Your Name: Jose Fuentes

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 into 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.

- 3) Identify when students are at risk of falling behind or are struggling with a particular concept in reading comprehension for effective intervention and provide necessary resources to help the students not fall behind
- 4) peer to peer collaboration between students where the students benefit from each other and complement one another

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

- 3) time between identifying a students struggles and implementation in countering those struggles (shorter time is better) as well as doing some experiments where we check improvement rate of students that did have that early intervention with students who without early intervention
- 4) both students that get paired up see an increase in score in various activities and test exams in subjects/topics that they once struggled in and feedback from the students themselves

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.

- For Tests/Assessments
    - us IDs to make students anonymous, using that as an alternative
    - implementing specific roles/permissions, so teachers would be able to access the students data
    - review data retention policies (deletion of that students data after 2 years for example)
- For student feedback
    - when given feedback, we won't know who wrote it, so more anonymity
    - can opt out/in for feedback collection and other metrics of data retention
- audio or visual data
    - get the consent of the parents and outline clearly what the data is being used for and what is being collected
    - higher order of encryption

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

- 3) over-reliance on the professors and parents for using this AI platform in helping students as sometimes the AI can be incorrect at times and give false information.
- 4) Algorithm Bias in terms of hindering the students ability to improve on a topic. For example, if a student is struggling with finding themes for a particular section of a book,

the AI might assign similar problems over and over again instead of suggesting strategies or other activities to help improve the students ability

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

- 3) can implement confidence ratings for each AI generated recommendations or feedback for the teachers. We can also have teachers check what the output of the AI is before sending it to the students and also have a flagging system so the teachers can flag bad outputs
- 4) ensure that the AI has a variety of instructional methods, activities so it doesn't repeat. Can also have a teacher override approach where they can assign new problems if they notice something strange or repetition is too much.

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.

- Feature Store component: a good way to manage features during our model training and inference
- some type of monitoring in the pipeline
- a dashboard or visual layer of some sort for teachers and parents

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

- i used AI and our notes in class to help me with the assignment and going over the transportation example we did in class

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."

- Question1: I identified another goal of the project which is more collaboration between students and teachers, something I neglected in my previous responses
- Question2: I learned about a new metric such as identifying the time when a student struggles with a topic and I probably didn't need my original metric of using students' test scores. Maybe redefining the scores/points system in evaluating students progress
- Question3: I went in a bit more detail about how to protect students data by making their ids anonymous and giving more permission to the teachers in seeing their data as well as on/off for data collection
- Question4: I ignored the fact that teachers can over-rely on the platform just as much as the students and maybe take their job less seriously or do less work.
- Question5: I included more feedback from my original design which incorporates the teachers more by rating the AI responses and having more human feedback in the pipeline for those generations

- Question6: I learned about feature storing which is something I forgot to mention in my original architecture design for model creation and importance

Turning it in

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