Google's advanced LLMs for various use cases, including text generation, summarization, and Q&A.

Easy Deployment:

Pre-integrated into Google Cloud, simplifying integration with existing infrastructures.

Scalability:

Designed for handling large-scale applications efficiently.

Customizable:

Allows fine-tuning for specific business or user needs.

# About the user interface:

## Camera:

This feature enables users to capture a photo in real-time, which is then processed using advanced AI algorithms to generate a detailed description of the image. The description provides a comprehensive understanding of the contents within the picture, including objects, people, activities, and the overall context. By analyzing visual elements and translating them into meaningful insights, this functionality helps users, particularly those with visual impairments, gain a clear idea of what the image depicts. For example, a photo taken of a park might be described as "a sunny park with green trees, a walking path, and people sitting on benches," offering an accurate mental picture of the scene.

## Daily Assist:

This feature empowers users to complete daily tasks efficiently by leveraging LangChain AI for intelligent assistance and a photo upload option for visual analysis. With this functionality, users can upload an image, and the system will process it to extract meaningful insights, such as identifying objects, interpreting text, or analyzing scenes. Additionally, users can engage in an interactive chat to refine their queries and receive accurate, context-aware information. For example, you can upload a photo of a product label, extract its details, and ask follow-up questions like where to find similar items or how to use the product. This combination of visual processing and conversational AI creates a seamless experience, making it easier to address queries and accomplish tasks in a smarter, more intuitive way. The voice assistant describes the identified objects, their relative positions, and any critical warnings, making the feature especially beneficial for visually impaired individuals or those navigating unfamiliar environments.

## Object Detection:

This feature allows users to upload a photo, which is then processed using advanced computer vision algorithms to identify objects, obstacles, or key elements within the image. The system highlights these items and provides a detailed analysis, offering valuable insights to improve user safety and situational awareness. For instance, if the photo depicts a room, the system might identify furniture, doorways, and potential hazards such as misplaced items on the floor. This analysis is further complemented by live voice assistance, which delivers the information audibly in real time. The voice assistant describes the identified objects, their relative positions, and any critical warnings, making the feature especially beneficial for visually impaired individuals or those navigating unfamiliar environments. By combining visual analysis with audio feedback, the feature ensures a seamless, accessible, and practical experience.