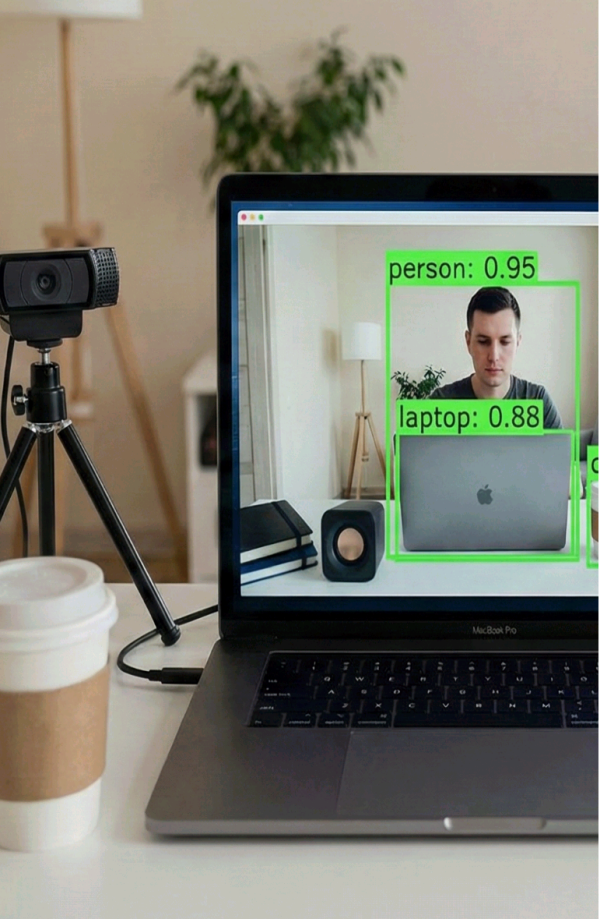



# Object Detection System Using pyttsx3 and SAPI5

A real-time visual recognition system with intelligent audio feedback for enhanced accessibility and automation

Team Members:

Md Rahmat Ansari 0131CL231051

**Guide:** Prof. Ruchi Chaturvedi

**Department:** CSE AI & ML

**Institution:** JNCT Bhopal

**Session:** 2025 – 26

# Introduction: Bridging Vision and Sound

## Project Overview

This project focuses on developing an advanced object detection system with an integrated auditory feedback mechanism, specifically using the pyttsx3 SAPI5 engine. Our goal is to create a more intuitive and accessible interaction with detected objects.

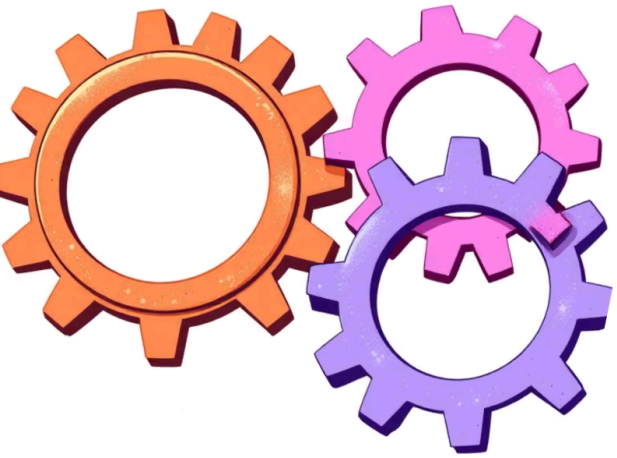
## Problem Domain & Motivation

Current object detection often relies solely on visual output. There's a critical need for systems that can verbalise identified objects, particularly for visually impaired users, or in environments where visual attention is divided. This project addresses this gap by enhancing situational awareness through spoken descriptions.

## Project Importance

The integration of object detection with text-to-speech provides significant benefits across various sectors, including accessibility tools, industrial safety, and smart environments, making visual information universally understandable. It represents a significant step towards more inclusive AI applications.

# Project Objectives: Our Guiding Principles



### Develop a Robust Object Detection Model

To implement and fine-tune a state-of-the-art object detection model capable of accurately identifying diverse objects in real-time video streams.

### Integrate SAPI5 Text-to-Speech

To seamlessly integrate the pyttsx3 library with the SAPI5 voice engine for clear and natural-sounding verbalisation of detected objects.

### Enhance User Accessibility

To provide an intuitive and accessible interface that delivers real-time auditory feedback, significantly improving the experience for all users, particularly those with visual impairments.

### Optimise Performance

To ensure the system operates with minimal latency, providing prompt object detection and verbalisation suitable for real-time applications.

# Methodology & Design: The Workflow

### Acquire Video Feed

Capture real-time video input from camera.

### Process Frames

Pre-process frames for model compatibility (resizing, normalisation).

