"MyGuide: A Real-Time Object Detection Walking Guide for the Visually Impaired“

Developed
By
FRANKBOTICS Tech. Ltd.
And

Rev. A. Ovwonuri
of Software Engineering Department, FUTO Owerri.

A demonstration and presentation made during the 3rd edition of the

Nigeria Computer Society AI and Robotics Summit

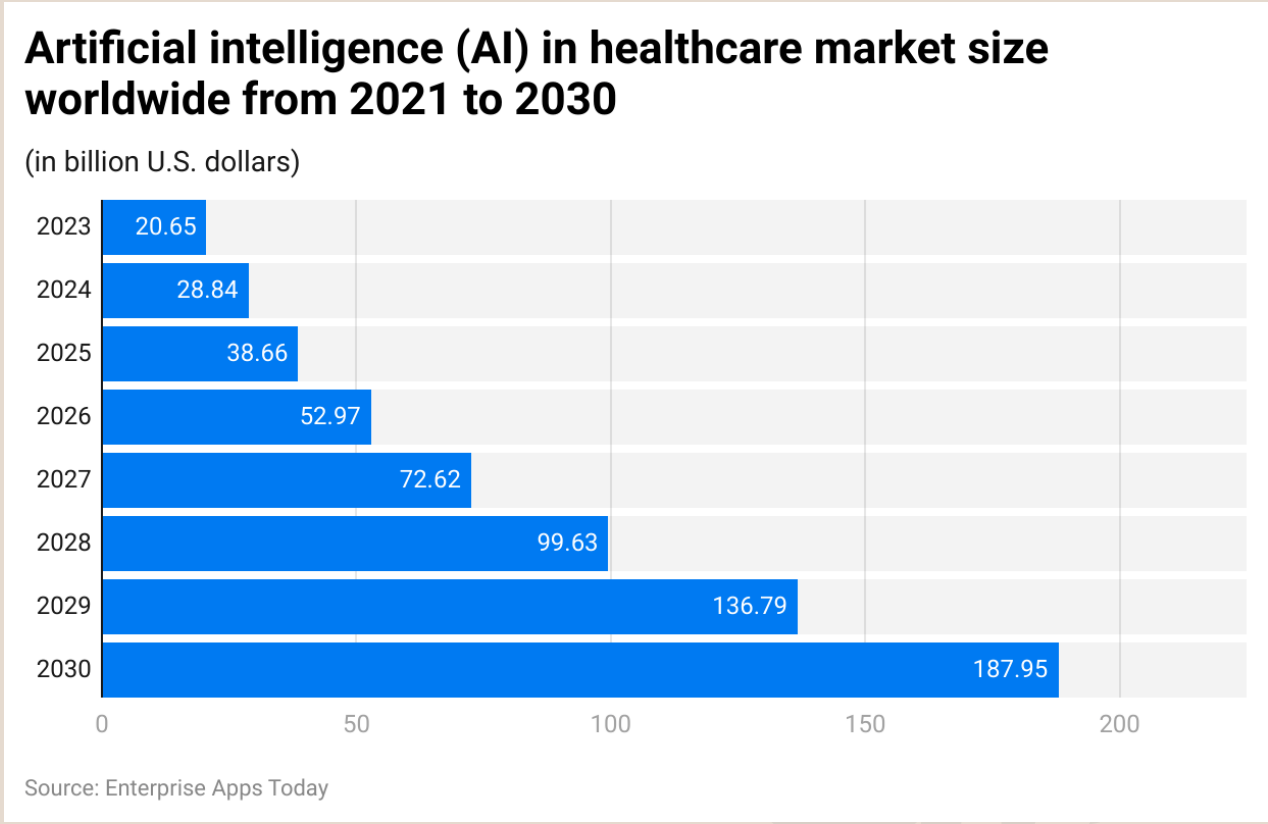
March 5th – 7th, 2024.

TABLE OF CONTENT

PREDICTED AI FUNDING IN THE HEALTHCARE INDUSTRY

According to Statista, the worldwide healthcare market length is expected to reach $187.95 billion by 2030, with projections reaching $20.65 billion in 2023.

