**AI202 – Trends and Techniques in Artificial Intelligence**

**Project Proposal- Abeera Najam 2022038**

**Project Title:** **Image Captioning System to Assist the Blind**

To assist the visually impaired, the area of assistive technologies has seen rapid growth in the past few years. This project aims to develop op a system that helps blind people cheaply get answers in their everyday lives with the help of state-of-the-art deep learning techniques such as image captioning, image recognition, and text generation.

**Dataset:** This project uses a publicly available dataset called VizWiz. The dataset was constructed from a natural visual question-answering setting where blind people each took a photo and recorded a spoken question about it, together with 10 crowdsourced answers per

Photo. The dataset includes 23,431 training images, 117,155 training captions, 7,750 validation images, 38,750 validation captions, 8,000 test images, and 40,000 test

captions. The annotations are written inside of a JSON file.

**Link to Dataset:** <https://vizwiz.org/tasks-and-datasets/vqa/>

(Images: [training](https://vizwiz.cs.colorado.edu/VizWiz_final/images/train.zip), [validation](https://vizwiz.cs.colorado.edu/VizWiz_final/images/val.zip), and [test](https://vizwiz.cs.colorado.edu/VizWiz_final/images/test.zip) sets)

**Real-Life Problem and Application:**

The real-life application of this project empowers visually impaired individuals to gain access to visual information in their environment, enhancing their independence and ability to navigate and interact with the world around them. By developing an image captioning system integrated into assistive technologies, this project seeks to empower visually impaired individuals to gain access to visual information without depending solely on others.

**Deep Learning Model:**

The project will incorporate computer vision and natural language processing techniques. It will develop a system using a CNN model along with an NLP model. These models will be concatenated to create a single image captioning system that takes image features as input and produces a text sequence describing the image. By implementing this project, it will aim to bridge the accessibility gap for visually impaired individuals and empower them to engage with visual information in a meaningful and independent manner.