**Authors:**

**Maime Pheello**

**Mololi Rammalane**

**Mothusi Project**

**Background**

In recent years, the integration of Internet of Things (IoT) technologies into everyday life has transformed how we interact with our environment. Smart devices equipped with sensors and AI capabilities are increasingly being utilized to enhance user experiences across various domains, including home automation, healthcare, and personal assistance. One of the most promising applications of IoT is in the realm of object recognition and voice interaction, which can significantly improve accessibility and usability for individuals, particularly those with disabilities or limited mobility.

The Mothusi project aims to leverage these advancements by creating an IoT system that combines object recognition, voice command processing, and real-time voice feedback. This system is designed to assist users in identifying objects in their surroundings and responding to voice commands, thereby enhancing their interaction with the environment.

**Problem Statement**

Despite the advancements in IoT and AI technologies, there remains a significant gap in the availability of intuitive systems that can seamlessly integrate object recognition and voice interaction for everyday users. Many existing solutions either focus solely on object recognition or voice command processing, lacking the ability to provide a cohesive user experience that combines both functionalists.

The limitation can hinder individuals, especially those with visual impairments or other disabilities, from effectively interacting with their surroundings. The absence of real-time feedback and the inability to initiate actions through voice commands can lead to frustration and reduced independence for these users. The Mothusi project seeks to address this gap by developing a comprehensive IoT system that not only recognizes objects but also allows users to engage with the system through natural voice commands. By providing immediate audio feedback upon object detection and responding to user inquiries, Mothusi aims to create a more accessible and user-friendly environment, ultimately enhancing the quality of life for its users.

**Motivation**

The motivation behind the Mothusi project stems from a desire to enhance accessibility and improve the quality of life for individuals, particularly those with disabilities. As technology continues to evolve, there is a growing need for innovative solutions that empower users to interact with their environment more effectively.

Many existing technologies either focus on object recognition or voice interaction separately, leaving a gap in comprehensive solutions that combine both functionalities. By addressing this gap, Mothusi seeks to create a more intuitive and user-friendly experience that meets the needs of users in their daily lives.

Also, Mothusi project is rooted in the aspiration to create a positive social impact. By providing users with tools to better understand and interact with their environment, the project aims to enhance the overall quality of life for individuals, fostering inclusivity and independence in a technology-driven world.

Ultimately, this project also serves as a platform for exploring advancements in AI and IoT technologies. By developing a system that combines these cutting-edge technologies, Mothusi contributes to ongoing research in the fields of machine learning, natural language processing, and human-computer interaction.

**Aim:**

The primary aim of the Mothusi project is to develop an innovative IoT system that seamlessly integrates object recognition, voice command processing, and real-time voice feedback to enhance user interaction with their environment, by providing them with the tools to identify objects and engage with technology through natural language.

**Objectives:**

1. Implement a camera-based solution for accurate and fast object detection using a pre-trained AI model.
2. Create a voice processing module to capture and interpret user commands for initiating object recognition.
3. Design a mechanism to deliver clear audio responses based on object detection and user commands.

**System Requirements**

Functional requirements

* The system shall capture images in real-time using a camera.
* The system shall process these images with a pre-trained AI model to detect and classify objects.
* The system shall trigger voice feedback to announce identified objects.
* The system shall capture voice commands using a microphone.
* The system shall interpret specific commands (e.g., "What is this?", "Identify object") using a speech recognition engine.
* The system shall initiate object recognition upon receiving a relevant voice command.
* The system shall allow users to initiate object recognition through voice commands
* The system shall provides real-time audio feedback using a speaker

Non-functional requirements

* The object recognition model shall achieve an accuracy rate of at least 90%.
* The system shall process images and provide feedback with a latency of no more than 2 seconds.
* The system should have an intuitive user interface that is easy to navigate and understand.
* Voice feedback should be clear, natural-sounding, and easily understandable for all users.
* The system should operate consistently under varying environmental conditions (e.g., lighting), with minimal false positives in object detection.
* The system should be designed for easy updates and maintenance, allowing developers to improve algorithms or fix issues efficiently.

Hardware Components:

* Camera: Used to capture live images for object recognition.
* Microphone: Captures voice commands for processing.
* Speaker: Provides audio feedback to the user.
* Raspberry Pi o for compact IoT solutions.

Software Components:

* YOLO (You Only Look Once): Fast and accurate object detection model suitable for real-time applications.
* Tensor Flow Lite: For running pre-trained models locally on small devices.
* Programming Language: Python

System architecture