In the coming year, the expected market size of this sector is followed by 2024 ($28.24 billion), 2025 ($38.66 billion), 2026 ($52.97 billion), 2027 ($72.62 billion), 2028 ($99.63 billion), and 2030 ($187.95 billion). The adoption of AI technologies in the healthcare region is predicted to enhance affected person effects, reduce scientific errors, and decorate general first-rate care.



Various applications of AI in healthcare:

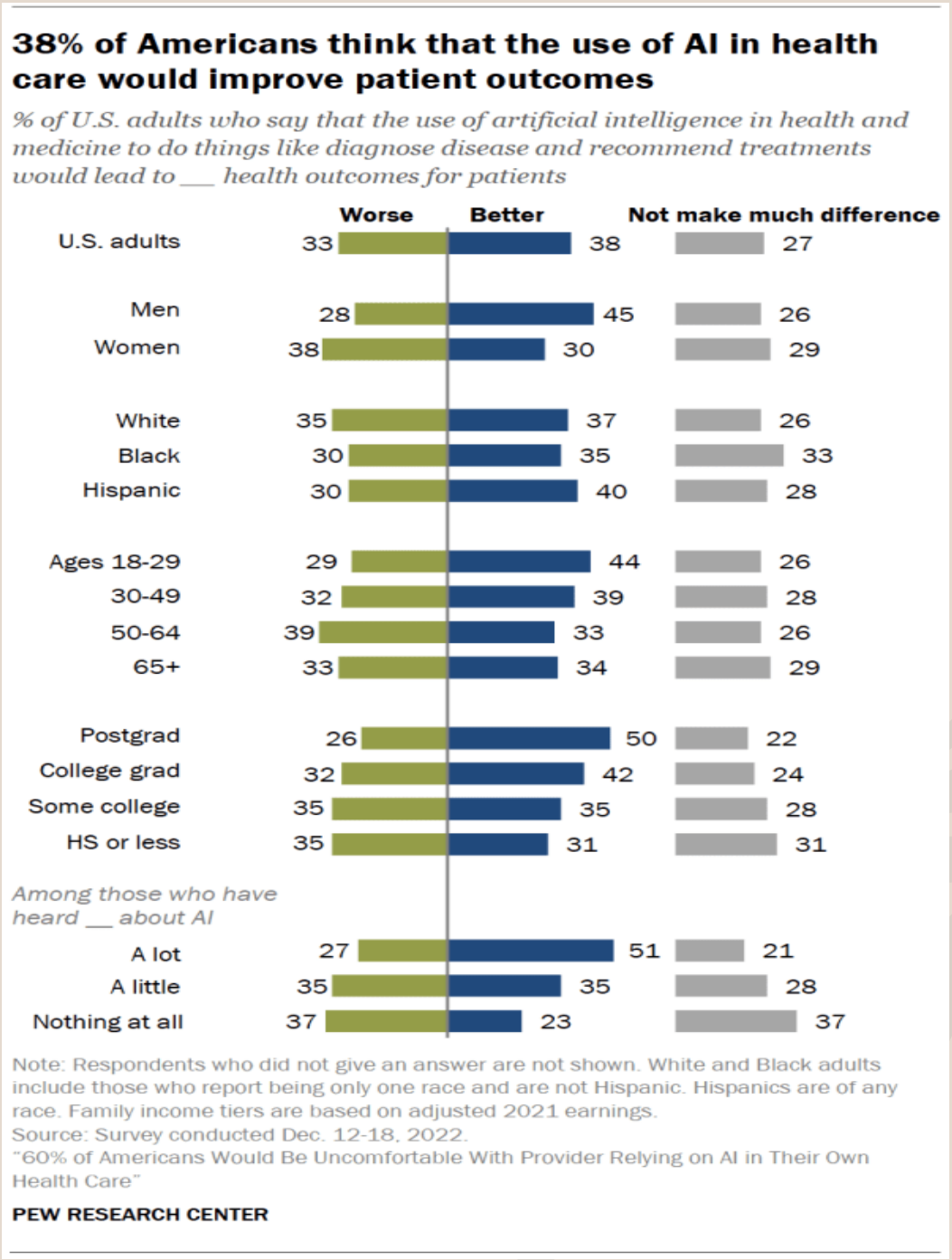
Precision Diagnostics

- Diagnostic imaging
- Diabetic retinopathy screening

Precision therapeutics

- Immunomics and synthetic biology
- AI-driven drug discovery

These are just to mention a few applications of AI in the healthcare industry.



