Your Name: Grant Nitta

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.
Improve reading student reading literacy
Improve student critical thinking abilities in parallel with literacy and comprehension.

Question 2: For each additional goal from Question 1; define a metric to measure success of that goal.
Literacy: Reading Rate (words per minute) and Vocabulary acquisition rate.
Critical Thinking: A common way to measure critical thinking is the Watson-Glaser Critical Thinking Appraisal.

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 wise all when using student, teacher, and historical school records it is important to anonymize the data so it cannot be tracked back to specific students. For student data it is important only necessary data is collected (just enough for measuring the metrics) It is also important to have a clearly defined time frame for how long student records are kept. For teacher records on how their students performed it would be the same as the students. For historical school performance If the data is anonymized enough where there are no identifiers it would depend on the district and their data rules.
For Data quality standardized tests are the typical way to gather student performance data, but they are not always the best way to collect data due to different learning and testing styles of students, so standardized tests shouldn't be the be all end all for how a student is performing. For student performance data both tests, projects and homework should be examined to get a better image of how a student is learning. Even this way leaves room for positive bias from students cheating, or negative bias from a student not participating even if they are absorbing content.
For Historical school performance there is the same worry about quality as there is for measuring students, but additional concerns is how performance is being measured year over year, if situations change year over year the relatability between years could be lost for student performance.

Question 4: Identify an additional two risks associated with this product and the potential impact of each risk.
Data security/Privacy for students/minors: Storing any data has the risk of being leaked or having unauthorized access and there is always a risk for data leakage.
Algorithmic Bias: Having the model trained on certain types of learners or being biased against specific demographics based on historical performance.

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

Algorithmic Bias: In building the algorithm train it on a wide array of data points (students) to try to reduce bias towards any one group. (this should be in both location (school), demographic, and performance to name a few), Have regular audits to ensure the algorithm is performing as intended across all demographics.

Data security/Privacy for students/minors: Ensure whatever third-party data storage provider is SOCII compliant to reduce the risk of data breaches. Implement MFA for all system accesses. Employ comprehensive encryption of data, and have role-based access for all users.

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.

Architectural portions that were missing from part 1 are, a student profile service, to aggregate the raw student data into a useable format, Content recommendation engine, is a model that predicts what content is best for the student given their scores. A difficulty adjustment engine to adjust content difficulty based on the student's score, Dashboard for parents/educators, and a Privacy control/access layer to ensure privacy of data.

The student profile service will connect to both the Database and the data cleaning sections from part 1, the model mentioned in part 1 would include the content recommendation system and difficulty adjustment models, while the deployment phase from part 1 would be the dashboard for parents and educators.

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

Used a mix of Google, and Claude.

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. Google helped me distinguish between what Literacy, comprehension and critical thinking was.
2. I learned about new ways to measure literacy, comprehension and critical thinking and I learned more about different tests/ways critical thinking is specifically measured both in an educational setting and for jobs/ business.
3. I didn't learn anything new but having time to sit down and think as well as get a clearer explanation of data governance I was able to have a more comprehensive answer then I did in part 1.
4. Having more access to resources allowed me to get different views on risks associated with collecting student data giving me a better answer then I did in part 1.
5. Having access to google allowed for different points of view and allowed for a better more comprehensive answer that is more thought out then what appeared in part1.
6. For data architecture I learned more about what was part of data architecture then my very very high level overview that I described in part1, Initially I thought it was the key components at a high level not specifically what each model/system that is being used.

Turning it in

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