Vision.ai – Design Process & Requirements Analysis

An Assistive AI System for Visually Impaired Users

Introduction

- Vision.ai is an AI-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

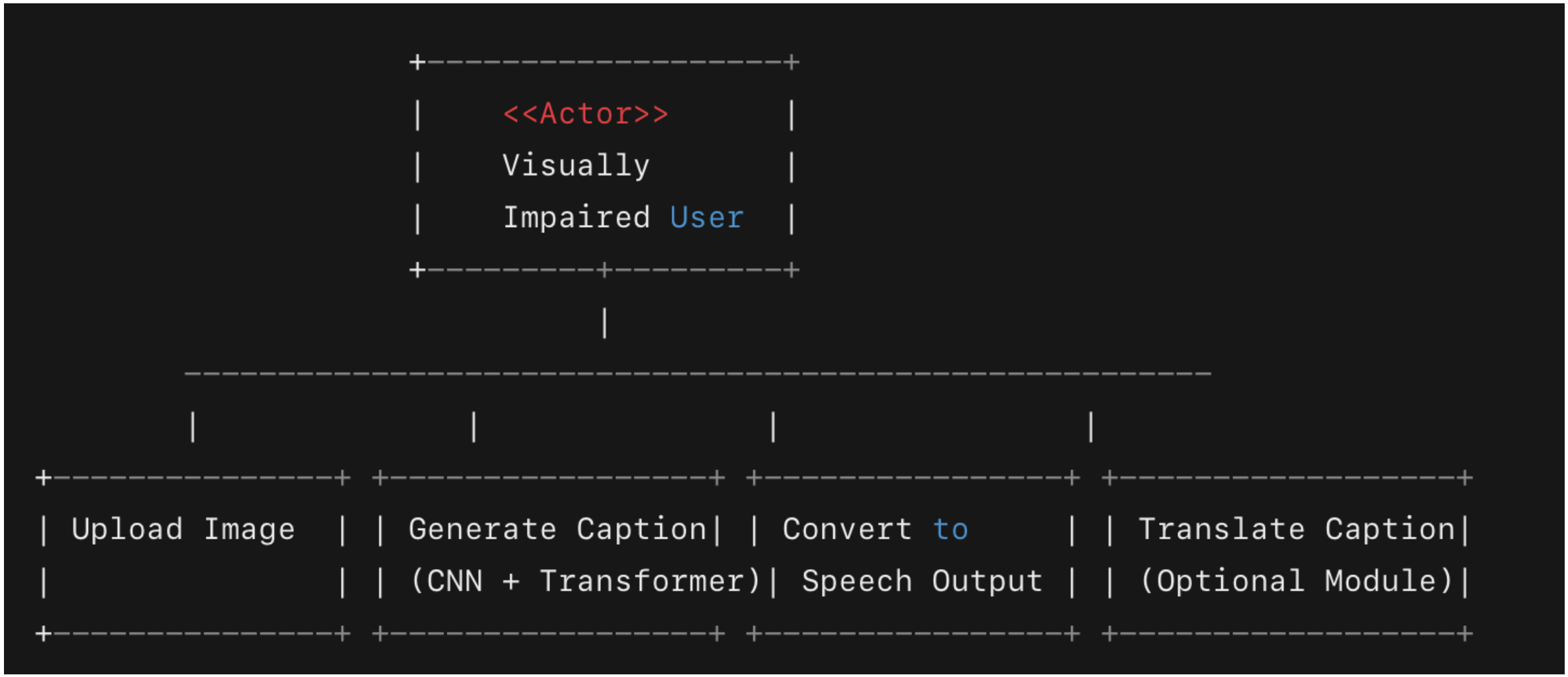
Text

- *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

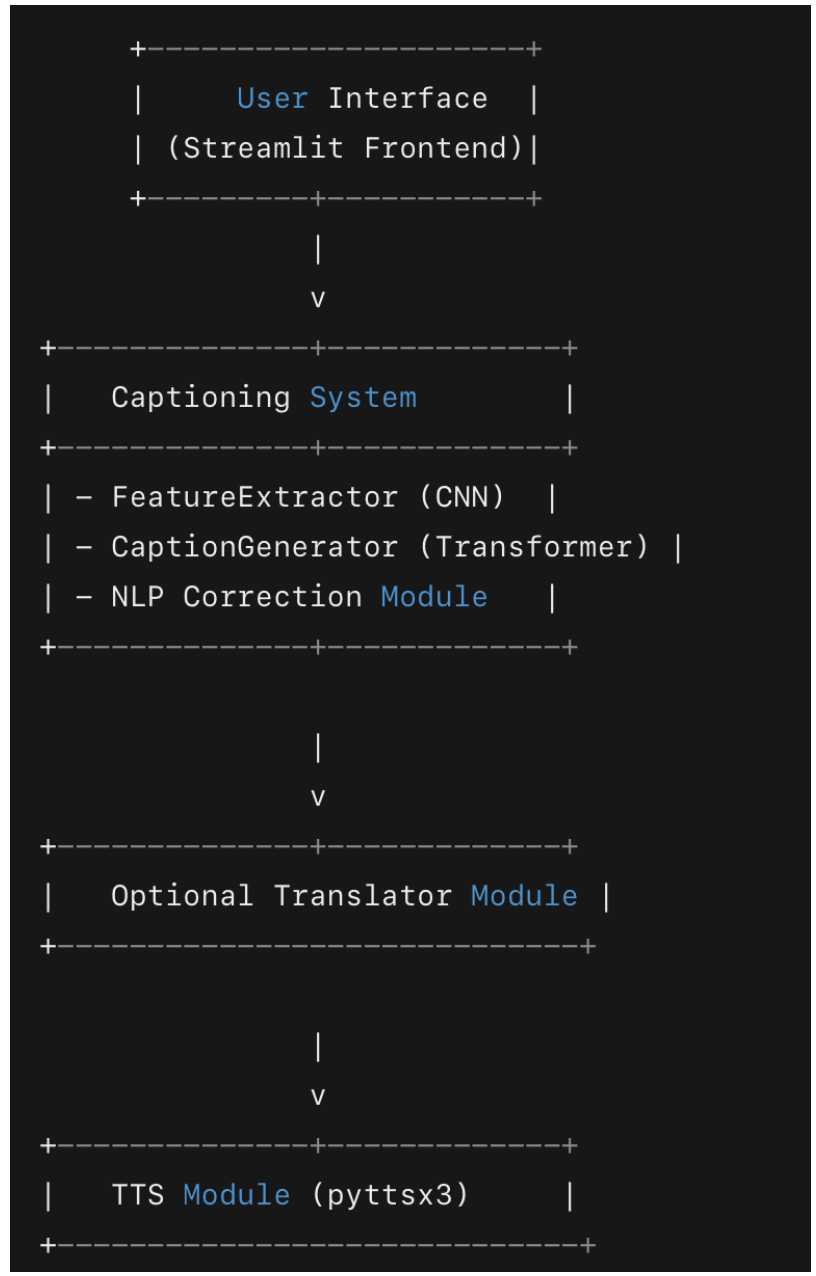
