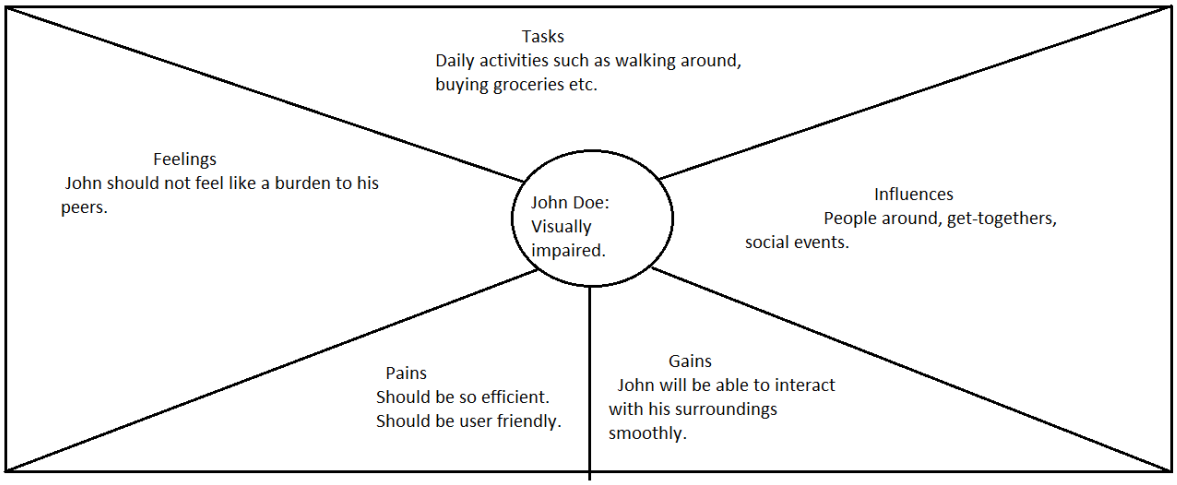
IMAGE CAPTIONING WITH VOICE

USE-CASE:

To help visually impaired people better understand their surroundings. People can use their mobile phones and capture photographs using which captions can be generated that can be read out loud to the visually impaired, so that they can get a better sense of the surroundings around them. Image caption generation can also make the web more accessible to visually impaired people.

EMPATHY-MAP:



PERSONA:

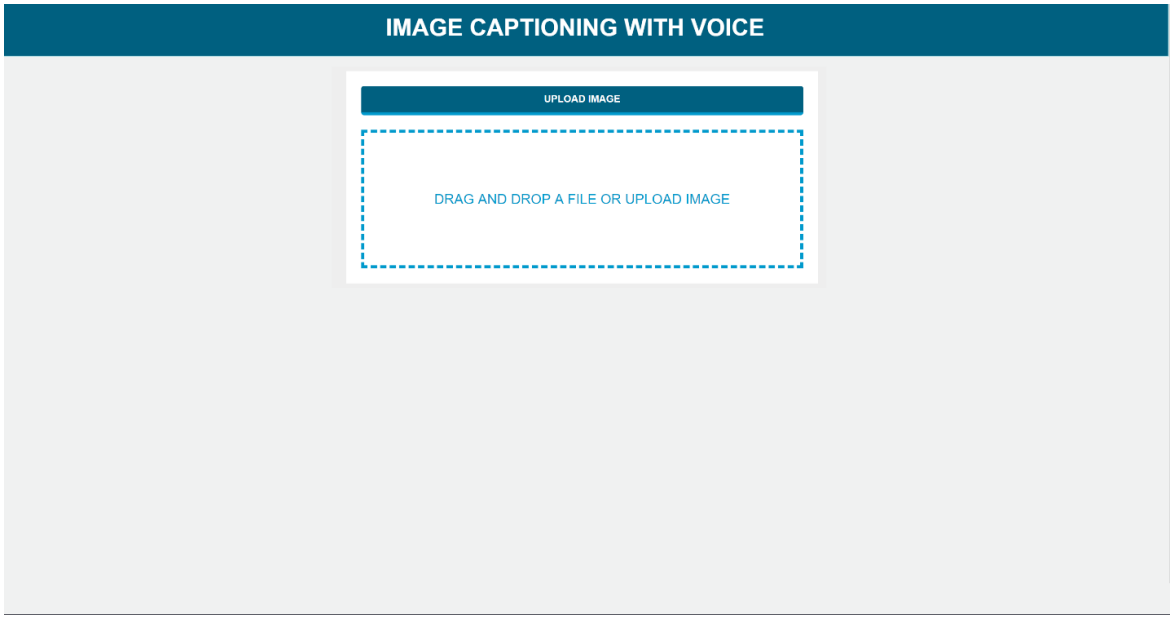
The output of the project will help the visually impaired (partially impaired as well).

