**GENERAL STRUCTURE OF THE PROPOSED MOBILE APPLICATION.**

This application will have three modules that are catering for specific groups of people in society namely;

* The partially blind
* The completely blind
* The helpers for the VIPs

**FUNCTIONALITIES FOR EACH OF THE GROUPS**

|  |  |  |
| --- | --- | --- |
| **PARTIALLY BLIND** | **COMPLETELY BLIND** | **THE HELPERS FOR THE VIPS** |
| This group of people will have both functionalities that suit the completely blind and the partially blind which include   1. Text recognition – ranges from reading articles, books, and other items that have text on them 2. Voice Assistant – this will help in the fast execution of instructions like sending messages and navigation through the app. 3. Object recognition features 4. object description features – helps in describing the surrounding of a person. 5. Speech-to-text features. These will help in speeding up interaction with the app. | 1. Text recognition – ranges from reading articles, books, and other items that have text on them 2. Voice Assistant – this will help in the fast execution of instructions like sending messages and navigation through the app. 3. Object recognition features 4. object description features – helps in describing the surrounding of a person. This uses captions for an area that are read out as speech feedback 5. Speech-to-text features. These will help in speeding up interaction with the app in the areas of sending messages plus accepting feedback during an interaction. 6. Speed dial features. – this involves contact for emergencies and sending messages to the contacts saved in the app 7. Wireless pairing functionalities to help them use other devices like Bluetooth to headphones and cameras to extended displays 8. Route navigation in the surrounding -helps the blind person know how if traffic is in the area nearing their location. | 1. User location features- these are used to see where exactly the user is and what is around them. 2. Directions and chat features 3. Monitoring movement features 4. Recording surrounding of user 5. Voice feedback texts. |

**HOW THE PROPOSED AP WILL WORK**

This app has three modules as listed above. It will have several technologies that will be used to come up s with a desired framework for the VIPs.

These technologies include;

**TensorFlow models for ML (machine learning)** - TensorFlow is an open-source library developed by Google primarily for deep learning applications. It also supports traditional machine learning. Using TensorFlow we cam develop our own models that will be used to achieve desired results in aspect of the following algorithms.

**Text To Speech algorithms** - Text-to-speech (TTS) is a type of assistive technology that reads digital text aloud. It's sometimes called “read aloud” technology. With a click of a button or the touch of a finger, TTS can take words on a computer or other digital device and convert them into audio.

The ML algorithm establishes the connection between phonemes and sounds, giving them accurate intonations. The system uses a sound wave generator to create a vocal sound. The frequency characteristics of phrases obtained from the acoustic model are eventually loaded into the sound wave generator.

Object detection algorithms - 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

Speech to text Algorithms- The voice commands are translated into real words that are later processed to desired destination.

Image caption generation algorithm – this algorithm helps us describe the surrounding inform to a camera to show the user or give them a clue on what is going on.

Face detection - Face-detection algorithms focus on the detection of frontal human faces. It is analogous to image detection in which the image of a person is matched bit by bit. Image matches with the image stores in database. Any facial feature changes in the database will invalidate the matching process. This works for people at home. Once the camera captures the human face, it will be matched the person in the database to identify who is in the surrounding

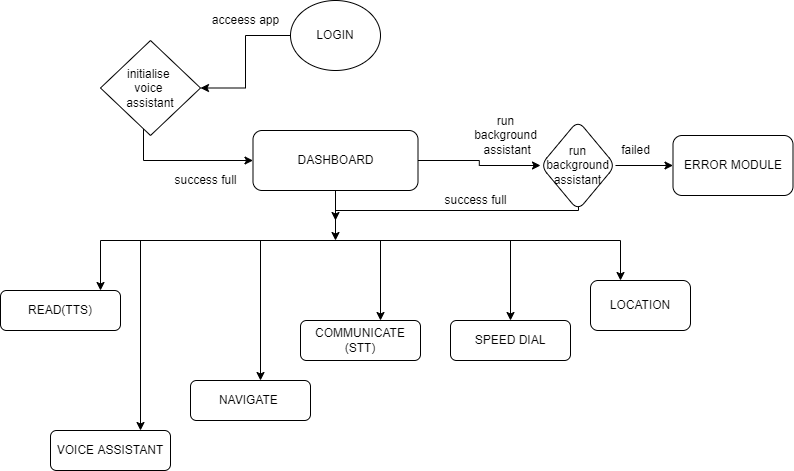
Google maps - these apis help us calculate the distance and the location of the VIPs. This eases routing during navigation.

