



INNOVATION. AUTOMATION. ANALYTICS

## PROJECT ON

# AI-Powered Solution for Assisting Visually Impaired Individuals

By- Yuvraj Kadam

# Problem Statement

## Objective

Design an AI-driven solution that provides:

- ***Real-time scene comprehension*** to understand surroundings dynamically.
- ***Visual-to-audio transformation*** by converting visual information into speech.
- ***Object and obstacle recognition*** to ensure safe navigation.
- ***Customized support*** for daily activities to enhance independence.

## Challenge

- *Visually impaired individuals encounter significant challenges in:*
- *Understanding their environment and navigating safely.*
- *Accessing written information and performing tasks requiring vision.*
- *Relying on existing assistive technologies, which often lack the capability to offer a holistic solution for real-time navigation and contextual awareness.*

# Project Scope & Requirements

## Key Deliverables

Develop a solution incorporating at least two of the following functionalities:

- ***Dynamic Scene Interpretation*** for understanding real-time environments.
- ***Visual-to-Audio Conversion*** to narrate visual content.
- ***Obstacle and Object Identification*** to facilitate safe navigation.
- ***Tailored Daily Assistance*** to support individual needs.

## Tools and Technologies

- ***Streamlit***: To create an interactive and user-friendly web interface.
- ***LangChain***: For seamless integration with generative AI models.
- ***Google Generative AI***: To power advanced AI functionalities.

# Main Page

The screenshot shows a web browser window with the title 'Vocalize AI' and the address bar displaying 'localhost:8508'. The browser's address bar includes navigation icons (back, forward, refresh) and a search icon. The page content is divided into a left sidebar and a main area. The sidebar, titled 'Features', lists three capabilities: Scene Understanding, Text-to-Speech, and OCR. The main area features the 'Vocalize AI' logo, a descriptive tagline, and an image upload section with a drag-and-drop area and a 'Browse files' button. A 'Deploy' button is located in the top right corner of the page.

Deploy

## Features

- **Scene Understanding:** Describe the scene in an image.
- **Text-to-Speech:** Hear the extracted text or scene description.
- **OCR:** Extract text from images.

# Vocalize AI

Assistive tool for visually impaired individuals. Upload an image to get started!

Upload an Image (JPG, PNG)

Drag and drop file here  
Limit 200MB per file • JPG, JPEG, PNG

Browse files

Upload an image to get started.

# Uploaded Image


Vocalize AI

localhost:8508

Deploy

## Features

- **Scene Understanding:** Describe the scene in an image.
- **Text-to-Speech:** Hear the extracted text or scene description.
- **OCR:** Extract text from images.



Type here to search

3:08 PM  
30-Nov-24



# Image To Text

The screenshot shows a web browser window with the address bar at `localhost:8510`. The application interface includes a sidebar with 'Features' and a main content area with three buttons: 'Describe Scene', 'Extract Text', and 'Text-to-Speech'. The 'Describe Scene' button is active, and a progress indicator shows 'Generating scene description...'. The 'Scene Description' section lists items and their purposes, such as the Statue of Ganesha, Pink Dhoti, Gold Ornaments, Hibiscus Flowers, Silver Platters, Incense Sticks, Gold Beaded Chains with Bells, Small Plant, and Raised Platform/Altar.

**Features**

- **Scene Understanding:** Describe the scene in an image.
- **Text-to-Speech:** Hear the extracted text or scene description.
- **OCR:** Extract text from images.

