This project processes patient inquiry audio files to generate a clinically appropriate and multilingual front-desk response for medical practices using speech-to-text, translation, intent classification, and LLM-based response generation. It supports multiple languages and distinguishes patient intents in the categories billing inquiry, appointment scheduling, prescription refill, general inquiry, medical emergency, and insurance coverage inquiry.

Tools and Libraries Used

Speech-to-Text	whisper	Transcribes multilingual audio to text
Translation	deep_translator	Converts non-English transcriptions to English
LLM (response generation)	Google Gemini	Generates appropriate responses to the patient's inquiry in the patient's language.
Date Handling	datetime	Timestamping json output
File I/O	os, json	File management and structured output generation

Assumptions and Shortcuts

- Created training data based on ChatGPT-generated ideas for possible patient inquiries for a medical practice. Determined commonalities to define labels and labeled data points manually.
- Assumes whisper base model is sufficient for transcription accuracy.
- Default translator used is deep translator, which is Google Translate's backend.
    - Gemini model could also be used, but unable to have 2 Gemini models on free plan
- Intent initially done with rule-based method, later replaced with transformer-based encoder and linear classifier
    - Training data can be expanded to include multi-lingual data points in the future if want to skip translating audio to English prior to intent classification
- API Key: need to have a valid Gemini API key
    - Currently removed API key from files, in future can put into .env file and add to .gitignore

Considerations for Scaling in a Production Healthcare Setting

- Ensure HIPAA compliance, encrypt all data at rest and in transit, restrict access
- Add confidence thresholds for Whisper language detection
- Validate LLM outputs through post-processing checks to prevent hallucinations and misinformation
- Integrate with phone number
- Log model decisions and output, alert for failed transcriptions, low-confidence intent classifications, abnormal LLM responses
- FastAPI endpoint
- UI to upload audio and display results