## **Project Task (Brainstorming during Interview):**

### **Building Patient Personas for Clinical Trial Simulation**

**Objective:**Design an ML-driven system to generate **User Personas** representing diverse patient pools, based on clinical trial data. These Personas will be used to:

* **Simulate patient-trial matching**.
* **Predict performance** (outcomes, eligibility, dropout risk, etc.) in new clinical trials.

**Problem Statement:**Clinical trials often suffer from recruitment inefficiencies and high dropout rates because of mismatched patient characteristics.  
You are tasked to brainstorm a way to:

1. **Cluster patients** into representative **Personas**.
2. **Map real patients** to these Personas.
3. **Pass Personas** through various **trial protocols** to **predict how they might perform** (e.g., success, risk, fit).

## **Guiding Questions for Brainstorming:**

* **Data Understanding:**
  + What patient-level data would be necessary? (Demographics, conditions, lab values, medications, etc.)
  + What trial-level data would be needed? (Inclusion/exclusion criteria, endpoints, interventions)
* **Personas Creation:**
  + How would you cluster or group patients into meaningful Personas? (e.g., K-means, hierarchical clustering, autoencoders, latent space models)
  + How would you ensure that Personas are both **diverse** and **clinically meaningful**?
* **Patient–Persona Mapping:**
  + How would you assign a new patient to a Persona?
  + What distance metrics or embedding techniques could be useful?
* **Trial Simulation:**
  + How would you simulate a Persona participating in a trial?
  + How could you predict trial success/failure/dropout per Persona? (e.g., classification models, survival analysis)
* **Evaluation:**
  + How would you validate if your Personas accurately represent real patient outcomes in clinical trials?
  + How could you iteratively improve the Persona definitions?

## **Bonus Discussion (Optional if time permits):**

* How would you incorporate **rare diseases** or **underrepresented subgroups** into your Personas?
* How might **real-world evidence (RWE)** (e.g., insurance claims, EHR data) augment the creation of better Personas?
* How would you adapt this system to **new incoming trial designs** automatically?

**Expectation during the Interview:**

* Think **aloud**.
* It’s okay to propose **multiple ideas** or **tradeoffs**.