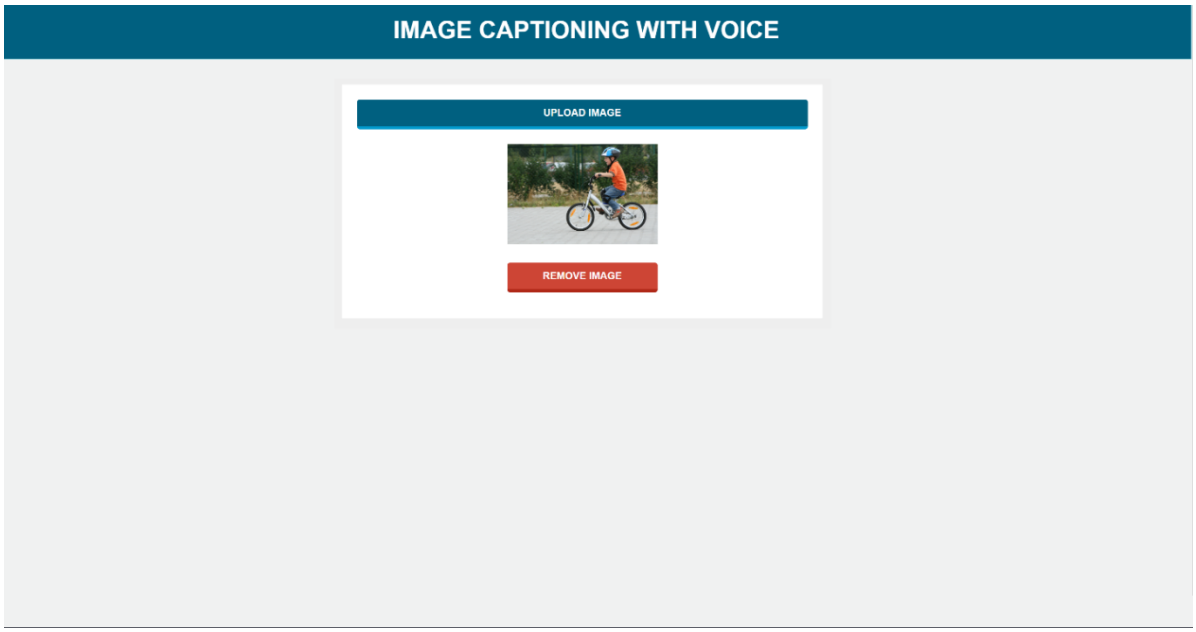
The following is the type of Engaging Personas through which we show the application of our project.

John Doe (A guy who loves his neighbourhood):