PROBLEM STATEMENT:

The proposal doesn’t eradicate but proposes a measure to assist and enhance efficiency, majorly healthcare. The proposed goal is to assist people with visual impairment. The model we implemented is just a base case for the proposed goal.

We are proposing an intelligent walking stick that utilizes object detection technology to identify objects in its surroundings. Upon detection, the device employs the Google Text-to-Speech (GTTS) API to pronounce the object's name. The system should support object detection and pronunciation in four languages: English, Igbo, Hausa, and Yoruba. The final model will be deployed using Streamlit for user-friendly interaction.

AIMS AND OBJECTIVES:

- Develop a walking stick integrated with object detection technology to enhance the user's awareness of their surroundings.

- Implement the Google Text-to-Speech (GTTS) API to enable the walking stick to pronounce the names of detected objects.

- Extend the functionality to support object detection and pronunciation in multiple languages, including English, Igbo, Hausa, and Yoruba.

- Integrate the system with Streamlit for seamless deployment and user interaction.

- Ensure user-friendliness and accessibility in the design and implementation of the walking stick solution.

Terms used:

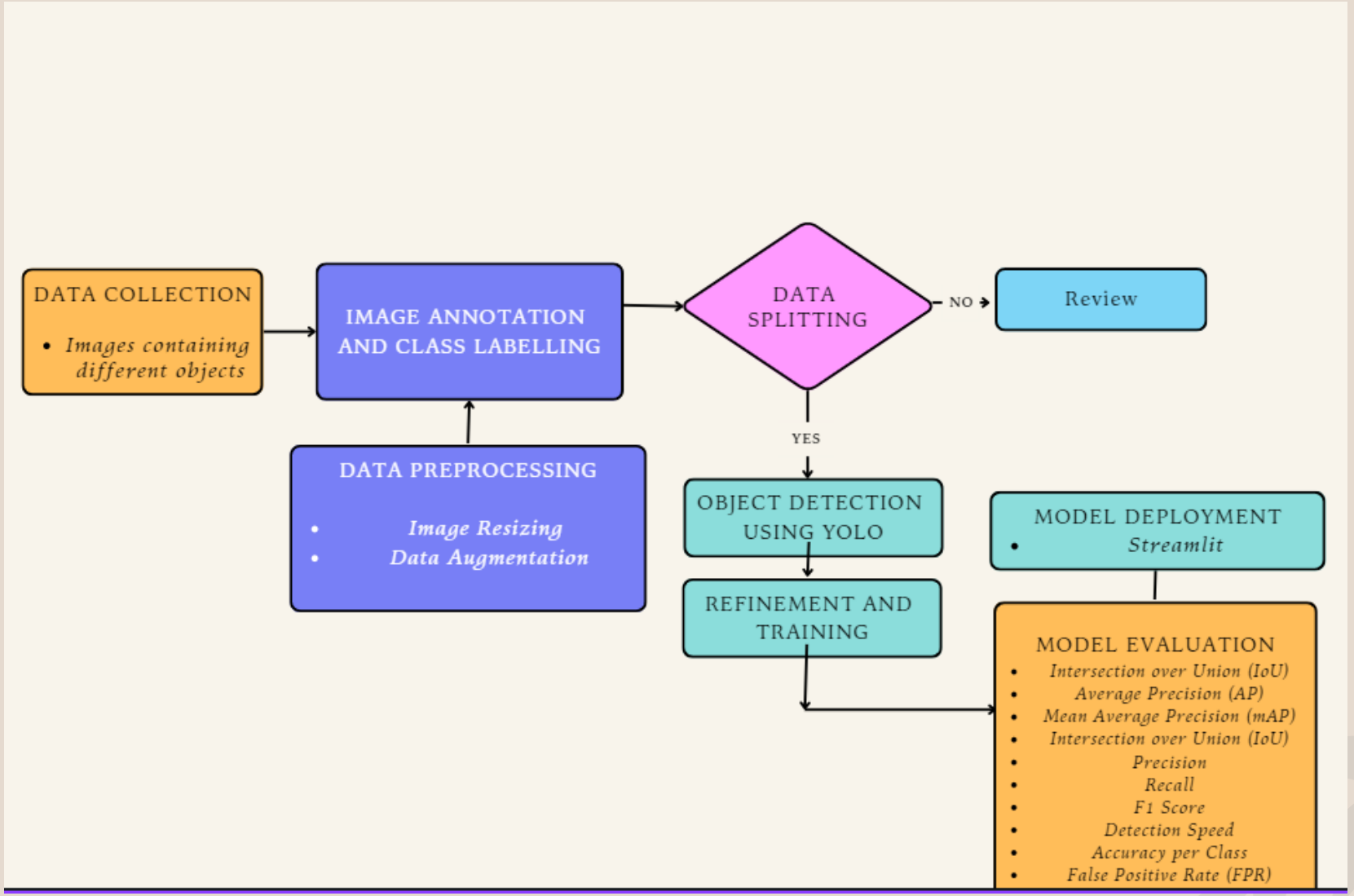
1. Object detection is a computer vision technique for locating instances of objects in **images** or videos. Object detection algorithms typically leverage machine learning or deep learning to produce meaningful results.

