Vision.ai – Design Process & Requirements Analysis

An Assistive Al System for Visually Impaired Users

Introduction

 Vision.ai is an Al-powered assistive system that helps visually impaired users understand their surroundings by generating image captions and reading them aloud.

Objective

- Provide real-time image captioning
- Convert captions into speech
- Enhance accessibility using AI and NLP
- Optional translation for multilingual support

Requirements Analysis

Text

• Functional Requirements:

- Upload and process images
- Extract features using CNN
- Generate captions using Transformer
- NLP refinement of captions
- Text-to-Speech conversion
- Optional translation of captions

• Non-Functional Requirements:

- High accuracy and speed
- User-friendly interface
- Cross-platform compatibility

Design Process Overview

- Step 1: Image Upload
- Step 2: Feature Extraction using CNN
- Step 3: Caption Generation using Transformer
- Step 4: NLP Refinement
- Step 5: Optional Translation
- Step 6: Text-to-Speech Conversion
- Step 7: Audio Output to User

Use Case Diagram

```
<<Actor>>
                      Visually
                      Impaired User
Upload Image | | Generate Caption| | Convert to | | Translate Caption|
                 (CNN + Transformer)| Speech Output | | (Optional Module)|
```

Component Diagram

```
User Interface
     (Streamlit Frontend)|
  Captioning System
FeatureExtractor (CNN)
- CaptionGenerator (Transformer) |
- NLP Correction Module
 Optional Translator Module
  TTS Module (pyttsx3)
```

Deployment Diagram

```
User Device (Laptop/Mobile)
 - Streamlit Interface
 - Audio Output
  Application Container
 (Docker Container)
- CNN Model
- Transformer Model
- NLP Engine
- TTS Engine (pyttsx3)
- Optional Translator API
  Cloud Deployment / GitHub |
 (Heroku / Render / AWS)
```

Sequence Diagram

```
Visually impaired | System | CNN+ Transformer | NLP Engine | Translator | TTS Engine
            I upload image
                      1 -- Extract Features --
                      1 -- Generate caption -----)
                      1 = ---- Raw Caption --- [
                      1 - Refine Caption ----
                      1 - Translate Caption (if opted) ---->1
                     | -- Convert Text to Speech -- ---
                     12 ..... Audio output ----
```

Class Diagram

Modular Architecture

- ImageUploader Module
- FeatureExtractor Module
- CaptionGenerator (CNN + Transformer)
- NLPEngine for language refinement
- TTSModule for speech output
- Translator (Optional Module)

Deployment Strategy

- - Frontend: Streamlit Interface
- Backend: Python modules (Modular structure)
- Model Hosting: Locally or on cloud
- Containerization: Docker for easy deployment
- Platform: Deploy on Web, Raspberry Pi, or desktop

Summary

- Vision.ai combines Computer Vision, NLP, and Speech Synthesis
- Designed with modular, scalable, and accessible architecture
- Aims to empower visually impaired users for independent interaction