
International Institute of Information Technology Bangalore

# Vis-Assist ChatBot for Teachers

By - Kedar Deshpande - IMT2020523,
Dewanshi Dewan - IMT2020549,
Aakash Khot - IMT2020512

# Problem Statement

**Aim** : To empower **visually impaired** teachers in special schools to access educational experience on par with mainstream schools, utilizing Generative AI technology.

Visually impaired teachers in special schools don't have the right tools and materials they need for teaching. They struggle to find accessible content and face fewer opportunities for learning compared to teachers in regular schools.

1. Lack of Tailored Resources and Tools
2. Limited Accessibility of Educational Content

# Proposed Solution

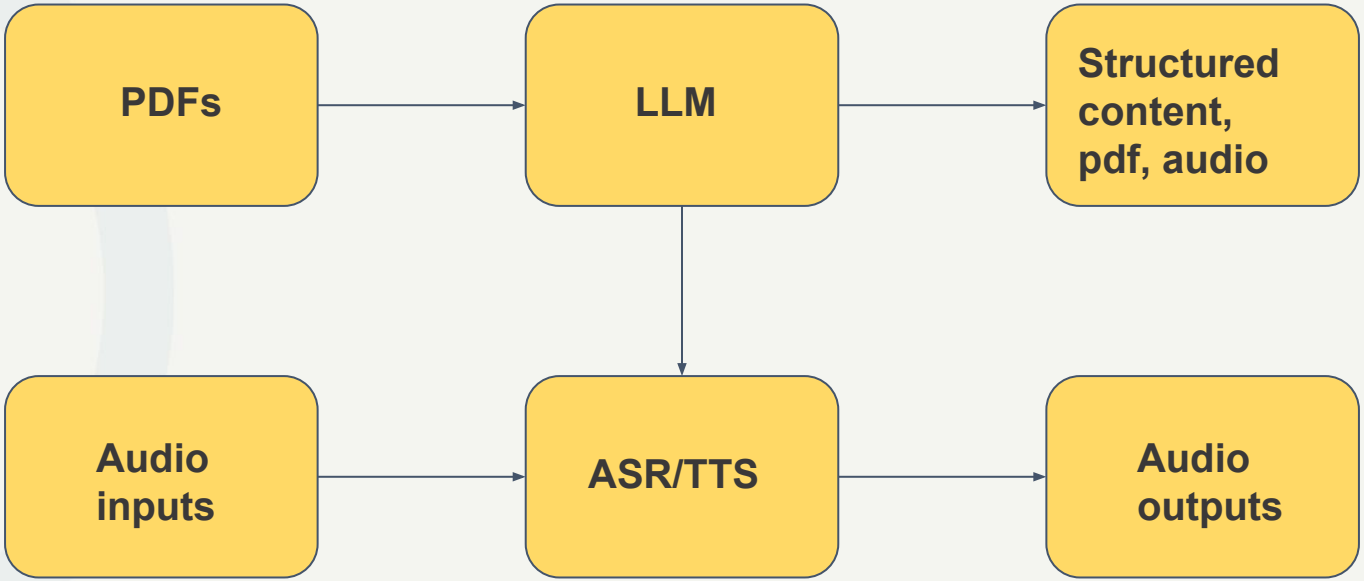
**A Learning Module application powered by Generative AI-** ***(LLM based tool to assist visually impaired teachers analyze and tailor content)***

Working:-

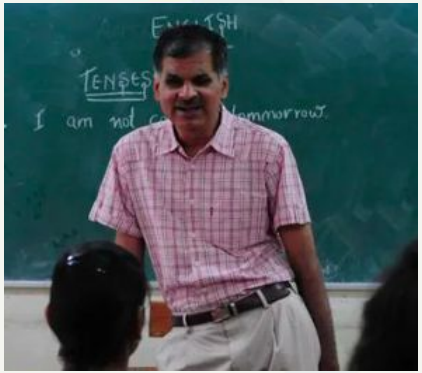
1) Visually impaired teachers upload documents or text materials onto the platform.
2) Generative AI processes the text, extracts key information, and generates a structured learning module.
3) The system will be accessible and interactive.

