Hear the World

AI-Powered Real-Time Scene Narration for the Visually Impaired

### **Abstract**

This project presents a deep learning-based tool designed to enhance the independence of visually impaired individuals by converting real-world visual scenes into detailed, context-aware audio descriptions. The system utilizes a combination of computer vision and large language models (LLMs) to process live camera inputs, recognize objects and text in the environment, and generate coherent and descriptive narratives. The process begins with capturing a video feed using wearable devices or smartphones, followed by object detection and scene understanding for textual information.

The extracted text data is then passed to an LLM, which generates a detailed, human-like description of the scene. The description is subsequently converted into real-time audio using a text-to-speech (TTS) engine, providing users with a natural and accessible audio feedback. By offering real-time, comprehensive information about their surroundings. This tool empowers visually impaired individuals to navigate and interact with their environments more effectively. The system is optimized for real-time performance and aims to work in diverse environments, with future enhancements and integration with mobile application.