

Part 4: Reflection & Workflow Diagram (10 points)

Reflection (5 points)

1. Most challenging part of the workflow & why

The most challenging part of the workflow was the data strategy and preprocessing phase.

This stage required understanding what data sources the hospital would realistically collect, identifying missing or inconsistent data, ensuring privacy compliance, and designing appropriate feature engineering steps. Healthcare data is often sensitive, noisy, and unstructured, making it difficult to create a clean, fair, and representative dataset. Additionally, balancing ethical considerations while still building a useful model required thoughtful decision-making.

2. How I would improve my approach with more time/resources

With more time and resources, I would improve the project by:

- Collecting a larger and more diverse dataset to reduce bias and improve generalization.
- Using automated feature engineering tools (e.g., FeatureTools) to extract richer clinical features.
- Collaborating with medical professionals to better understand domain-specific variables and ensure the model reflects real clinical needs.
- Testing multiple advanced models (XGBoost, LightGBM, calibrated models) and comparing performance.
- Conducting a deeper fairness analysis using tools like AIF360 to identify disparities and ensure equitable predictions.

Diagram (5 points)

AI Development Workflow

