IntelliDoc Voice

Voice-Powered Intelligent PDF Assistant

Project Overview

IntelliDoc Voice is a CLI-based voice-enabled AI assistant that allows users to upload a PDF document and interact with it using voice commands.

The system uses:

- LiveKit for real-time voice interaction
- LangChain to manage and coordinate the AI agent
- RAG (Retrieval-Augmented Generation) for document-based Q&A
- Tool Calling for executing utility functions

It intelligently decides whether to:

- Retrieve information from the PDF (RAG)
- Call a tool (calculator, word counter, summarizer, etc.)
- Or combine both

Problem Statement

Traditional PDF readers only allow manual reading and searching. Users cannot:

- Ask natural language questions
- Perform analysis on document content
- Use voice to interact
- Execute contextual operations (e.g., count occurrences, calculate totals)

IntelliDoc Voice solves this by enabling intelligent, voice-based document interaction.

Core Features

1. Voice Interaction

- User speaks a question
- Speech converted to text
- AI processes request
- Response converted back to voice

2. PDF-Based Question Answering (RAG)

- User uploads a PDF
- Document is chunked
- Embeddings created
- Stored in FAISS vector database
- Relevant chunks retrieved during query

3. Intelligent Tool Calling

The system automatically detects when to use tools.

Implemented Tools:

Tool	Purpose
Calculator	Solve math expressions
Word Counter	Count word occurrences in document
Summarizer	Summarize entire document
Number Extractor	Extract numeric values
Time Tool	Return current system time

4. CLI-Based Execution

Everything runs from:

- VS Code
- Windows terminal