## **DOCUMENTATION**

## **Voice Vision**

Introduction:

The introduction discusses the evolution of technology towards virtual assistants as a solution for simplifying tasks in the busy lifestyle of blinds, emphasizing the role of data from the web in building intelligent virtual assistants interconnected with the mobile camera.

Study objectives:

Develop a virtual assistant for Windows users, utilize machine learning, deep learning, computer vision and natural language processing concepts, increase the accuracy of speech to text software, and provide a smart working experience for desktop users over the web. CV is majorly used for the identification of any objects.

Main findings:

The main findings are related to the efficiency and effectiveness of the virtual assistant in improving user experience and productivity through natural language processing and artificial intelligence. It is connected with Computer vision for the use of recognition.

Methodology:

Our methodology leverages advanced technologies including speech recognition, neural networks, natural language processing (NLP), and artificial intelligence (AI) to enhance the accuracy of speech-to-text software. This approach encompasses various steps of NLP and inter-process communication between the operating system and ongoing processes. Additionally, we utilize Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and computer vision (CV) for image and video recognition.

Summary:

The paper discusses the development of a voice assistant for Windows users, emphasizing the convenience and efficiency of performing tasks through voice commands and accessing information on the internet, with a focus on enhancing the accuracy of speech to text software in addition to it, there is a connection in camera to find out the object for detection and navigation.

Outcome measured:

Accuracy of speech to text software and voice recognition capabilities using pattern recognition techniques, based on artificial intelligence, natural language processing, and deep learning. Adding on, it enables the user to navigate to places with the help of voice assistant where input is obtained through camera.