LLM Engineer Task: Clinical Note Summarizer

Task Title: Building an Intelligent Clinical Note Summarizer with Healthcare Context

Objective:

Create a Python-based summarizer for clinical notes using a pre-trained LLM (such as GPT, BioBERT, or other NLP models). The summarizer should extract key medical information such as diagnoses, medications, procedures, and clinical findings from raw text notes. After extraction, it should generate a concise summary that prioritizes the most critical information, like conditions requiring urgent attention or uncommon diagnoses.

Instructions:

1. Brainstorming (Approx. 30 minutes):

- Explore possible ways to summarize clinical notes using NLP models. Consider how you can prioritize critical information (e.g., prioritizing certain diagnoses like sepsis or conditions with specific markers such as high lactic acid).

- Identify how you would integrate medical knowledge (e.g., using ICD-10 codes or clinical decision rules like SIRS criteria).

- Come up with a plan for handling noisy or unstructured clinical data and how to clean it up for better model input.

2. Coding (Approx. 1 hour 30 minutes):

- Write a Python script that:

- Loads raw clinical notes (mock data will be provided).

- Extracts key information from the notes using a pre-trained LLM (you can use GPT-4,

BioBERT, or any other pre-trained healthcare-specific model available).

- Summarizes the extracted information into a concise, readable format.

- You should handle common issues like abbreviations (e.g., "CHF" for congestive heart failure), structured data (like lab results), and unstructured free text.

- Optional: If possible, prioritize urgent conditions based on certain keywords or clinical flags (e.g., "severe sepsis" if SIRS criteria are mentioned).

3. Test Input:

Here is a mock clinical note for testing purposes:

45-year-old male with a history of CHF, diabetes mellitus, and chronic kidney disease. Presented to ER with fever, tachycardia (HR: 120 bpm), and hypotension (BP: 90/60). Labs: WBC 18.3, lactate 4.5 mmol/L, pH 7.32, serum bicarbonate 18 mEq/L. Blood cultures pending. Suspected diagnosis: sepsis. Plan: Start broad-spectrum antibiotics and fluids. Admit to ICU.

4. Output Example:

The script should generate a summary like:

Summary:

"45-year-old male with a history of CHF, diabetes, and CKD presents with suspected sepsis (tachycardia, fever, hypotension, high lactate). Initial labs: WBC 18.3, lactate 4.5 mmol/L, pH 7.32. Plan: Admit to ICU, start antibiotics and fluids."

Deliverables:

- Python code that implements the above solution.
- A brief explanation of how you prioritized the information in the summary.

Notes:

- Make sure the task is achievable within the timeframe by keeping the code structure simple, and leveraging available pre-trained models for the NLP tasks.

- Feel free to use any libraries like spaCy, transformers, pytorch, etc.