Teachers can interact with the system using audio to further refine the content according to their need.

# Architecture Followed



# Real-Life Accounts




*Prof. Rohit Trivedi of Sarojini Naidu Girls Post Graduate Government Girls' College, Nutan who set up a software to convert text books and notes into audio format.*

*Reetu Mansi, teacher at Navodaya Vidyalaya, Saharanpur who records Hindi audio study material for impaired students to prepare for exams.*

*Rakesh Kumar Soni, visually impaired teacher from Basti district, UP Uses Alexa in the classroom*

# Technologies to be used

1. Mixtral 8x7B (LLM) → Llama3
2. Whisper Large v3 (ASR)
3. XTTS v2 (Text to Speech)

Whisper to be used through Inference API provided for free by Hugging Face.

# Mixtral 8x7B – LLM

**LLMs** ( Large Language Models) - “Giant Language learners”

- What’s Mixtral ? What’s Special about it ?

  Mixtral stands out because it uses a technique called **Mixture-of-Experts (MoE).**
- In Mixtral 8x7B, “8x7B” stands for 8 groups of experts, and each group has about 7 billion parameters.

# Llama3 – LLM

**Llama3 -** Latest LLM developed by **META**

- Matches or exceeds the performance of proprietary models like Gemini Pro 1.5, thanks to its extensive training data, larger context size.
- Can handle larger prompts, improving context understanding.

| | Meta Llama 3 70B | Gemini Pro 1.5 Published | Claude 3 Sonnet Published |
|---|---|---|---|
| MMLU 5-shot | 82.0 | 81.9 | 79.0 |
| GPQA 0-shot | 39.5 | 41.5 CoT | 38.5 CoT |
| HumanEval 0-shot | 81.7 | 71.9 | 73.0 |
| GSM-8K 8-shot, CoT | 93.0 | 91.7 11-shot | 92.3 0-shot |
| MATH 4-shot, CoT | 50.4 | 58.5 Minerva prompt | 40.5 |

# Whisper Large v3 – ASR

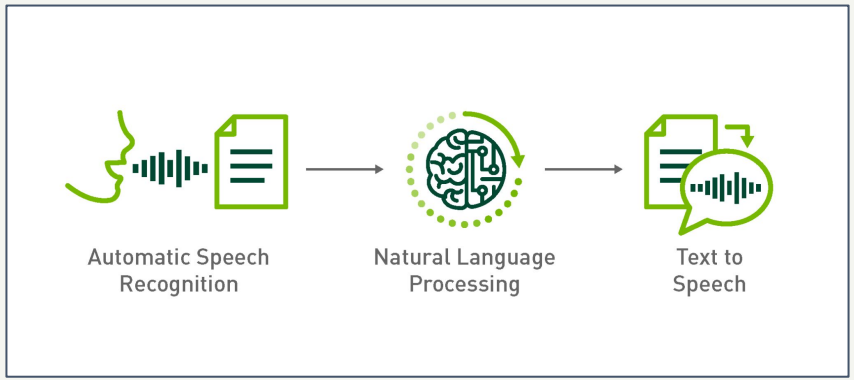
Automatic Speech Recognition (ASR) utilizes AI&ML technologies to interpret and transcribe spoken language into written text.

Various ASR Models - Whisper Large v3, Distil-Whisper, SeamlessM4T v2 etc.

**Whisper Large v3**

- Reliable and understands speech

  even in noisy environments



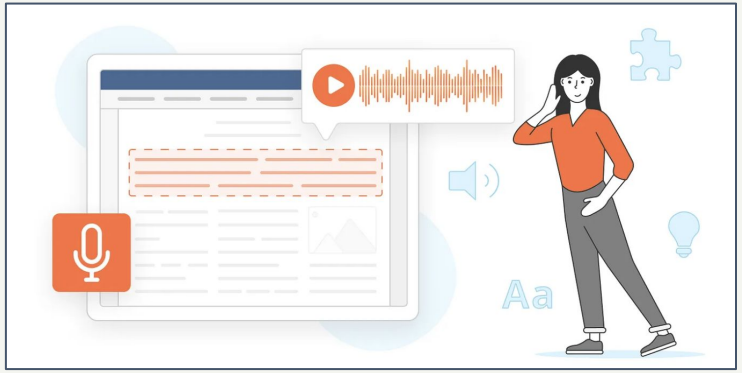
# XTTS v2 – TTS

Text to Speech (TTS) - Text to Spoken words

Various TTS Models - VALL-E, XTTS v2, FluentSpeech etc.