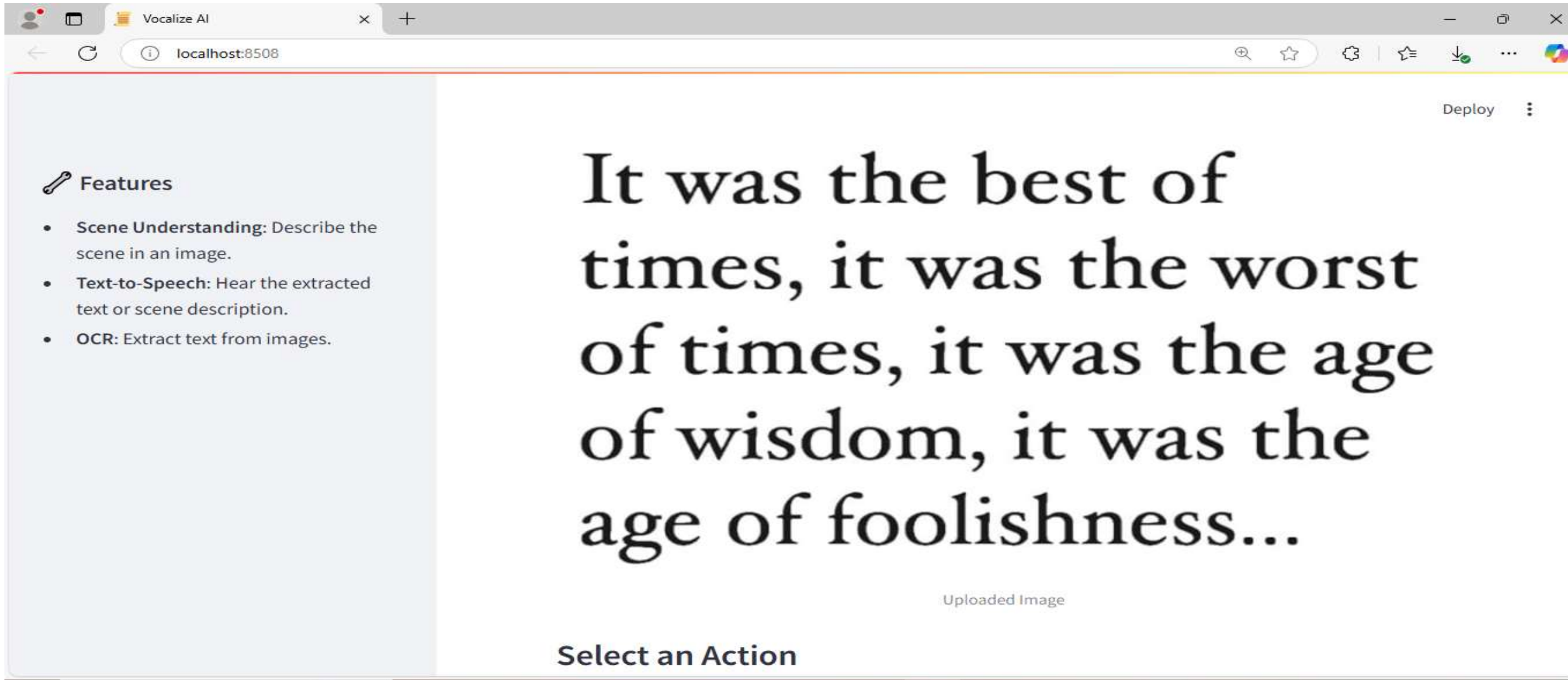
**Describe Scene** **Extract Text** **Text-to-Speech**

Generating scene description...

## Scene Description

1. **Items and their purpose:**
  - **Statue of Ganesha:** The central figure, a Hindu deity revered as the remover of obstacles and the god of beginnings.
  - **Pink Dhoti:** A lower garment worn by Ganesha, often brightly colored.
  - **Gold Ornaments:** Jewelry adorning Ganesha, signifying his divine status.
  - **Hibiscus Flowers:** Placed as offerings to Ganesha, symbolizing devotion and purity.
  - **Silver Platters:** Likely containing offerings of sweets or other items.
  - **Incense Sticks:** Used for creating a fragrant atmosphere during worship.
  - **Gold Beaded Chains with Bells:** Decorative elements hanging on either side, adding to the festive atmosphere.
  - **Small Plant:** Potted greenery in the corner, contributing to the decorative setup.
  - **Raised Platform/Altar:** The surface on which Ganesha is placed, elevating the deity and creating a dedicated space for worship.

# Text Image Upload



The screenshot shows a web browser window with the address bar at `localhost:8508`. The page title is "Vocalize AI". On the left, a sidebar lists features: Scene Understanding, Text-to-Speech, and OCR. The main area displays a large text block extracted from an image, with the label "Uploaded Image" below it. At the bottom, there is a button labeled "Select an Action".