Other technologies include

1. Machine Learning sdk kits for flutter android development
2. CNN algorithms
3. Geolocation and mapping APIs
4. Firebase platforms

DATA FLOW DIAGRAMS FOR THE APPLICATION

**APPLICATION DASHBOARD**

****

**EXPLANATION**

Once the user has been logged in to the app they will gain access to the dashboard. In the background a voice assistant will be initialized to help the user navigate through the app.

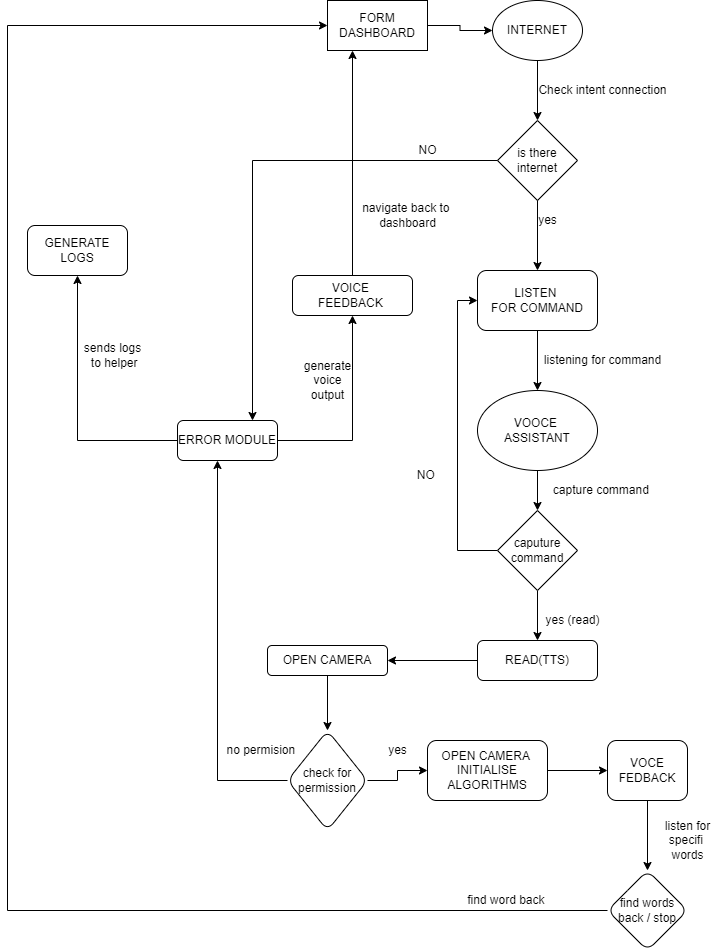
The voice assistant will listen for major or key words like **start , end , call , read, navigate , location, communicate, dial** words in order to perform a corresponding action for the user. Also in addition while an option for voice assistant will be available to help the user navigate through the app and perform functionalities. In case of any errors during the running of the app , they will be rendered as voice feedback from the error module.

The dashboard has tiles with labels which have corresponding identifiers that will be used during the instruction execution.

**OPERATION**

After the user is logged in, a screen reader and a voice assistant will help the user select desired functionality by either listening to the screen reader or the voice assistant.