2. **Object detection** is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well-researched domains of object detection include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance.

3. It is widely used in computer vision tasks such as image annotation, vehicle counting, activity recognition, face detection, face recognition, video object co-segmentation. It is also used in tracking objects, for example tracking a ball during a football match, tracking movement of a cricket bat, or tracking a person in a video.

WORKING METHODOLOGY:

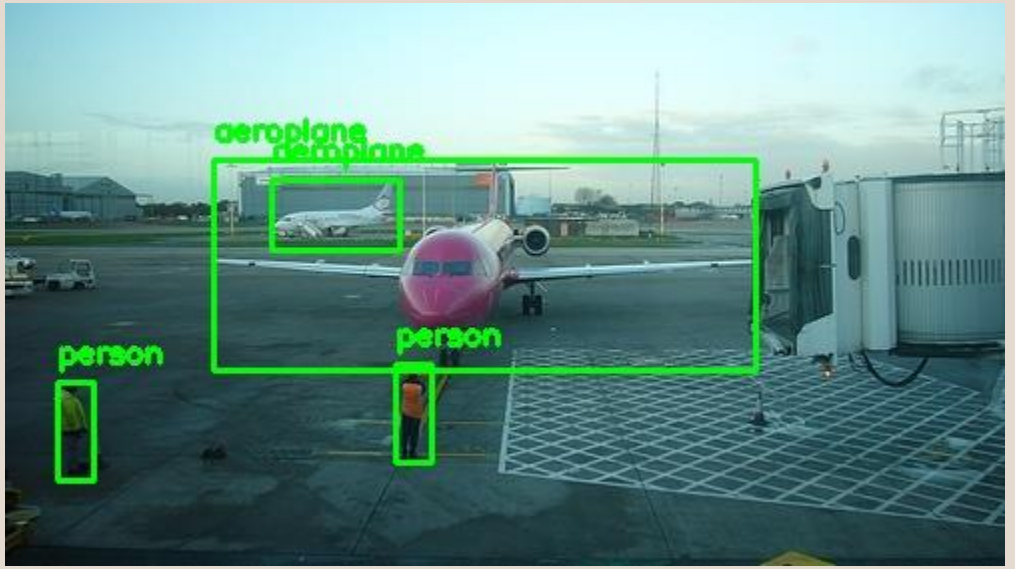
*

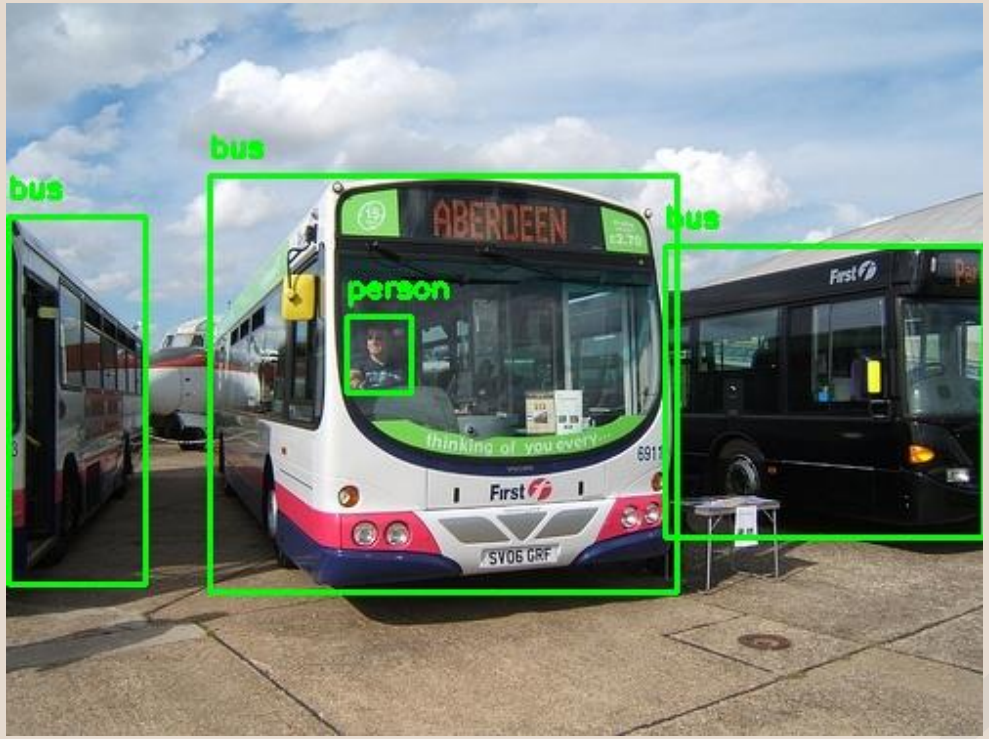


GOAL COMPOSITIONS:

- Sensors
- Object recognition model
- Translation model (Object labels translated in various languages)
- Google-Text-Speech

  (Pronunciation of Object names)

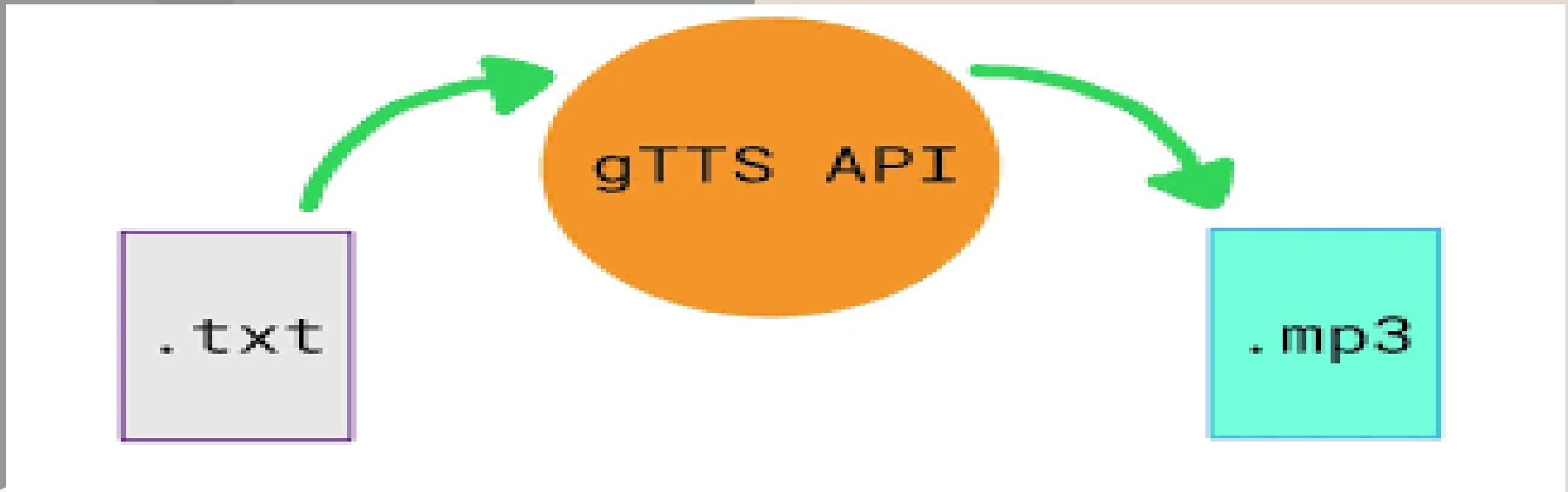




What is text to speech. Text to speech is also known as TTS, read aloud, or even speech synthesis. It simply means using artificial intelligence to read words aloud be; it from a PDF, email, docs, or any website.