**XTTS v2**

- It allows us to create spoken versions of educational materials with high-quality audio output.



# Some Frameworks used - **Streamlit** & **Langchain**

**LangChain** is a toolkit for creating chatbots and conversational AI models.

In our project: We used LangChain to implement a retrieval-based question-answering system, allowing our chatbot to provide informative responses to user queries based on the input text.

**Streamlit** is a Python library for building interactive web apps.

In out project: We utilized Streamlit to create an interactive web application that allows users to upload PDF documents, record audio inputs, and download tailored content generated by the chatbot.

# How the Models are working-

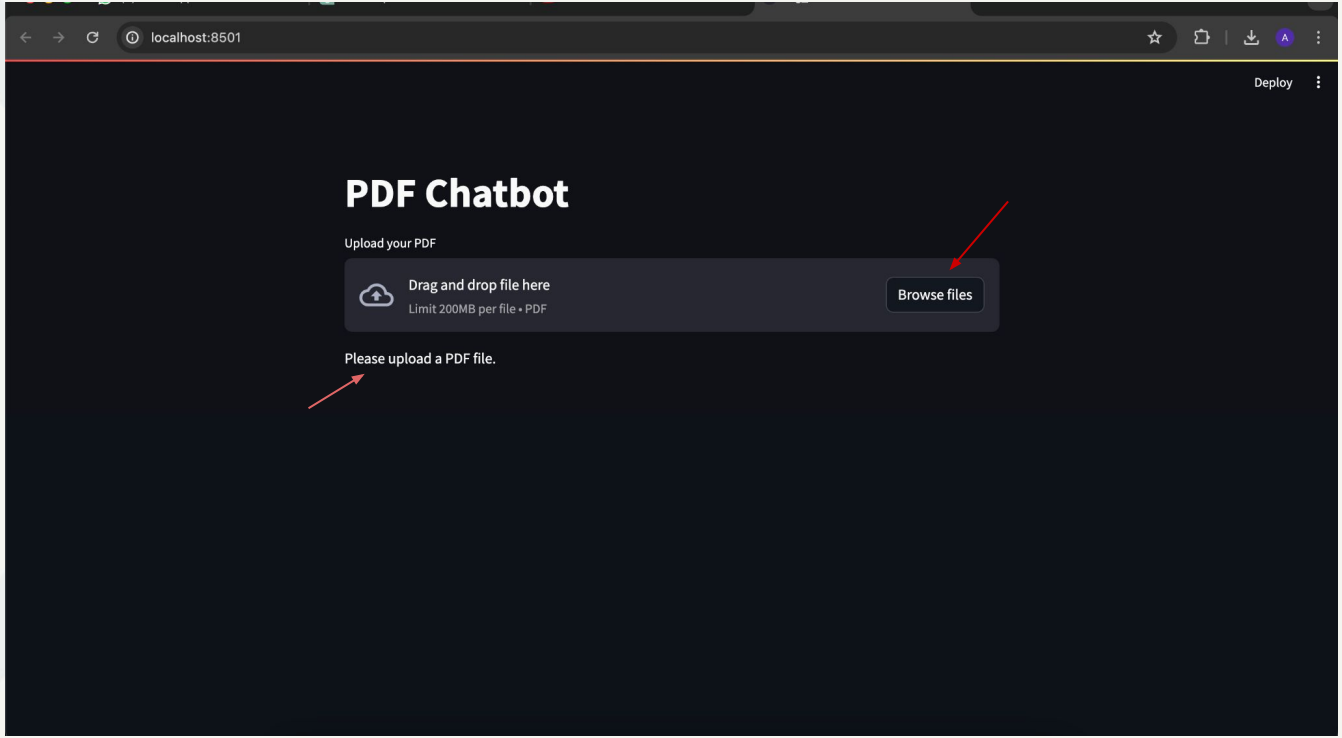
1) **Model Loading**: The code loads various AI models for language understanding and generation, such as retrieval QA models and language models like Ollama.
2) **Document Processing**: When a PDF is uploaded, the code analyzes its content using text processing techniques like text splitting to prepare it for further interaction.
3) **Interactive Chatbot**: It employs a chatbot interface where users can interact via text or audio. The chatbot generates responses using the loaded language models.
4) **Speech Synthesis**: The generated responses are converted into speech using a text-to-speech (TTS) model, allowing users to hear the bot's replies.
5) **PDF Generation**: Tailored content created is then structured and converted into a PDF format for download.

