Your Name: Vani Singh

MSDS 603 MLOps Assignment 1 – Part 2 (2.5%)
Requirements Gathering

In this assignment, you will gather the requirements for building a specific AI/ML-powered product. You will need to identify the business and technical requirements, assess potential risks, propose mitigation strategies, and outline the high-level components needed for successful implementation of the product. You **will not** need to actually build the product.

Learning Objectives

- Apply MLOps principles to a real-world product scenario
- Practice requirements gathering and analysis for ML systems
- Identify potential risks in ML systems and develop mitigation strategies
- Understand the core components required in an ML product pipeline

Scenario

An EdTech company is developing a personalized learning platform for K-12 students. The platform will use machine learning and AI to analyze student performance data from standardized tests and ongoing assessments within the platform to create customized learning pathways for each student. The system should adapt in real-time to student progress, identifying knowledge gaps, recommending appropriate learning activities, and adjusting difficulty levels to maximize learning outcomes while maintaining student engagement. The platform must eventually work across various subjects, but for now we will focus only on ***reading comprehension***. Ideally, it should accommodate different learning styles, comply with educational privacy regulations (like FERPA), and provide actionable insights to teachers and parents through intuitive dashboards.

Requirements

This assignment is done in **two parts**. Part One was already completed in class, and your answers to Part One should be available to you in Gradescope. Complete Part Two below at home and turn in to Canvas. If you did not attend class for Part One, you must accept a zero grade for this assignment since Part Two depends on your answers to Part One.

Part Two

In this part, use ***any resources you want*** (e.g. team members, internet, AI) to help you answer the below questions. Type your answers directly in this word doc.

Question 1: Define an additional two goals for this project.

- Provide actionable insights and progress tracking for teachers and parents to support student learning.

- Ensure the system can scale to accommodate growing student populations and expand across different subjects in the future.

Question 2: For each additional goal from Question 1; define a metric to measure success of that goal.

- Percentage of teachers and parents who report increased understanding of student performance (via surveys), and dashboard usage rate (measured by logins and engagement time).
- System latency under increasing user loads (e.g., 95th percentile response time stays under 1 second with 10x user base), and successful onboarding of new subjects/modules (tracked via release timeline and performance benchmarks).

Question 3: Briefly describe data governance considerations for the data sources you previously identified in Part One. Be sure to include data privacy and data quality requirements.

- Data Privacy:
    - All student performance data must comply with FERPA (Family Educational Rights and Privacy Act) to ensure that student records are handled securely. Personally identifiable information (PII) must be encrypted and access-controlled.
    - Consent from parents/guardians must be collected before data is used, especially for minors.
- Data Quality:
    - The platform should ensure that data collected from assessments and standardized tests is accurate, up-to-date, and representative.
    - Regular audits and validation rules should be in place to detect and correct anomalies, such as incomplete records, inconsistent scoring scales, or outdated materials.

Question 4: Identify an additional two risks associated with this product and the potential impact of each risk.

- **Bias in recommendations due to unequal data representation:** If the training data overrepresents certain demographics or regions, the platform may give skewed or unfair recommendations.
- **System fails to update content based on latest educational standards:** The platform may fall out of alignment with state or national curriculum guidelines, leading to irrelevant or outdated learning suggestions.

Question 5: For each additional risk identified in Question 4; propose a strategy to mitigate the risk.

Question 6: Describe, in words, any additional major architectural components needed for this product that you did not already include in Part 1 and how those components interact with each other and with components that you described in Part 1.

- **Content Management System (CMS):** A backend system to store, update, and tag educational materials by subject, grade level, difficulty, and learning style. The recommendation model interfaces with the CMS to pull relevant materials based on the student's profile.

- **Real-time analytics pipeline**: This component collects and processes student interaction data in real-time (e.g., quiz responses, time spent per activity). It feeds into both the evaluation and recommendation models to allow real-time updates to the learning pathway.
- **Dashboard & Notification Engine:** Works alongside the UI to deliver insights, alerts, and feedback to parents and teachers. It integrates with the evaluation model outputs and analytics data to surface meaningful progress updates.

Question 7: What other resources did you use to help answer these questions this time?

- Collaboration with peers in the MSDS cohort
- Online resources on EdTech best practices and FERPA compliance
- Help from ChatGPT for brainstorming risks and metrics

Question 8: Reflect on how you answered each question in Part One when you were working solo and compare it to Part Two. For each question 1-6, write down one thing you learned by answering the question again with assistance and resources. For example: "I learned about the existence of metric X, and that the metric I wrote down in Part One is actually not that useful for this problem."

- **Question 1:** I learned to think more broadly about the product impact—not just on students but also on stakeholders like parents and teachers.
- **Question 2:** I realized that some goals, like "engagement," need concrete, measurable metrics to be actionable.
- **Question 3:** I gained a deeper appreciation for the regulatory side of ML systems and the complexity of handling children's data.
- **Question 4:** I learned to consider systemic risks like bias and curriculum misalignment, not just model-level risks.
- **Question 5:** I learned that risk mitigation often includes both technical (e.g., audits) and organizational (e.g., content review processes) strategies.
- **Question 6:** I learned that having a modular and scalable architecture is crucial, especially for products aiming to expand across subjects and grades.

Turning it in

Please type your name at the top of the first page, save as ***pdf***, and submit to Canvas.