

Your Name: Serigne Diaw

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.

1. Provide real-time, actionable feedback to educators and parents.

The system should deliver insights on student progress, strengths, and areas of improvement, allowing teachers and parents to intervene or support when needed. These insights should be provided through an interpretable dashboard.

2. Enable seamless integration with existing school information systems and assessment tools.

The platform should ingest performance data from standardized tests or district tools to create a learning profile for each student.

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

1. For the real-time feedback, a metric will be the average time from student activity completion to dashboard update.
2. For the integration with school information systems and assessment tools, the metric will be the percentage of successful data sync events that do not require manual intervention.

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 the eyetracking data, recordings might be classified as personally identifiable information under laws like FERPA. They also may require consent, especially for young children. As far as data quality, we will need to verify that the systems are properly calibrated and that recordings are synchronized with learning content for inference.

For the reading comprehension accuracy, we will need to treat performance data as protected records and be transparent about how the data is used. We will also need to standardize how accuracy is measured across assessments and include relevant metadata like time spent, question difficulty, etc.

Finally, for the test scores, we will need to restrict access to authorized parties and ensure data-sharing agreements are in place. We also need to handle re-tests and updating scores.

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

1. One additional risk is that the ML model will reinforce biases due to training data imbalances (eg. socioeconomic status, English language proficiency). The impact of this is that students may be steered into easier learning paths if they fall into a certain demographic group, limiting their growth.
2. Another potential risk is an overdependence on AI over human judgment, where educators may overly rely on AI-generated insights without considering professional judgment. The impact of this is that students may receive inappropriate interventions or receive less human support.

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

1. To mitigate the first risk, we can curate datasets that reflect diverse backgrounds, including students from all demographic groups. We can also run routine fairness checks to make sure the product doesn't have any biases.
2. We can frame the product as a co-pilot and emphasize that it should not be relied upon by itself. We can also display confidence levels with each recommendation.

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.

Firstly, I forgot to include a data ingestion layer to pull in things like scores and eye-tracking feeds. We will also need an ETL pipeline to clean and transform the data before uploading it to the database. We can use a feature store to store reusable features like reading speed, test scores, etc. We will also need an integration layer (API) to connect the user interface with external systems such as the standardized test databases. We will also need a security layer to manage encryption and role-based access. We need a model deployment component to serve the inference.

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

I used ChatGPT and Perplexity AI.

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 did not think about the real-time element of this product in my original response. With other resources, I learned that this should be emphasized and is an important consideration for the architecture of the system.

Question 2. I refined metrics to be specific and measurable using my resources, such as time from activity completion to displaying in the UI.

Question 3. With resources, I gained a more comprehensive view of data privacy. For example, I learned about the specific laws such as FERPA and COPPA, such as the fact that children under the age of 13 are subject to specific rules regarding personally identifiable data. My resources led me to emphasize the need for encryption.

Question 4. As far as risks and impacts, I added some key concerns regarding the ML algorithm that I hadn't anticipated. This includes algorithmic bias and an overreliance on AI. We need to be careful with this product because an AI algorithm can not be completely trusted to provide an all-inclusive learning environment for students.

Question 5. I originally suggested that we just delete the eyetracking data after use, but I didn't anticipate the privacy concerns while the data is actually in the system. With resources, I learned that encryption is key so that the data can not identify students while it is in the system.

Question 6. I originally described a high-level architecture, but I missed several key components such as feature stores, and real-time inference components.

Turning it in

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