**ON SELECTION OF: READ (TEXT TO SPEECH )**

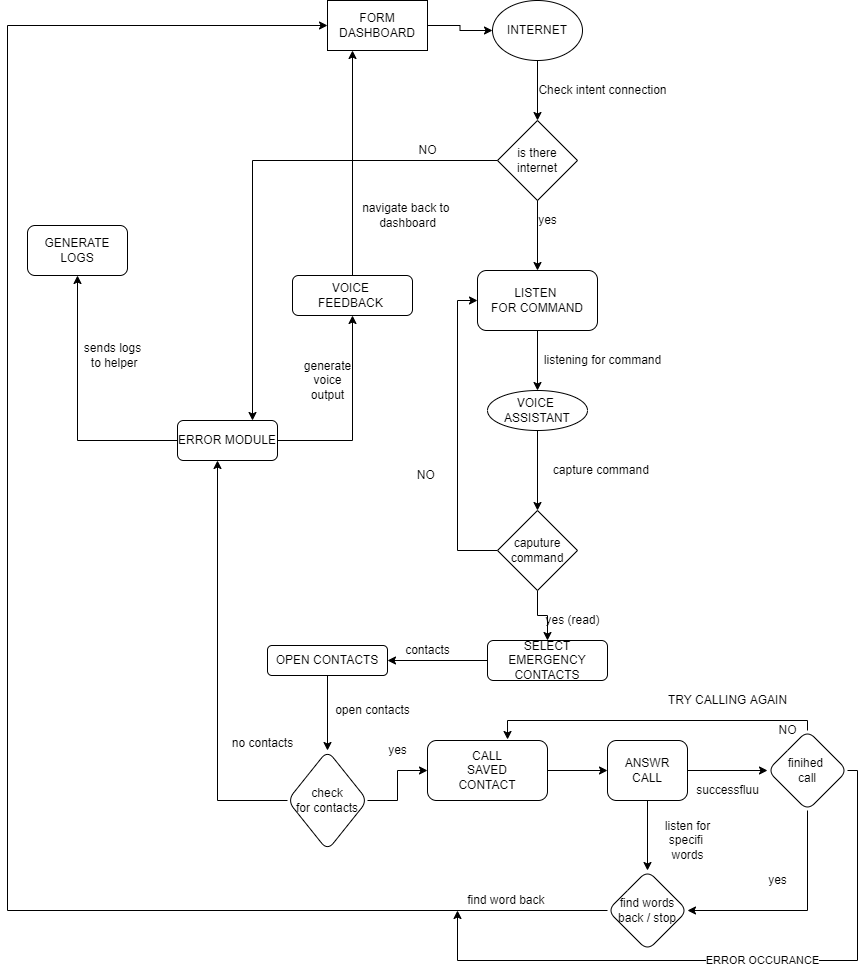
****

**To** select the **READ functionality**, a user will either say read or once u select read, from the dashboard the ap will check if an internet connection is present, if true the voice assistant will keep listening In the background for key works like stop end start according to the stage of instruction execution.

The camera will be opened and the Text To Speech algorithms will be loaded in the background. Once camera permissions are granted to the app the camera will open and load the algorithms to detect text in what the user is trying to view. Text-to-speech (TTS) is a type of assistive technology that reads digital text aloud. It's sometimes called “read-aloud” technology. With a click of a button or the touch of a finger, TTS can take words on a computer or other digital device and convert them into audio.

These words are then rendered as feedback to the user in audio form. While this is happening the voice assistant is listening and waiting for keywords. Once they are detected, the corresponding action will be performed back to the dashboard for a new instruction. In case of any errors, they are tracked back to the error module and rendered as error logs back to the VIP helper for review.