# Problem Faced

1. User Interface (UI) Complexity: Managing the UI can become complex, especially when dealing with multiple interactive elements such as file uploaders, chatbots, and download buttons. Ensuring a smooth and intuitive user experience can be challenging.
2. Difficulties integrating various libraries and APIs can lead to compatibility issues and delays.
3. API rate limits

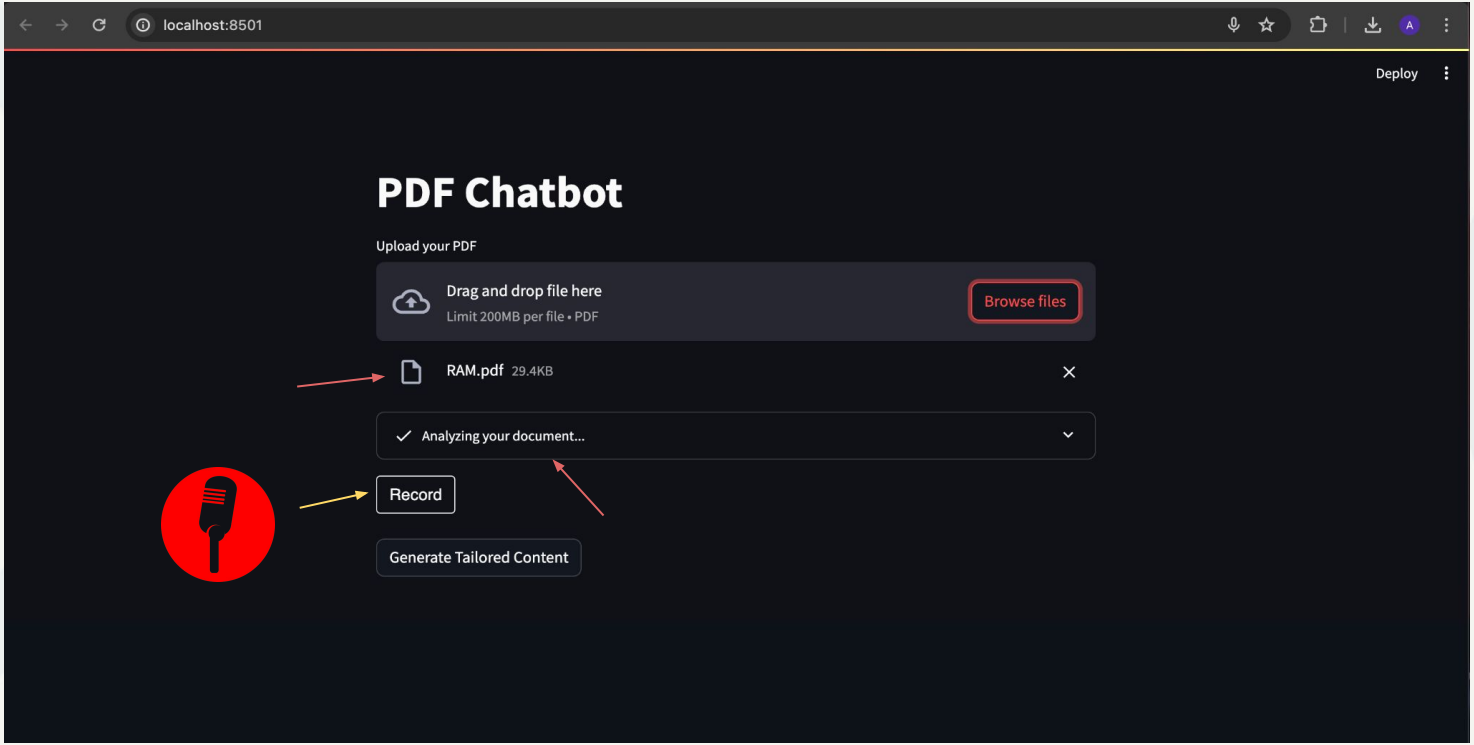
# Step-by-Step Walkthrough of Application