gTTS (Google Text-to-Speech), a Python library and CLI tool to interface with Google Translate's text-to-speech API. Write spoken mp3 data to a file, a file-like object (bytestring) for further audio manipulation, or stdout .

http://gtts.readthedocs.org/



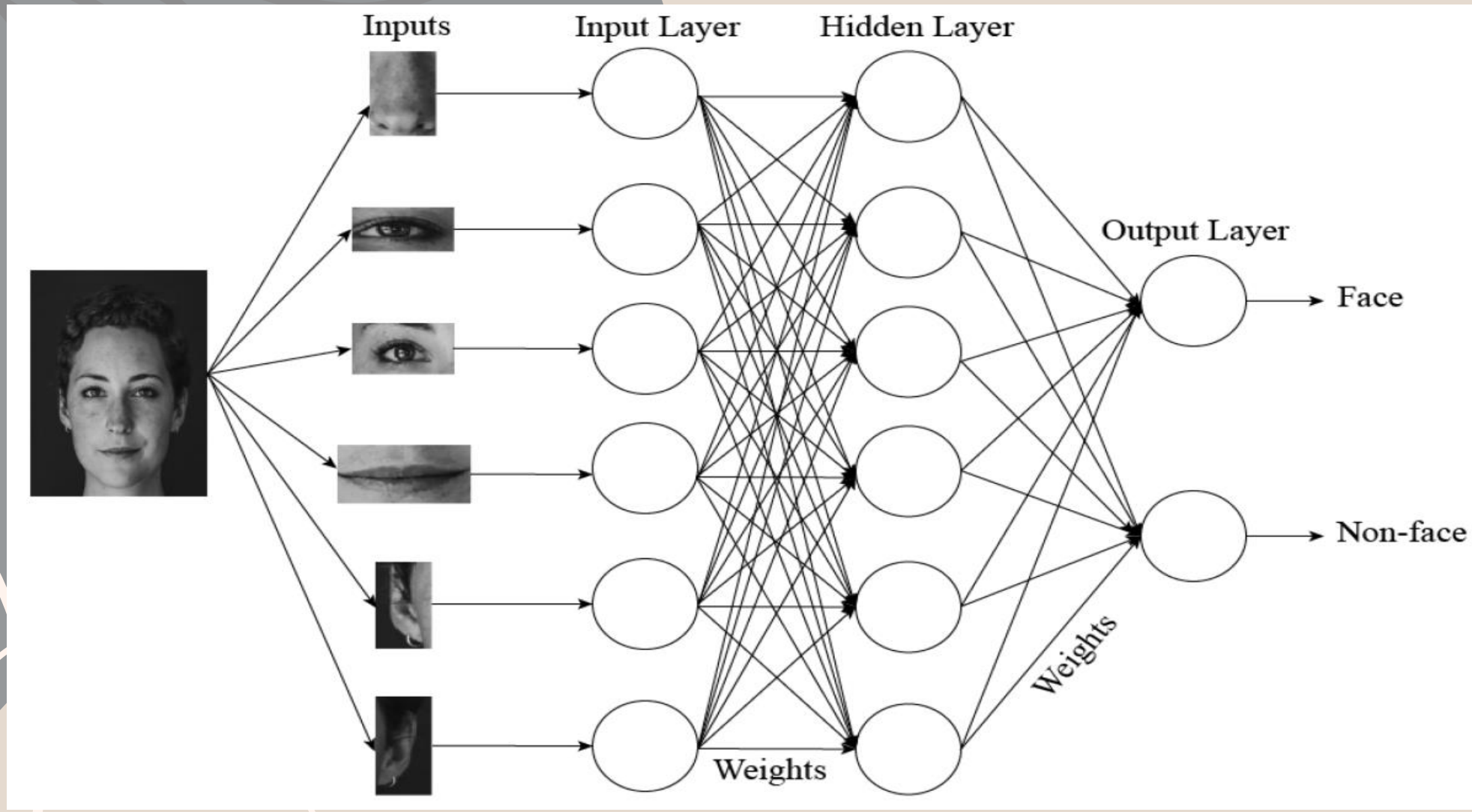
The Python code that generated the text-to-speech using gSpeak API library

```
IMPORT TORCH
FROM PIL IMPORT IMAGE
IMPORT OS
IMPORT TORCHVISION.TRANSFORMS AS TRANSFORMS

# IMPORT YOLOV5 DEPENDENCIES
FROM MODELS.EXPERIMENTAL IMPORT ATTEMPT_LOAD
FROM UTILS.GENERAL IMPORT NON_MAX_SUPPRESSION

# SET THE PATH TO THE YOLOV5 MODEL
MODEL_PATH = 'YOLOV5S.PT'  # SET THE PATH TO THE YOLOV5 MODEL WEIGHTS FILE

# LOAD YOLOV5 MODEL
DEVICE = TORCH.DEVICE('CUDA' IF TORCH.CUDA.IS_AVAILABLE() ELSE 'CPU')
MODEL = ATTEMPT_LOAD(MODEL_PATH)

# SET MODEL TO EVALUATION MODE
MODEL.EVAL()

# SET THRESHOLD FOR OBJECT DETECTION CONFIDENCE
```

