Project Name: Transcripto

Team Members:

Harika J
Haikoo Khandor
Dhruv Patel
DarshnilSinh Rana

Project Proposal

What exactly is the function of your tool?

Transcripto is an AI-powered speech-to-text and summarization application that transcribes audio files into accurate text using Gemini 2.0. Beyond transcription, it also generates concise summaries and answers user queries based on the transcribed text. The interface, built using Flet, provides a simple and intuitive platform where users can upload audio files and receive real-time transcriptions, summaries, and intelligent responses.

Why would we need such a tool, and who would use it?

- **Students and Researchers:** Convert recorded lectures, interviews, and discussions into summarized text for easy referencing.
- **Businesses and Professionals:** Generate meeting notes, concise summaries, and action points automatically.
- **Podcasters and Journalists:** Create summarized transcripts from interviews and discussions for articles and reports.
- **General Users:** Get key insights and answers from long recordings without reading the full transcript.

This tool reduces manual transcription and summarization time, making content consumption more efficient across various domains.

Does this kind of tool already exist?

There are tools like Otter.ai and Descript, but Transcripto offers:

- **Real-time processing using Gemini 2.0, ensuring higher accuracy.**
- **Automated summarization and question-answering capabilities.**
- **Lightweight and accessible UI using Flet, requiring no complex setups.**
- **Cost-effective by leveraging open-source frameworks and free-tier AI APIs.**

Users will appreciate the differences in affordability, real-time accuracy, and the additional functionality of summarization and Q&A.

How do you plan to build it?

- **Data Collection:** Users upload audio files (MP3, WAV, etc.) for processing.
- **Core Algorithm:**
    - Use Gemini 2.0 API to transcribe speech into text.
    - Pre-process audio using Pydub for better recognition.
    - Generate text summaries using NLP models.
    - Implement a question-answering module based on the transcribed content.
- **User Interface:**
    - Flet-based web UI for easy file uploads, viewing transcriptions, and summaries.
    - Option to edit, save, and download transcripts and summaries.

What existing resources can you use?

- Google Gemini 2.0 API for transcription and summarization.
- Flet framework for UI development.
- Pydub library for audio processing.
- OpenAI or Hugging Face models for question-answering.
- Cloud hosting options like Heroku or Streamlit for deployment.

How will you demonstrate the usefulness of your tool?

- **Live Demo:** Users upload audio and receive instant transcriptions and summaries.
- **Comparative Testing:** Benchmark accuracy and summarization quality against Otter.ai and Whisper AI.
- **User Feedback:** Testing with students, professionals, and podcasters.

Rough Timeline & Milestones

Week	Task
1-2	Set up Gemini API & Flask
3-4	Implement transcription, summarization & audio processing
5-6	Build user-friendly UI & testing phase
7-8	Final testing, evaluation, and deployment

This project will provide an efficient, real-time speech-to-text solution with automated summarization and Q&A capabilities, making transcription more accessible, cost-effective, and accurate for a wide range of users!