**Step 1**



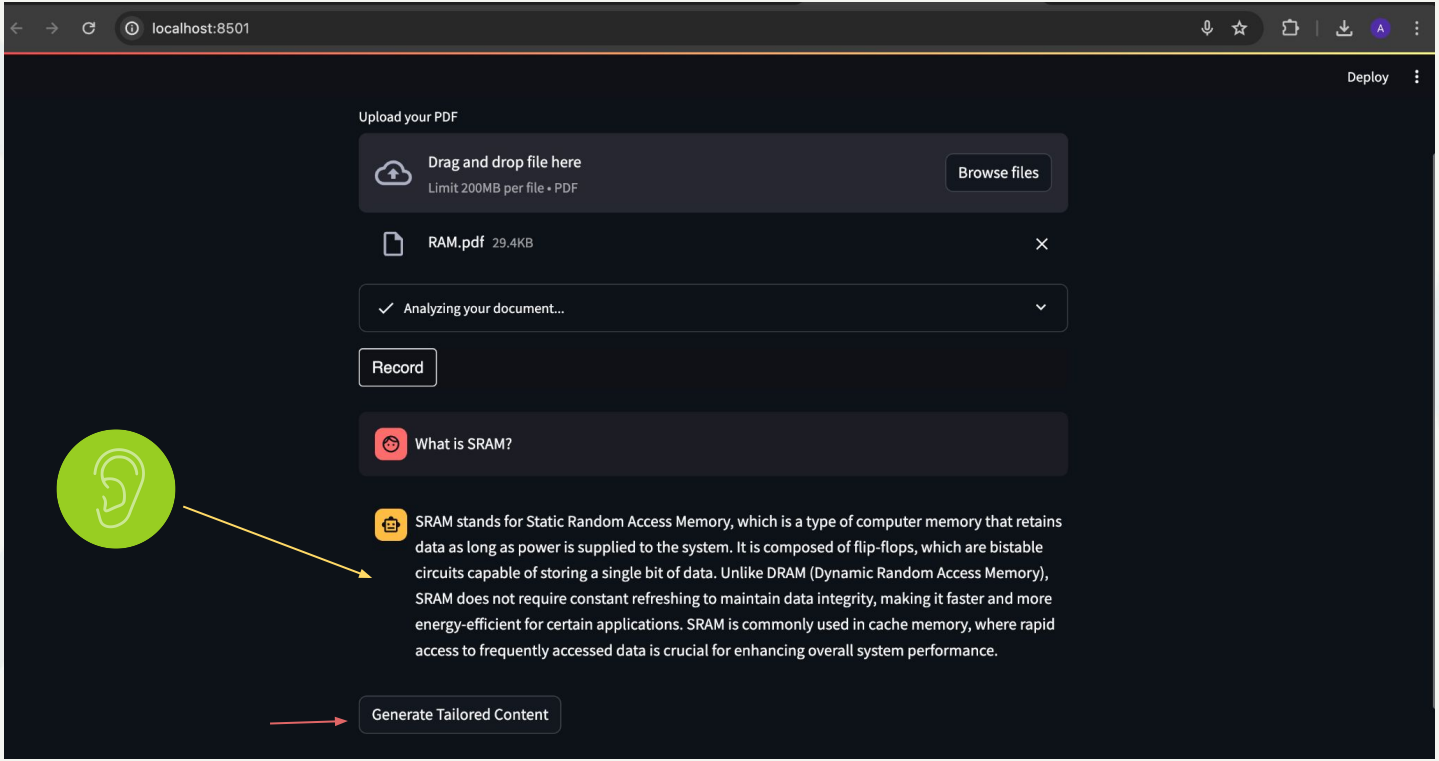
# Step-by-Step Walkthrough of Application

