

**Your Name: Venkatachalam Subramanian Periya Subbu**

## **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.**

Ans :

Goal 3: Maintain high student engagement throughout the learning experience.

Goal 4: Maximize the accuracy of personalized learning recommendations.

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

Ans:

Metric 3: Average session duration

Metric 4: Precision and recall of recommended learning activities

**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.**

Ans:

- For **Online Libraries** and **Question Banks**, we must ensure the content is properly licensed, high-quality, and free of bias.
- For **Student Platform Product Data**, strict FERPA compliance is needed, which could be done by encrypting all student data, anonymize personal information, and control access.

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

Ans:

Risk 1: Model Bias Toward Certain Demographics. The ML model may perform better for students from specific backgrounds. For instance, recommendations and model performance for native English speakers might be different to non-native English speakers for students from same school level. This could widen achievement gaps, reduce fairness, and lead to dissatisfaction among students, parents, and educators.

Risk 2: The model might learn patterns specific to the format of standardized tests and not generalize well to different styles of comprehension questions. Students may perform well on platform assessments but poorly in real-world testing environments.

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

Ans:

1. Regularly audit model performance across different demographic groups, apply fairness constraints during model training, and augment training datasets with more diverse examples to balance performance.

2. Train the model on a wide variety of question types, sources, and assessment styles. Include validation and generalization tests to ensure that the model performs well on unseen formats.

**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.**

Ans:

1. Model monitoring
2. Dashboard for parents and teachers

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

Ans:

1. Classmates
2. GenAI

**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."**

Ans:

For question 1: In part 1, I only focused on system goals and not business goals (student outcomes and product outcomes)

For question 2: In part1, my answers were generic to metrics, in part2 I was able to be more specific (precision and recall)

For question 3: I initially thought of data privacy mostly as "keep it safe,". In part 2 I focused more on "accuracy" of the data.

For question 4: In Part One, I focused only on technical failures. In part2, I considered even the potential risks of being bias and poor student performance due to poor generalization, focusing on more business aspects.

For question 5: In Part One, I wasn't thinking about active mitigation. I learned that auditing models, diversifying datasets, and generalization testing are practical ways to reduce these risks.

For question 6: I had missed out on 'Model Monitoring' and 'Dashboard'.

## **Turning it in**

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