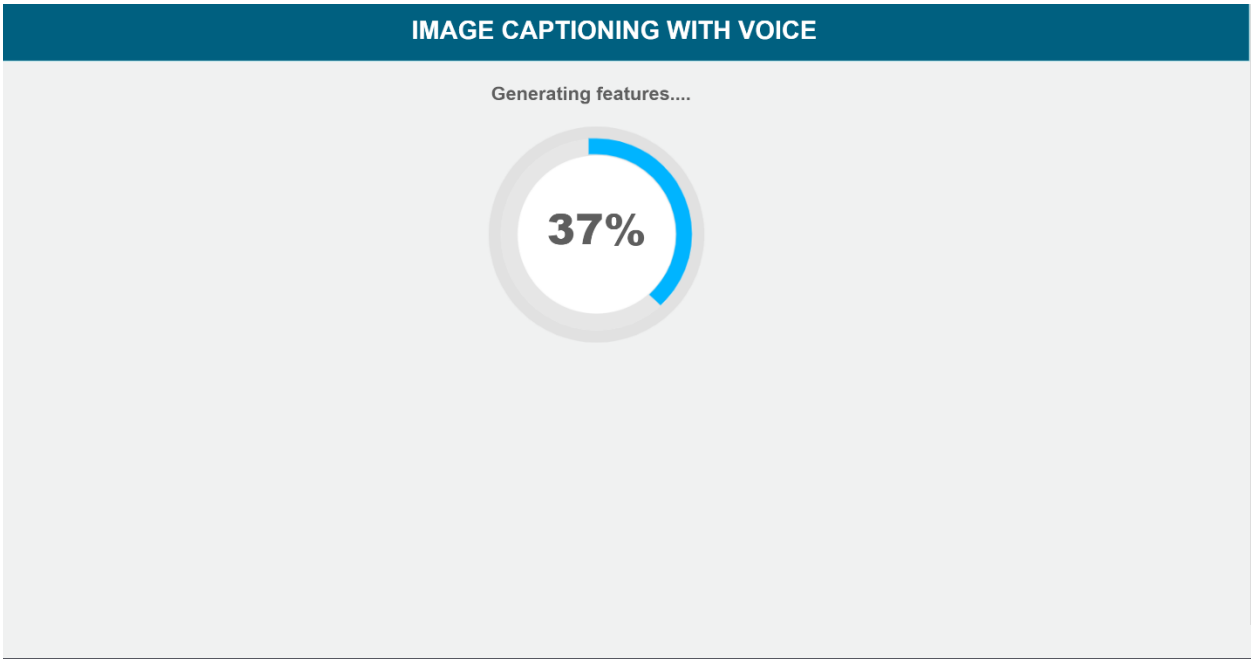
John loves spending his time in his neighbourhood but however,due to his limited visual aid he finds it difficult to walk in his neighbourhood. He becomes too cautious that he might walk into someone or something. However, it would be better if he can somehow know what is in front of him or around him.Thus, the audio output of our software will let john know about the things around him.

