Chelyah Miller
9 May 2025

MSDS 603 MLOps Assignment 1, Part 2

Question 1: Define an additional two goals for this project.

1. Detect and adapt to different learning styles
2. Provide actionable insights to teachers and parents

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

1. I could use user engagement patterns such as time spent on interactive vs. textual content and videos vs. reading to infer preferred modalities.
2. I could aggregate student progress into dashboards and alert educators to students who are struggling, excelling or disengaging.

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.

1. Standardized Test Scores
    a. Considerations: The students' scores have to be restricted to authorized parents so it would have to be aggregated so that the individual scores aren't known. Additionally, it would be necessary to ensure consistency across districts so that there could be accurate comparisons which might be difficult. It's also difficult to determine the best way to handle missing/incomplete data so that we're not adding bias when analyzing.
2. Engagement data from other educational tools
    a. Considerations: There would have to be clear consent mechanisms and limitations on third-party data sharing since engagement tracking may involve user behavior and these would be minors primarily. There would also have to be clear metrics for measuring engagement and false positives would have to be monitored.
3. Student socioeconomic status (SES)
    a. Considerations: SES data is highly sensitive so it would have to be anonymized or aggregated wherever possible and there would again have to be clear consent mechanisms in place. The data would have to be self-reported which could lead to inaccuracies or bias.

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

1. Educators might start to overly rely on the technology or the technology's recommendations without using their own judgment which could result in a one-size-fits-all style that disregards students' needs.

2. There might be algorithmic bias if the model is trained on biased data which reinforces existing inequalities and could limit opportunities for students from marginalized communities.

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

1. Provide onboarding so teachers learn how to integrate the recommendations into their professional strategies.
2. Use datasets that reflect a wide range of student backgrounds and regularly test the model for biased patterns such as consistently lower recommendations for certain groups.

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.

1. Servers to run the AI models
2. Databases for storing user profiles and results
3. Authentication and security to protect student data
4. Front end interface that students, teachers, and parents interact with

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

1. I used ChatGPT and Claude to answer my questions this time

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. I didn't think about the application having the ability to determine different learning pathways which it definitely could. This got rid of one of my concerns immediately.
2. I kept thinking about having to use outside tools to determine best tools of engagement but just observing how the user engages would be enough to make inferences.
3. There are so many considerations that have to be made when dealing with data from minors. I think I underestimated the difficulty of having to ensure everything is consistent and aggregated while also trying not to introduce bias.
4. While I thought of educators overly relying on the data (in a different way), I didn't think about the algorithmic bias that could occur due to existing inequalities in education/education data and how that would affect the model. This is one of the aspects that I think would be significantly harder to tackle.

5. While I really like the idea of providing professional education to the educators, I also think about the difficulty of integrating this in a way that doesn't lead to over reliance, especially when facing limited resources and burn-out. I'm considering the benefits of regular audits by a principal/members of the school district to manage this.
6. I severely under thought this question because basic aspects like security, servers, and databases are a key component of this and I forgot to list them initially.