Intelligent Consent Flow for Medical Care
Short Design Document

Overview

This project is a Streamlit-based web application that uses modern AI and speech technologies to help patients understand a medical procedure, ask questions, and provide digital or verbal consent in a simple, friendly, and privacy-sensitive way.

Tech Stack Decisions

- **UI Framework:** Streamlit was chosen because it allows fast prototyping of user interfaces for audio, text, and signature capture, with minimal code and rapid iteration.
- **Large Language Model (LLM):** Perplexity AI API ("sonar-pro" model, via Pro subscription) was selected for generating medical consent summaries and answering patient questions. It supports state-of-the-art quality, reliable plain-language outputs, and multi-model flexibility, and is easy to call from Python.
- **Speech Recognition:** OpenAI Whisper (run locally) provides accurate and private speech-to-text (STT) transcription for voice-based questions and detecting verbal consent.
- **Text-to-Speech (TTS):** gTTS was used for spoken feedback of summaries and answers since it works offline, is simple, and supports many languages.
- **Digital Signature:** streamlit-drawable-canvas enables in-browser, device-agnostic signature capture using mouse or touchscreen.
- **Logging:** Minimal, privacy-first file logging (medical record and timestamp only) to avoid storing sensitive personal or conversation data unless explicitly required.

Integration of Speech and LLM

- **Input:** Users can enter questions by typing or by speaking. If speaking, the app records audio, processes it with Whisper for transcription, and then uses the resulting text.
- **LLM Conversation:** All medical notes and patient questions are sent to the Perplexity AI LLM via a secure API call. The LLM produces patient-friendly summaries or clear answers, which are shown in the UI.
- **Output:** All LLM-generated text (summaries or answers) can be played aloud by generating an audio file with gTTS and streaming it back in the browser.
- Both text and spoken interactions are supported at every stage, and the system maintains in-memory Q&A history within the session for user reference (but does not log it for privacy).

Edge Cases Considered

- **Patient is Unsure/Repeats Questions:** The app allows repeated, free-form questioning. The AI never complains if a question is asked again or is unclear, and always provides a calm, simple answer.
- **User Speaks Unclearly (Speech-to-Text Fails):** The app displays what was transcribed and prompts the user to retry if their question or verbal consent could not be understood.
- **Partial/Inaccurate Consent Phrase:** If Whisper does not detect the phrase "I consent" in audio, the app asks the user to speak more clearly and re-record.
- **Skipped Steps or No Input:** All buttons are disabled when their required input is missing, and the user is guided with clear instructions for what to do next.
- **API/Network Failure:** Any failure in the LLM API, Whisper, or TTS triggers a friendly error message and a prompt to retry without losing current progress.
- **Accessibility:** Both text and audio (TTS) output are always available for summaries and answers.
- **Privacy:** Only the minimal procedure record is stored—no conversation logs, voice, or signature files are retained.