I have an exciting idea. I want to modify the code below to generate a synthetic database of clinical notes for biomedical informatics research. We're going to need to first define our workflow for the LLM, then we define how we leverage existing data to create lifelike notes, and then we need to write the code.

My rough idea for the workflow is the following

1. Create a table of physicians, we need to define a variety of characteristics about each of them. Those characteristics include are they detail oriented or do they tend to document the minimal amount required. How often do they make a mistake when typing where characters are reversed. What is their background, east/west coast trained, ethnicity, etc. essentially anything that will result in a more distinctive writing style. We need to define whether they follow a template or take creative liberties with the templates they are provided. We also need to define which clinicians cover for one another as we will have PCP's assigned but occasionally will have them see other providers. (The workflow to assign the provider seeing someone will be outside the LLM's control, a random number generator with a percentage assigned might be appropriate. We're going to use OMOP tables where we might just sync their provider table with the one we generate.)

2. We are going to assume we have an OMOP database already populated with patients When we go through the workflows we want to start with a single patient and go through all their encounter records and generate notes for those encounters. We need to craft a number of queries that will gather all the information we need for those clinical notes to be built.

3. Before we generate the clinical notes we need to create a SQL table (Not in omop) for that specific note where we craft the prompt to include extraneous information that OMOP would not collect but that would be relevant if we are trying to extract data from clinical notes. Because we need to have that data discretely captured so we can compare it in future steps when we're extracting clinical data from the notes. Please make sure to include a hook in this step where if we're interested in generating notes that include specific freetext data we can inject it into the generation, EG we might want mentions of guns in home or illegal drug usage specifically. Don't forget the table created needs to be able to link to OMOP's tables

4. When we generate the clinical notes the prompt that will generate the clinical notes needs to include all the custom provider information, a clinical note template for that clinic, and all the patient specific information, all the note specific information. We will generate the note, then we grade the note generated to ensure its meeting all the objectives, if it does not we throw it away and generate another attempt.

5. We iterate through an individual patient and then we go to the next patient in the queue. The notes that are generated should be saved to a JSON object which we are appending/adding to with each new note.