**Step 2**



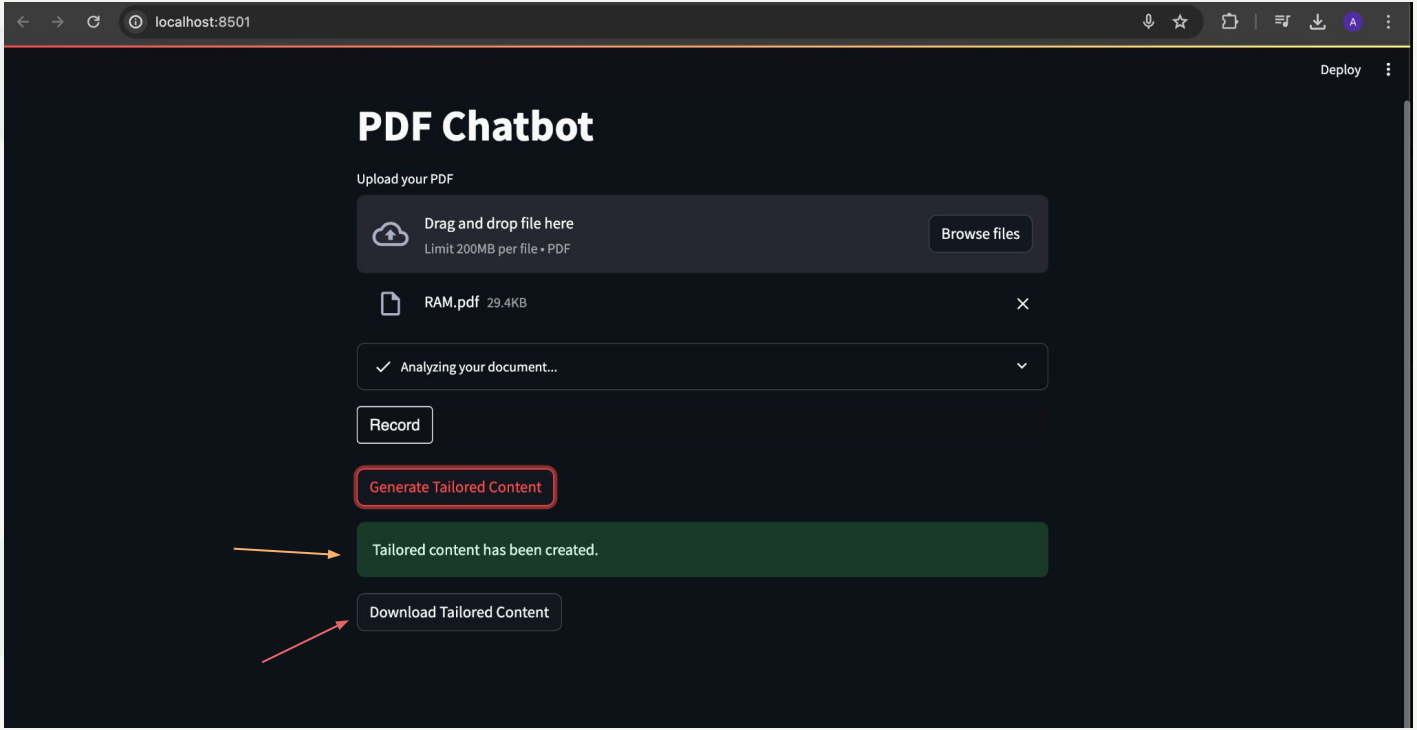
# Step-by-Step Walkthrough of Application