**Features**

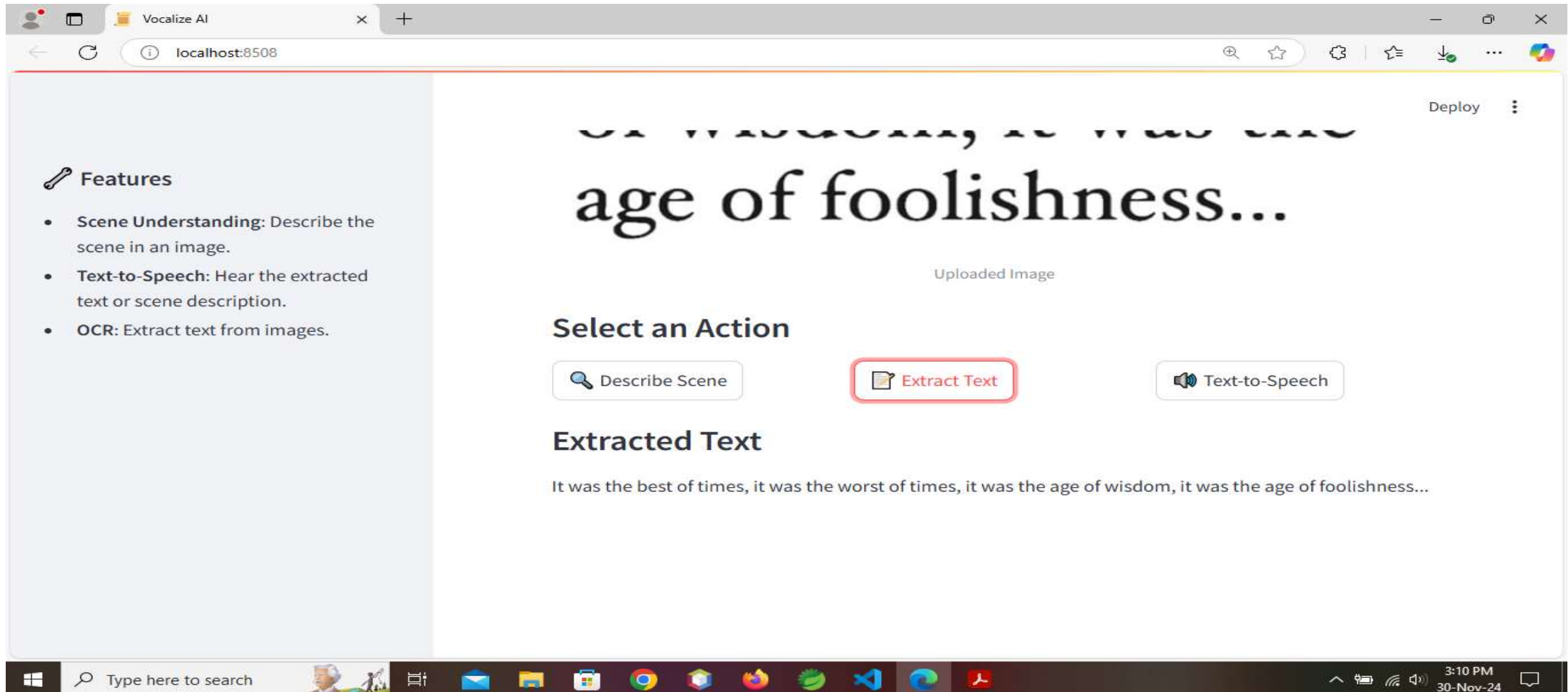
- **Scene Understanding:** Describe the scene in an image.
- **Text-to-Speech:** Hear the extracted text or scene description.
- **OCR:** Extract text from images.

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness...

Uploaded Image

Select an Action

# Extracted Text



The screenshot displays the Vocalize AI web application running on a local host. The interface includes a sidebar with features, a main content area with an uploaded image, action buttons, and the extracted text.

**Features**

- **Scene Understanding:** Describe the scene in an image.
- **Text-to-Speech:** Hear the extracted text or scene description.
- **OCR:** Extract text from images.

**age of foolishness...**

Uploaded Image

**Select an Action**

- Describe Scene
- Extract Text**
- Text-to-Speech

**Extracted Text**

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness...



# Key Challenges and Proposed Solutions

## Challenge 1: Accurate Scene Understanding and Text Extraction

- ✓ **Solution:** Utilize cutting-edge AI models for image recognition and Optical Character Recognition (OCR) to ensure precise visual analysis and text extraction.

## Challenge 2: Intuitive Interaction for Non-Technical Users

- ✓ **Solution:** Develop a user-centric interface using Streamlit, incorporating clear step-by-step guidance and real-time feedback to enhance accessibility and ease of use.

## Challenge 3: Maintaining Real-Time Processing Efficiency

- ✓ **Solution:** Optimize performance with streamlined API calls and advanced pre-processing methods to deliver a seamless and uninterrupted user experience.

THANK  
YOU