**ON SELECTION OF: EMERGENCY (SPEED DIAL)**

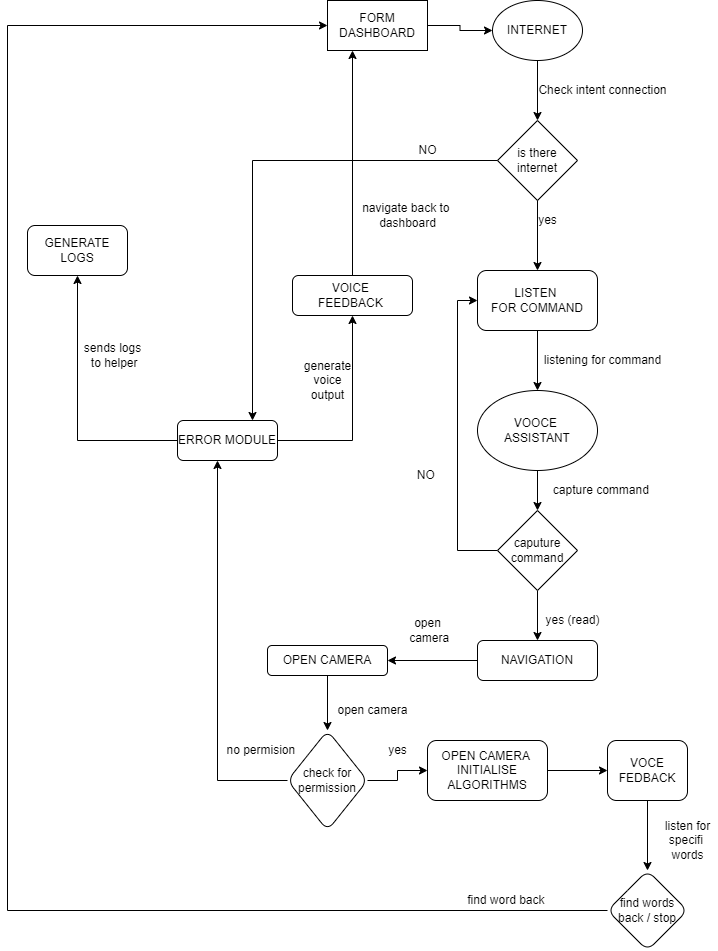
****

**To** select the **EMERGENCY / SPEED DIAL functionality**, a user will either say speed dial, or once u select emergency, from the dashboard the app will check if an internet connection is present, if true the voice assistant will keep listening In the background for key works like stop end start according to the stage of instruction execution.

The app will open the speed dial contact list and read out which contact the user would like to call, once the contact is selected, the cal will begin to ring. In case an error occurs it's tracked and rendered as feedback to the user in audio form. While this is happening the voice assistant is listening and waiting for keywords. Once they are detected, the corresponding action will be performed back to the dashboard for a new instruction they are also tracked back to the error module and rendered as error logs back to the VIP helper for review.

In case of a successful call, the calls will be recorded and later used for other purposes while the voice assistant waits and listens for keywords to perform specific actions on the dashboard

**ON SELECTION OF: NAVIGATION**

****

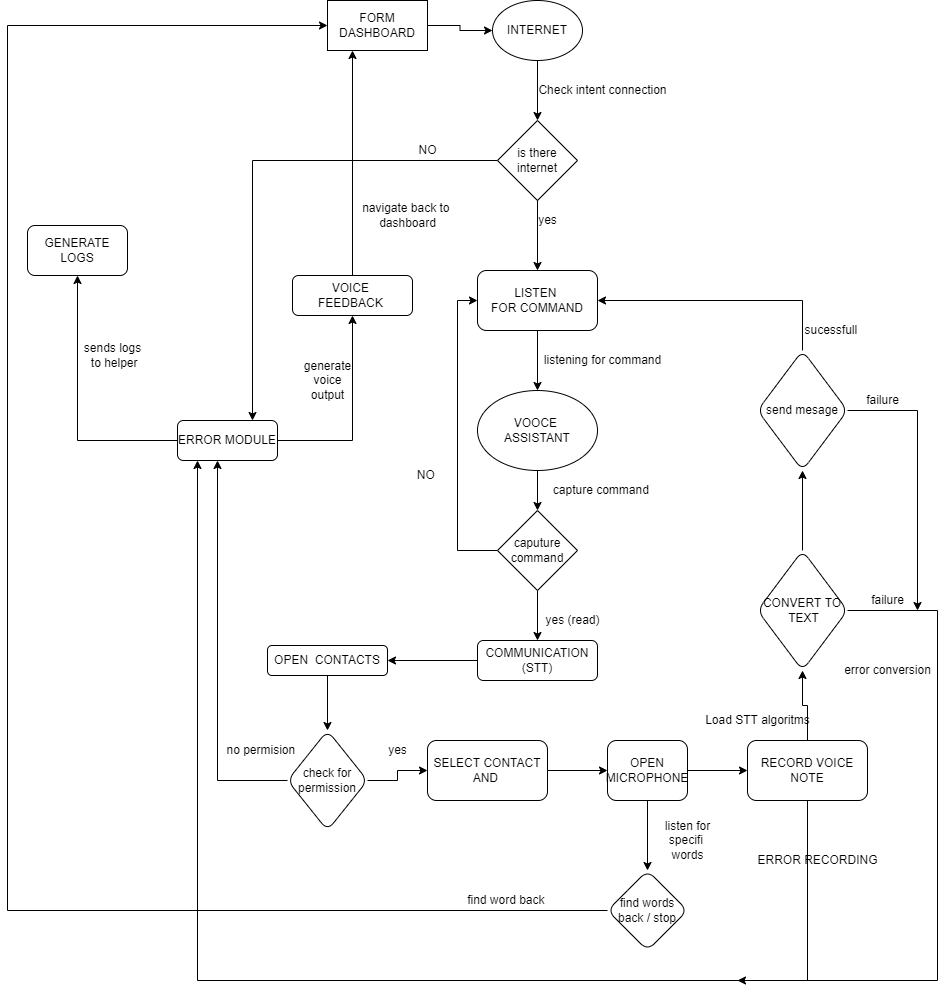
**To** select the **NAVIGATION functionality**, a user will either say navigate or once u select navigate, from the dashboard the app will check if an internet connection is present, if true the voice assistant will keep listening In the background for key works like stop end start according to the stage of instruction execution.