Here, The image is being given as an input to the model

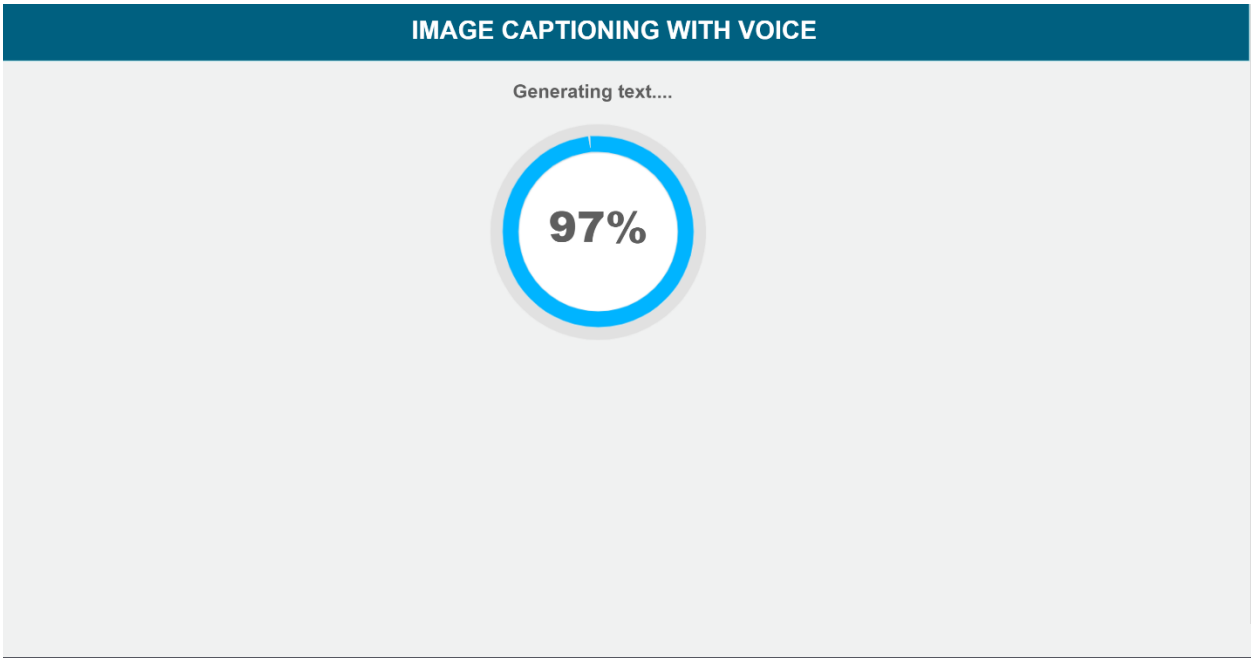




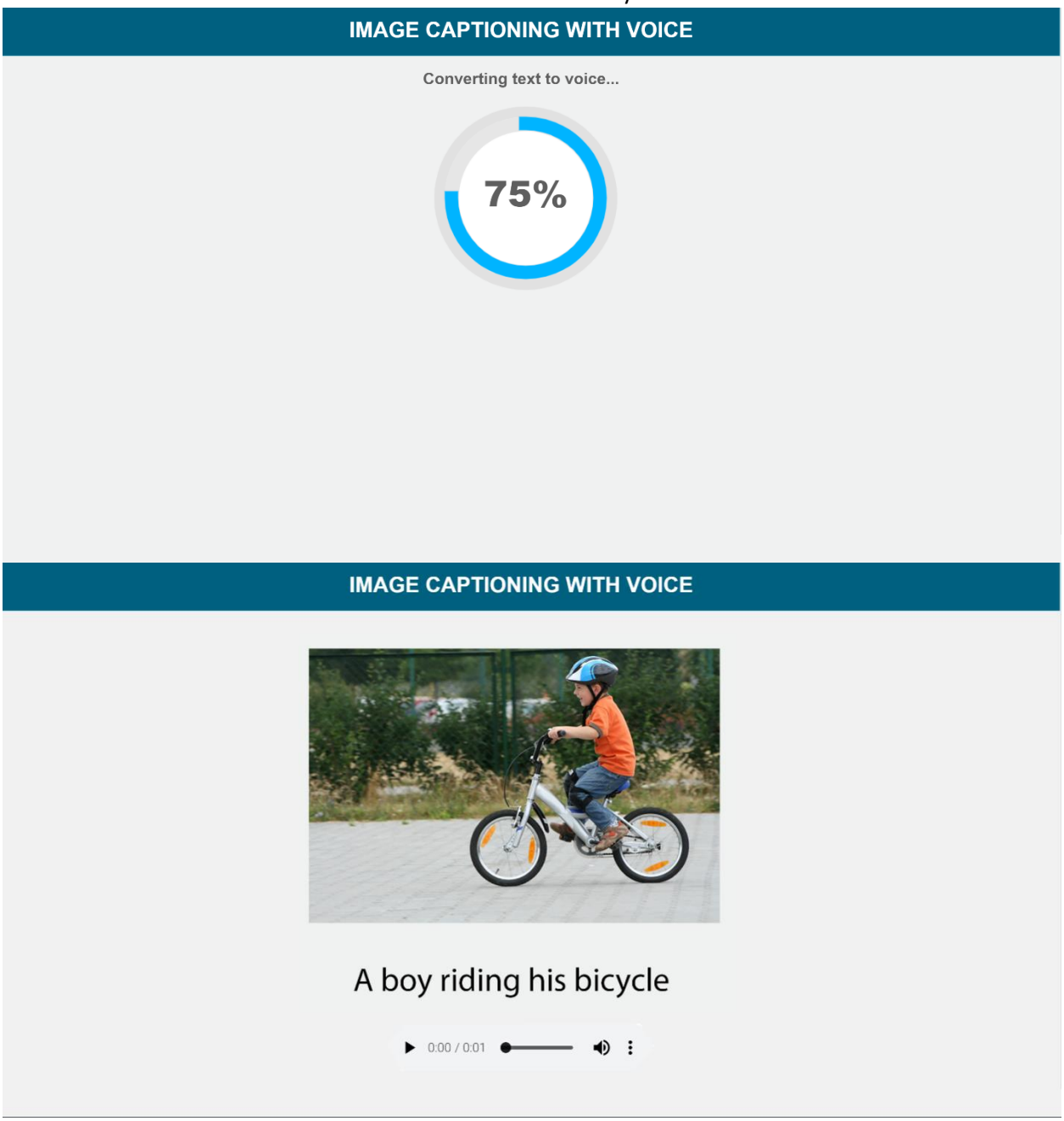
The image is then processed, objects in the image are recognized and features are generated.



The features generated are then converted into meaningful sentences.



The text is then converted to voice. Voice assistant will convey this to the user.



Future scope:

This project can be extended to the video version and can be integrated with Google glasses and Bluetooth audio device. With this we can generate live audio streaming description of the scenarios happening around.