Design Process Overview

- - 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

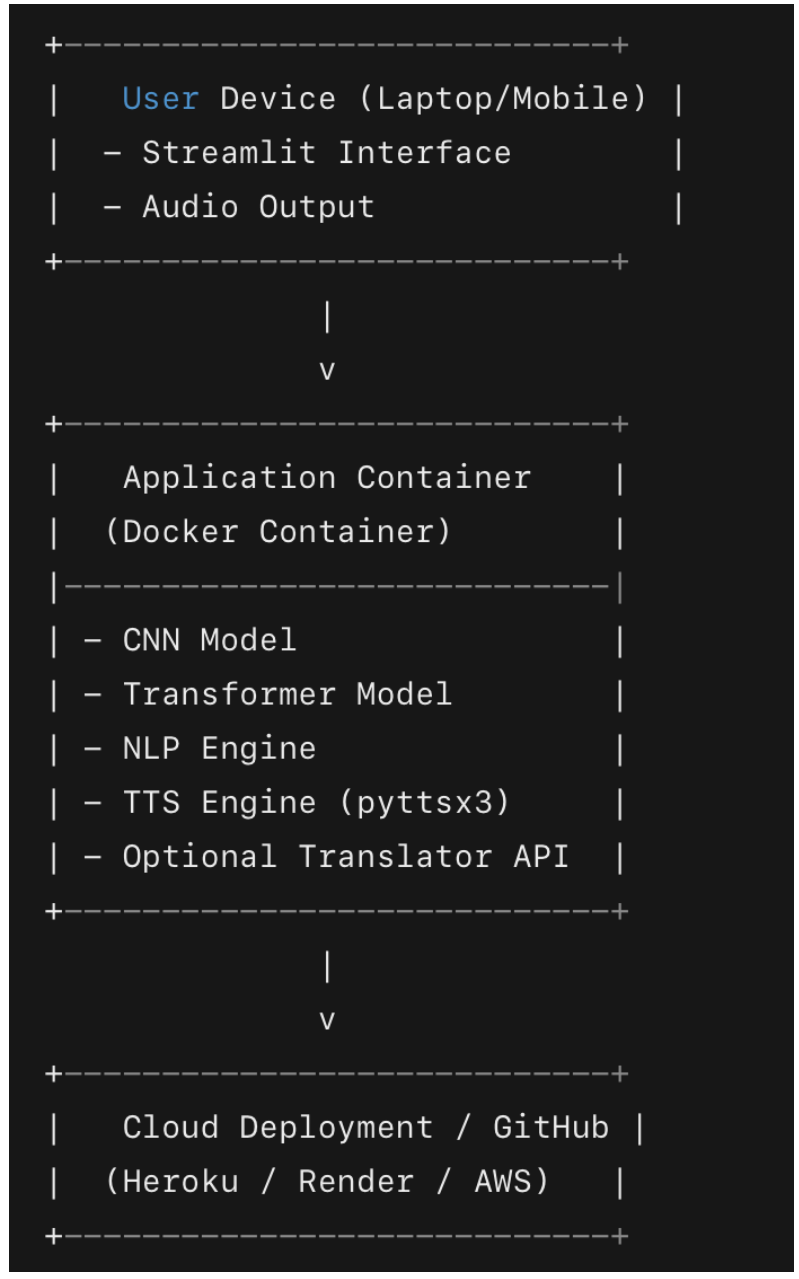
Use Case Diagram



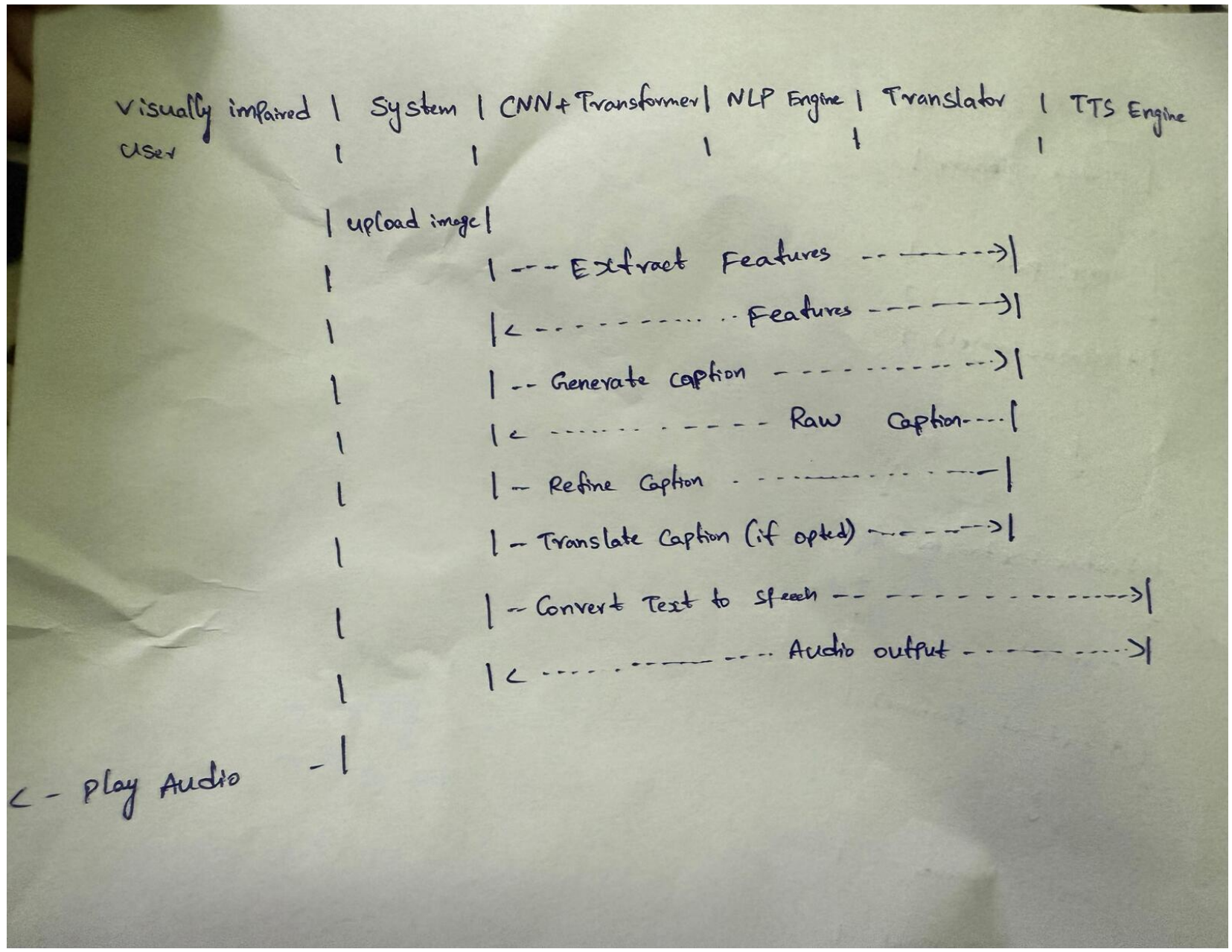
