AI Powered Document Summariser

Design Document

October 9, 2023
Version Major 1.0

Anvit Gupta

22114009

anvit_g@cs.iitr.ac.in

Utkarsh Lohiya

22114103

utkarsh_l@cs.iitr.ac.in

Krishna Aggarwal

22114048

krishna_a@cs.iitr.ac.in

Sarvasva Gupta

22114086

sarvasva_g@cs.iitr.ac.in

Vineet Kumar
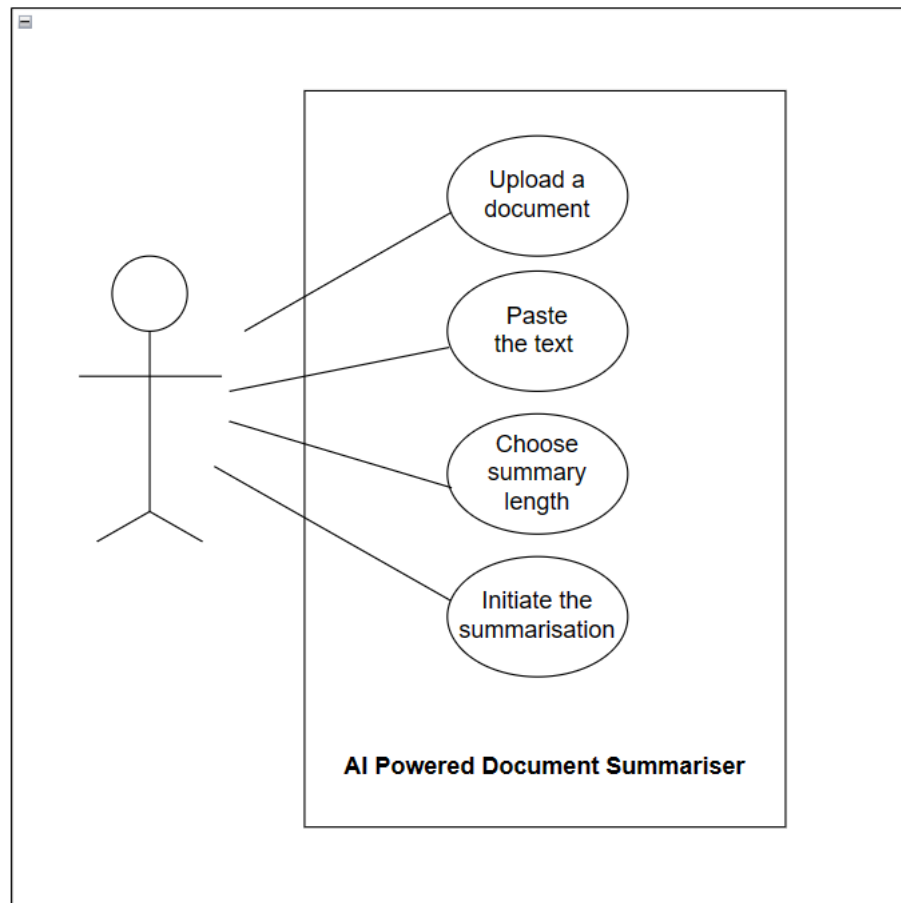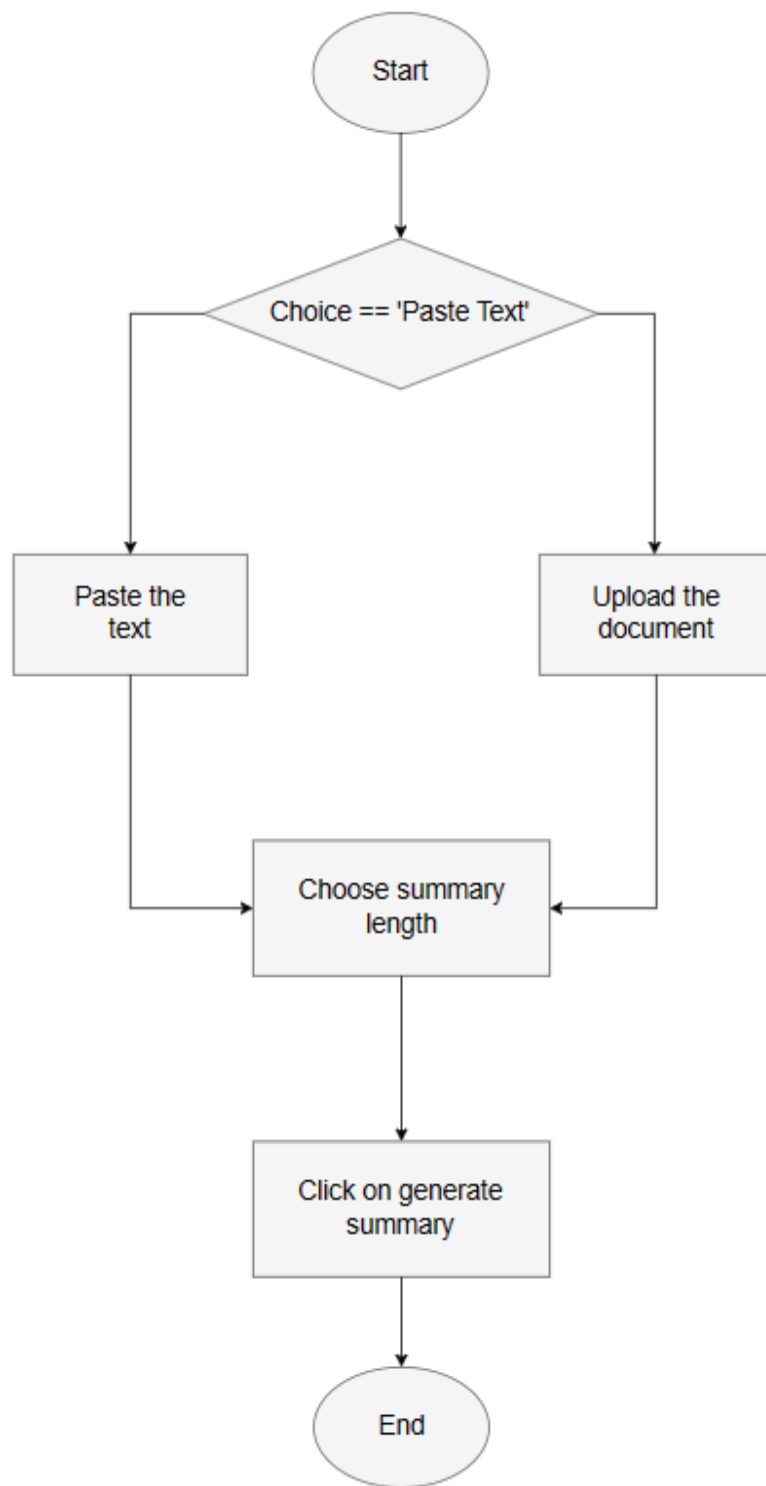
22114107

