Your Name: Iris Yu

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.

- Enable personalized study comprehension content recommendation based on student interests to increase engagement and motivation.
- Track and support the development of metacognitive reading strategies (e.g., summarizing, predicting) through in-platform activities and prompts**.**

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

- Personalized content engagement:
  - Metric: Click-through rate (CTR) and completion rate of recommended reading materials
  - Supplementary: Increase in average time spent on reading tasks after personalization.
- Metacognitive strategy development:
  - Metric: Improvement in rubric-based evaluation of strategy application (e.g., number of successful summarizations per activity).
  - Supplementary: Frequency and correctness of self-explanation prompts completed

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:
  - Ensure compliance with FERPA by anonymizing.
  - Secure storage using encrypted databases and role-based access control.
  - Obtain explicit parental/guardian consent for data usage.
- Data Quality:
  - Implement validation rules for input scores (e.g., 0–100 for grades, timestamps for assessments).
  - Regular audits to detect and correct missing, inconsistent, or duplicated entries

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

- Over-personalization may cause students to be exposed only to materials that match their current interests, limiting their exposure to diverse topics or challenging content.
  - Impact: Reduced development of well-rounded reading skills and critical thinking.
- Teacher dashboard overload due to excessive data visualization or non-actionable insights.
  - Impact: Teachers may ignore the dashboard or make suboptimal instructional decisions due to confusion or information fatigue.

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

- Over-personalization:
  - Allow teachers to override and adjust recommended content manually.
- Teacher dashboard overload:
  - Use flashcards and visual alerts instead of raw data tables to guide attention effectively.

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.

- Recommendation Engine: A component using collaborative and content-based filtering to suggest reading materials. It pulls data from the real-time database and student profile, and outputs to the UI.
- Student Profile Management: Stores student preferences, past performance, learning style, and interaction history. Feeds into both the ML model and recommendation engine.

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

- Course slides on MLOps and data governance
- Notes from class discussion on education data risks

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."

1. Goal definition: I learned how to design goals that are both educationally meaningful and measurable using behavior-based metrics.
2. Metrics: I realized that qualitative strategies like metacognition can still be quantified using rubrics and frequency tracking.
3. Data governance: I learned more about FERPA and the need for encryption and pseudonymization, not just permission.
4. Risks: I now see how product design can accidentally narrow learning if diversity isn't baked into recommendations.
5. Mitigation strategies: I learned that co-design with stakeholders (like teachers) is essential for usability and adoption.
6. Architecture: I discovered how modular design allows for feedback loops and personalized interactions while keeping scalability in mind.

Turning it in

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