The camera will be opened and the object detection algorithms like CNN, YOLO, and other algorithms will be loaded in the background. Once camera permissions are granted to the app the camera will open and load the algorithms to detect objects and give descriptions according to the surrounding of the user's view.

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. These results will be rendered as voice feedback to the user. During the object detection, google maps will automatically generate a map view of the user's location and give insights into where the user is and where they would like to go. In case they select a destination location, route suggestions will be given in audio form according to the best route to take and which route has fewer motor vehicles

These words are then rendered as feedback to the user in audio form. While this is happening the voice assistant is listening and waiting for keywords. Once they are detected, the corresponding action will be performed back to the dashboard for a new instruction. In case of any errors, they are tracked back to the error module and rendered as error logs back to the VIP helper for review.

**ON SELECTION OF: COMMUNICATION**

****

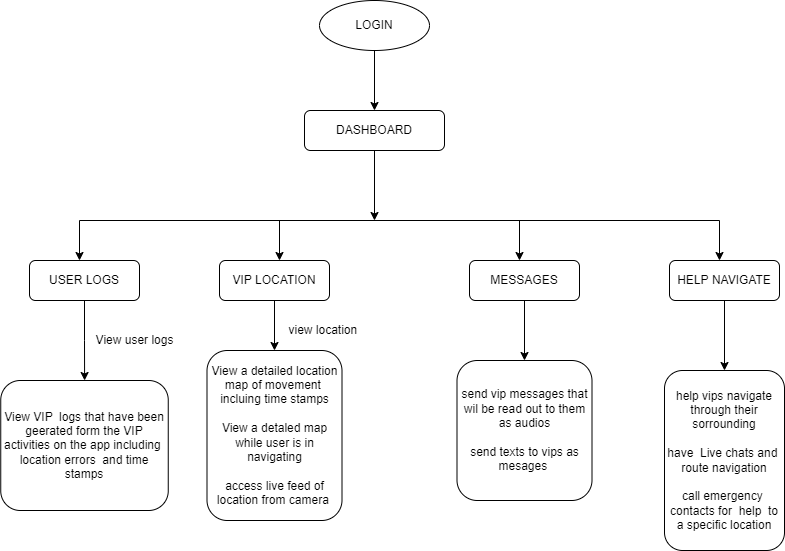
**To** select the **COMMUNICATION functionality**, a user will either say communicate, or once u select communicate, from the dashboard the app will check if an internet connection is present, if true the voice assistant will keep listening In the background for key works like stop end start according to the stage of instruction execution.

The app will open the contact dial contact list and read out which contact the user would like to send a message to, once the contact is selected, the microphone will open and ask for permission to record audio. In case an error occurs it's tracked and rendered as feedback to the user in audio form. While this is happening the voice assistant is listening and waiting for keywords. Once they are detected, the corresponding action will be performed back to the dashboard for a new instruction they are also tracked back to the error module and rendered as error logs back to the VIP helper for review.

In case of a successful message capture, the message will be recorded and later used for other purposes while the voice assistant waits and listens for keywords to perform specific actions on the dashboard. The voice commands are translated into real words that are later processed to the desired destination.

These results will be rendered as text feedback to the desired user in message form. During the process, google maps will automatically generate a map view of the user's location and give insights into where the user is and this data is attached to the message being sent. While this is happening the voice assistant is listening and waiting for keywords. Once they are detected, the corresponding action will be performed back to the dashboard for a new instruction. In case of any errors, they are tracked back to the error module and rendered as error logs back to the VIP helper for review.

**The helper module**

****