vineet_k@cs.iitr.ac.in

Project Abstract

The proposed software is an AI-powered document summarization tool that accepts various input formats, including PDF, DOCX, and text pasting and generates extractive summaries in real-time. Users can choose from short, medium, or large summary options, and the tool supports multiple languages for English summarization. Additionally, it offers word count statistics for both input and output, enhancing the user's control and understanding of the summarization process.

Design

Upload a document

Paste the text

Choose summary length

Initiate the summarisation

AI Powered Document Summariser

Use case model for Document Summariser

Flowchart representation for use cases

Various use cases are defined as follows:

1. Upload a document: This use case describes the scenario where a user wants to input a document into the system for summarization. The software should handle the uploaded document and extract its text content.

2. Paste Text: This use case covers the scenario where a user prefers to directly input text by pasting it into the system. The software should accept the pasted text and process it for summarization.

3. Summary length: Users can customize the summarization process level of detail required in the summary. This use case captures these choices.

4. Initiating summary generation: When the user is satisfied with their document input and summarization preferences, they trigger the system to generate a summary. This use case represents the core functionality of the system.

Questions

- What machine learning or NLP algorithms are used for summarization?

- Are there specific models or techniques employed, such as transformer-based models like BERT or GPT?

- Describe the step-by-step process of how the system generates summaries.

- How are sentences or key information selected for inclusion in the summary?