...continuation of the code

```
# IMAGE TRANSFORMATIONS
TRANSFORM = TRANSFORMS.TOTENSOR()

# ITERATE OVER IMAGES IN THE DIRECTORY
FOR IMAGE_FILE IN OS.LISTDIR(IMAGE_DIR):
    IF IMAGE_FILE.ENDSWITH('.JPG') OR IMAGE_FILE.ENDSWITH('.PNG'):
        IMAGE_PATH = OS.PATH.JOIN(IMAGE_DIR, IMAGE_FILE)

        # LOAD IMAGE
        IMG = IMAGE.OPEN(IMAGE_PATH)
        IMG_TENSOR = TRANSFORM(IMG).UNSQUEEZE(0).TO(DEVICE)

        # PERFORM INFERENCE
        PRED = MODEL(IMG_TENSOR)[0]

        # APPLY NMS
        PRED = NON_MAX_SUPPRESSION(PRED, CONF_THRESHOLD, 0.5)

        # DRAW BOUNDING BOXES AND LABELS ON IMAGE
        FOR DET IN PRED:
            IF DET IS NOT NONE AND LEN(DET):
                FOR *XYXY, CONF, CLS IN DET:
```

The most common type of Deep Learning algorithm used for object recognition is a deep learning Convolutional Neural Network (CNN). CNNs are a type of artificial neural network that are well-suited for object detection tasks.



LIMITATIONS

- Sensor Accuracy and Reliability
- Complexity of Object Recognition
- Language Translation Accuracy (Accent)
- Speech Synthesis Quality
- Computational Resources and Performance

FUTURE SCOPE

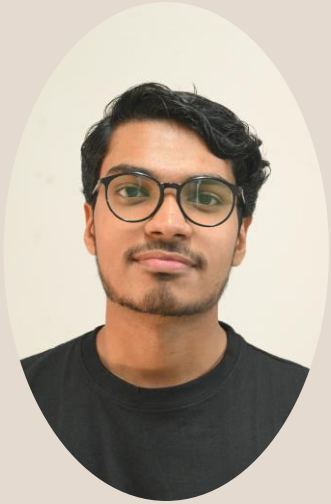
The future scope of the walking stick system involves integrating advanced object recognition algorithms and multi-sensory capabilities to enhance environmental perception. Further advancements include expanding language support, integrating augmented reality features, and fostering collaboration for continuous innovation in assistive technologies.

CONTRIBUTORS:

NDUBUISI GODCARES .C.

SYMBIOSIS INSTITUTE OF TECHNOLOGY

(AFFILIATION OF SYMBIOSIS INTERNATIONAL UNIVERSITY)



NAJEEB SAIYEED

SYMBIOSIS INSTITUTE OF TECHNOLOGY

(AFFILIATION OF SYMBIOSIS INTERNATIONAL UNIVERSITY)

Conclusion:

" MyGuide " aims to empower visually impaired individuals by providing them with a reliable, real-time object detection walking guide. This innovative solution leverages cutting-edge technology to enhance independence, safety, and accessibility in daily life, ultimately contributing to a more inclusive society.


Thank you