AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

HACKxAMRITA
IGNITE · INSPIRE · INNOVATE

ECHO-VISION

FutureX

ED-TECH

BY

SIVAPURAM VENKATA SAI SASHANK
RANGA SAI VARUN TALLURI
EDAPALAPATI VENKATA BALA JYOTHI SWAROOP
TELAGAMSETTY VISWAJITH GUPTA
KOUSHIK SIVAPURAM

PROBLEM STATEMENT

HACKxAMRITA
IGNITE · INSPIRE · INNOVATE

Problem Statement:
Voice-activated AI summarization and reading assistant for blind learners.

Problem Context:
Blind or visually impaired learners often face difficulty accessing digital documents like PDFs, especially when they want to extract information or summarize specific sections. Current AI summarization tools require manual selection or clicking, which makes them inaccessible to people who rely on voice commands.

Core Challenge :
Design a voice-controlled AI system that allows blind users to select text or sections in a document using natural spoken commands, summarize the selected content accurately, read the summary aloud in a clear, human-like voice. The system must work across multiple file types (PDF, Word, e-books).

PROPOSED SOLUTION

HACKxAMRITA
IGNITE · INSPIRE · INNOVATE

➢ **Objective:** Enable digital independence for the visually impaired by transforming static files into interactive, voice-controlled conversations.

➢ **Key Features:**

- **Instant Audio Response:** Uses LLM streaming (Groq/Llama 3) for zero-latency AI replies.
- **RAG Intelligence:** Ask questions ("Summarize this", "What's the conclusion?") and get precise answers.
- **Human-Like Audio:** Used Coqui tts for human like voice generation.
- **Making an Extension:** Voice generation while user moves his cursor on the screen

➢ **Workflow:**

- **Ingest:** User uploads a file; the system extracts and structures the text locally.
- **Command:** User speaks *"Ask what the main topic is"*.
- **Interact:** System retrieves the exact context, streams the AI response, and synthesizes it into natural audio on the fly.

TECH STACK

HACKxAMRITA
IGNITE · INSPIRE · INNOVATE

- **Frontend/UI**
  - **React (Vite):** Fast, responsive UI framework.
  - **Web Speech API:** Native browser API for Speech-to-Text (Voice Commands).
  - **High Contrast CSS:** Custom styling for visual accessibility.
- **Backend**
  - **Python (FastAPI):** High-performance async backend serving API endpoints.
  - **Groq Cloud (Llama 3):** Ultra-low latency LLM inference for real-time document Q&A and summarization.
  - **Local File System:** Direct OS-level file manipulation.
- **Core Tech**
  - **RAG Pipeline:** Retrieval-Augmented Generation for intelligent Q&A.
  - **Coqui VITS:** Advanced Neural Text-to-Speech (TTS) for offline, human-like audio output.
- **Any other technology used**
  - **PyPDF / Python-Docx:** Robust text extraction libraries.
  - **Uvicorn:** ASGI Server for fast, async Python execution.

UNIQUE SELLING PROPOSITION

- **Research Gaps**
  - **The "Context Gap":** Standard screen readers (JAWS, NVDA) are linear; they can't summarize or skip to "the important part" without tedious navigation.
  - **The "Input Gap":** Current GenAI tools (ChatGPT, ChatPDF) rely on visual interfaces (drag-and-drop, typing), making them inaccessible to the blind.
- **Differentiation (Differentiation between others and Novelty)**
  - **vs. Screen Readers:** We don't just read; we *understand*. We replace "Next Paragraph" commands with "Summarize Chapter".
  - **vs. AI Chatbots:** We are **Voice-First & Local**. No need to find the upload button or type queries. The file system is integrated directly into the conversation.
- **Innovation**
  - **"Voice-to-Vector" Pipeline:** First-of-its-kind workflow that converts spoken natural language directly into semantic vector search, bypassing the keyboard entirely.
  - **Autonomous Agent:** The system doesn't just wait for input; it actively navigates the OS to find and retrieve knowledge.

REAL WORLD APPLICATION

HACKxAMRITA
IGNITE · INSPIRE · INNOVATE

➢ Target Audience

- **Blind & Visually Impaired Students:** Who need to study from textbooks and research papers without relying on scribes or slow screen readers.
- **Professionals with Vision Loss:** Lawyers, researchers, and office workers who need to review contracts and reports efficiently.
- **Elderly Users:** Who struggle with small text on screens and complex technology interfaces.

➢ Real-life Impact

- **Academic Independence:** A blind student can now say "Find my history book and quiz me on Chapter 4," studying entirely on their own.
- **Workplace Equality:** A blind lawyer can quickly find specific clauses in a 100-page contract using voice search, working at the same speed as sighted colleagues.
- **Digital Inclusion:** Bridges the gap between static digital content and non-visual users, making the digital world truly accessible.

FEASIBILITY & VIABILITY

HACKxAMRITA
—IGNITE · INSPIRE · INNOVATE—

- **Practicality**
  - **Low Barriers to Entry:** Runs on standard laptops (Windows/Mac) without requiring expensive specialized hardware (like Braille displays).
  - **Offline Capability:** Core features (Navigation, TTS) work offline, ensuring reliability even without internet.
  - **Ease of Adoption:** Zero learning curve—if you can speak, you can use it.
- **Scalability**
  - **Modular Architecture:** The RAG pipeline can easily switch from local models to cloud APIs (GPT-4) as user needs grow.
  - **Format Agnostic:** Can be extended to support Emails, Webpages, and Code files with minimal changes.
- **Business Impact**
  - **Cost Reduction:** Reduces the need for human scribes and expensive assistive technology.
  - **Productivity Boost:** Enabling blind employees to work faster directly increases their employability and economic contribution.
  - **Market Potential:** Massive underserved market of 2.2 billion vision-impaired people globally (WHO).

FUTURE SCOPE & ROADMAP

HACKxAMRITA
IGNITE · INSPIRE · INNOVATE

➢ Further Innovation Possibilities

- **Advanced OCR Integration:** Incorporate vision-language models (like GPT-4V or LLaVA) to describe complex charts, graphs, and images within PDFs, not just text.
- **Multi-Lingual Support:** Expand TTS and Speech-to-Text to support more languages, making it a global accessibility tool.
- **Cloud Sync & Cross-Device:** Allow users to start reading on their laptop and continue on their phone.
- **Contextual Memory:** The AI remembers past conversations ("What did we read last week?").

➢ Conclusion

- EchoVision is more than just a reader; it is an intelligent companion that restores autonomy to the visually impaired.
- By combining Voice Navigation with RAG Intelligence, we are not just solving a usability problem—we are opening doors to education, employment, and independence.

AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

HACKxAMRITA
IGNITE · INSPIRE · INNOVATE

THANK YOU