Component Diagram



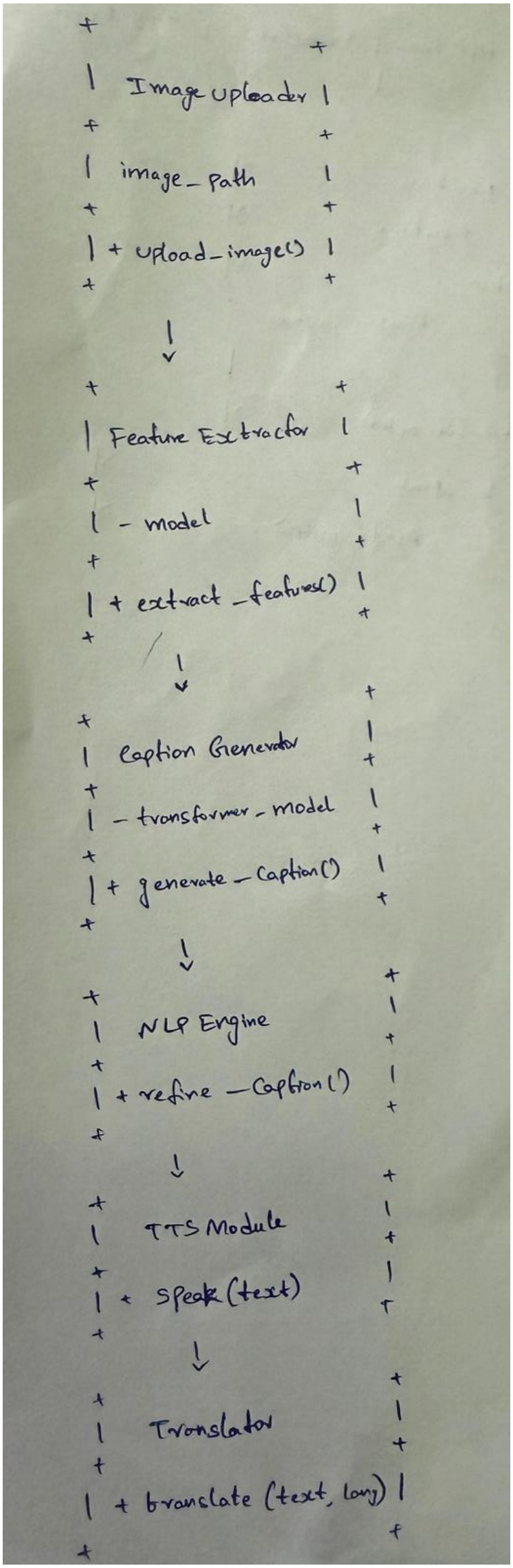
Deployment Diagram



Sequence Diagram



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