**Step 3**

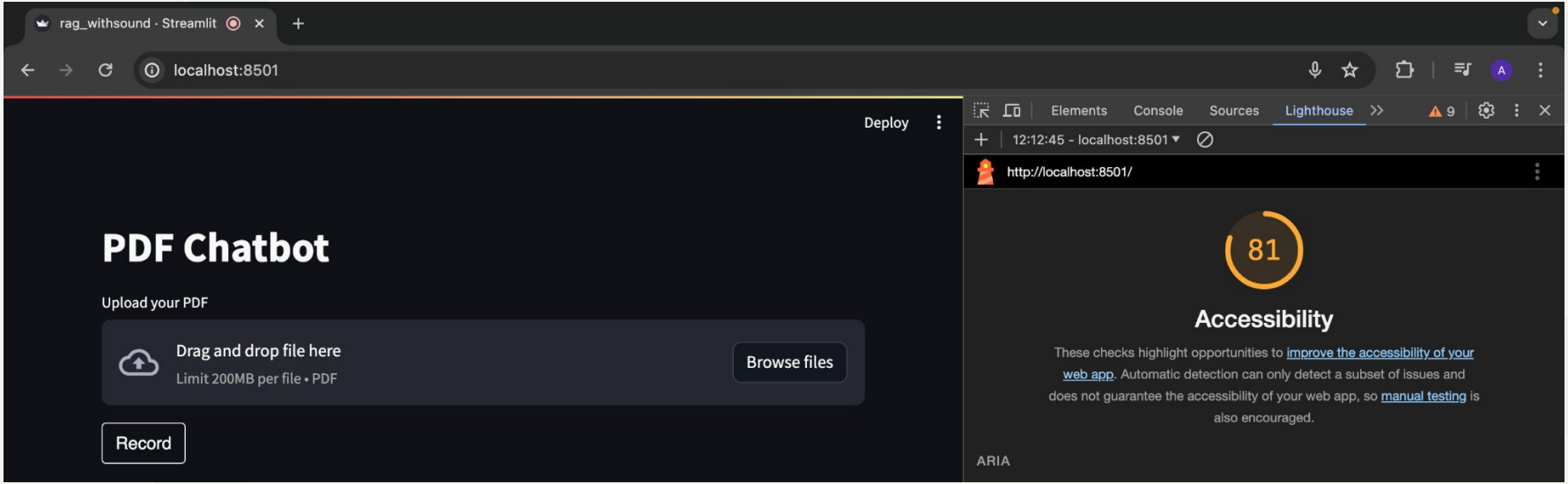


# Step-by-Step Walkthrough of Application

**Step 4**



# Google Lighthouse Accessibility Score Obtained




**81%**

# Some Valuable Feedback

- Different buttons for voice generation, give user needs a voice assist or the user is using any type of screen reader.

  → Overlapping of screen reader and our voice assist.

- Setting the page order correctly for better navigation.
- Error handling

# Future Scope

- Collaborate with visually impaired educators to gather feedback and insights for version 2.
- Conduct pilot testing in special schools with visually impaired teachers and students.
- Incorporate features to bookmark and save answers/passages as well as synchronize better with screen reader applications.

# References


1. Generative AI for Special Educators - https://www.youtube.com/watch?v=Yp0KdkZpaVU&ab_channel=AIforEducation
2. https://www.hindustantimes.com/bhopal/visually-challenged-prof-gives-a-new-vision-to-students-in-bhopal/story-WgWCJndpINtX7ofTtrxj8I.html
3. https://www.timesnownews.com/education/teachers-of-india-story-of-rakesh-and-how-he-effectively-used-ai-tools-to-teach-visually-impaired-children-article-103388570
4. https://thelogicalindian.com/education/meet-reetu-mansi-visually-impaired-teacher-who-creates-study-material-to-help-differently-abled-people-crack-entrance-exams-32392
5. https://medium.com/@vimalathithanswetha/ollama-unleash-the-power-of-large-language-models-on-your-local-machine-df8907e13580

# Thanks