### Run Object Detector

Apply selected CNN model (e.g., YOLOv5, EfficientDet) to identify objects.

### Generate Text Output

Extract object labels and confidence scores; form descriptive sentences.

### Verbalise via SAPI5

Use pyttsx3 with SAPI5 for audio playback of descriptions.

### Continuous Feedback

Loop the process for continuous, real-time auditory updates.

# Key Technologies Utilised

Our project leverages a robust stack of tools and frameworks to achieve its objectives.

### Python

Primary programming language for development.

### PyTorch / TensorFlow

Deep learning frameworks for model implementation.

### OpenCV

For real-time video processing and image manipulation.


### pyttsx3

Python text-to-speech library for SAPI5 integration.

### SAPI5 (Windows)

Speech Application Programming Interface for natural voices.

### GPU Acceleration

Leveraging CUDA/cuDNN for efficient model inference.

# Audio Feedback: Text-to-Speech with pyttsx3

### Intelligent Conversion

Transforms detected object labels into clear, spoken alerts for immediate user awareness

### Offline Capability

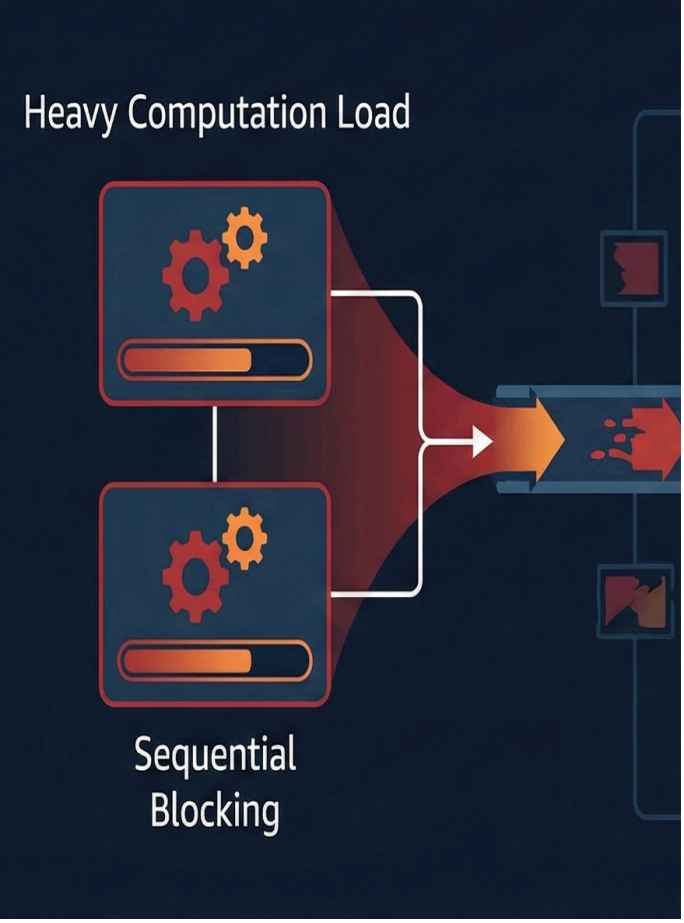
Functions entirely offline, ensuring low latency and zero internet dependency for reliable operation

### Customisable Properties

Adjustable voice parameters including speech rate, volume levels, and voice type selection

### Enhanced Accessibility

Significantly improves system accessibility and creates intuitive user interaction through multimodal feedback




# Technical Challenge: Video Lag

### Heavy Computation Load

Real-time video processing demands intensive model inference computations that can overwhelm single-threaded architectures

### Sequential Blocking

Traditional sequential processing causes frame capture and display operations to block each other, creating visible lag

### Degraded Experience

System lag reduces responsiveness, creates frustrating delays, and compromises the real-time nature of detection feedback

# Real-World Applications

### Smart Surveillance

Intelligent monitoring systems with audible alerts for detected intruders, suspicious objects, or unusual activities in secured areas

### Assistive Technology

Life-changing tools for visually impaired users through spoken object identification and environmental awareness feedback

### Industrial Automation

Real-time monitoring of production lines with audio warnings for quality control, safety hazards, and process anomalies

### Robotics Integration

Enhanced environment awareness for autonomous robots with multimodal feedback combining visual detection and audio responses

# Conclusion & Future Scope

## Key Achievements

- Successfully integrated YOLOv10 and pyttsx3 with SAPI 5 architecture for seamless real-time operation
- Multithreading and queue system effectively eliminate video lag and ensure smooth performance
- Created accessible, efficient embedded vision solution with multimodal feedback

## Future Enhancements

- Optimise model size and computational requirements for SAPI 5 hardware constraints
- Add multilingual text-to-speech support for broader accessibility
- Expand system to support multi-camera setups for comprehensive coverage