**Synopsis**

**Title: RAG Driven Video Summarization with Context Aware Chatbot**

**Abstract:**

The RAG Driven Video Summarization with Context Aware Chatbot is a web-based application that extracts key insights from YouTube videos and allows users to ask questions about the content. It automates transcription, semantic search, and response generation using advanced AI techniques. The system extracts audio from YouTube videos, transcribes speech to text using Whisper, and converts the transcriptions into embeddings stored in a FAISS vector database. When a user submits a query, the system retrieves relevant context and generates answers using a Hugging Face Transformer model. This approach enhances accessibility and improves efficiency in extracting useful information from video content.

**Objectives:**

* Automate transcription of YouTube videos using Whisper AI.
* Enable real-time Q&A based on video content using RAG (Retrieval-Augmented Generation).
* Improve retrieval efficiency using FAISS for vector search.
* Provide a user-friendly Streamlit-based interface for seamless interaction.
* Enhance the summarization and contextual understanding of video content.

**Methodology:**

**Audio Extraction:**

* Downloads only the audio from a YouTube video using yt-dlp.
* Saves the extracted audio in MP3 format for further processing.

**Transcription:**

* Uses OpenAI’s Whisper model for speech-to-text conversion.
* Processes the audio and converts it into accurate text format.

**Embedding Generation and Storage:**

* Uses SentenceTransformers to generate semantic embeddings from transcriptions.
* Stores embeddings in a FAISS vector database for fast retrieval.

**Question-Answering System:**

* Retrieves relevant context using FAISS based on user queries.
* Uses a transformer-based model (Google FLAN-T5) to generate precise answers.

**Deployment:**

* Developed using Streamlit for an interactive UI.

**Tools & Technologies Used:**

* NLP Models: Whisper, SentenceTransformers, FLAN-T5
* Vector Database: FAISS
* Frameworks: Streamlit, Hugging Face Transformers
* Languages: Python

**Expected Outcomes:**

* Automated and accurate transcription of YouTube videos.
* Efficient retrieval of relevant video segments for answering queries.
* Interactive and user-friendly interface for video-based Q&A.
* Improved accessibility by eliminating the need to watch entire videos.

**Applications:**

* Education: Helps students extract and summarize lecture content.
* Research: Assists researchers in quickly obtaining insights from academic videos.
* Content Creation: Supports YouTubers in generating captions and metadata.
* Accessibility: Improves content reach for hearing-impaired users.

**Conclusion:**

The RAG Driven Video Summarization with Context Aware Chatbot offers an innovative approach to extracting and interacting with YouTube video content. By integrating speech-to-text conversion, semantic search, and AI-driven question-answering, the system enhances accessibility, improves efficiency, and provides users with a seamless way to gain insights from videos.