Part 3: Critical Thinking (20 points)

1. **Ethics & Bias (10 points):**

   - How might biased training data affect patient outcomes in the case study?

If the training data underrepresents certain patient groups, e.g. ethnic minorities, low-income patients, the model may systematically mispredict their risk. For example, insufficient samples for a subgroup can cause “suboptimal performance” or underestimation of risk. Left unchecked, such bias “can lead to substandard clinical decisions and the perpetuation and exacerbation of longstanding healthcare disparities”. In practice, this could mean that some patients who actually have high readmission risk are overlooked, widening care gaps.

   - Suggest **1 strategy** to mitigate this bias.

To counter the ethics and bias, we would ensure *data representativeness* and fairness. This could involve collecting a larger, more diverse dataset (oversampling underrepresented subgroups or gathering new sources) and applying bias-aware techniques like re-weighting or fairness constraints. As recommended in the literature, “collection of large and diverse data sets” and using statistical debiasing methods can mitigate biases. Regularly monitoring model performance across demographic groups using fairness metrics also helps detect and correct emerging biases.

2. **Trade-offs (10 points):**

   - Discuss the trade-off between model interpretability and accuracy in healthcare.

Interpretability vs. accuracy: More complex models, e.g. deep neural networks or ensembles often capture subtle patterns and achieve higher accuracy, but they act as “black boxes.” Simpler models like logistic regression or decision trees are easier for clinicians to understand

and justify, but may miss some nuances. In healthcare, this trade-off is critical: when a wrong decision can cause serious harm, higher transparency is often prioritized, even if it costs some accuracy. In practice, clinicians may prefer a slightly less accurate but explainable model so they can understand **why** a patient is flagged as high-risk.

- If the hospital has limited computational resources, how might this impact model choice?

Computational constraints: With limited compute resources e.g. no powerful GPU or on-device processing, the hospital may need to use more efficient models. Achieving “high accuracy” usually requires complex models and lots of data, which increases computational demands. In a low-resource setting, one might choose a lightweight model like a shallow decision tree or logistic regression or use techniques like model pruning and quantization to reduce size. The trade-off is that extreme efficiency could slightly reduce accuracy, but it ensures the system runs reliably on available hardware.

Part 4: Reflection & Workflow Diagram (10 points)

1. **Reflection (5 points):**

   - What was the most challenging part of the workflow? Why?

Most challenging part: Balancing these competing priorities was toughest. For example, ensuring fairness and interpretability while still meeting accuracy targets is inherently difficult, especially given limited patient data. Data privacy and compliance e.g. HIPAA add complexity too, because they restrict data usage. In short, navigating ethical trade-offs: bias vs. performance, privacy vs. data richness requires careful consideration at each stage.

- How would you improve your approach with more time/resources?

Improvements with more time/resources: Given more data and stakeholder input, we would iterate further. For instance, we could collect additional patient records or include broader demographics to reduce bias. We'd also allocate time for thorough validation (possibly a pilot study) and involve clinicians to refine feature engineering. With extra compute or data access, we could experiment with advanced techniques (e.g. ensemble models with explainability tools, or federated learning to preserve privacy). Such enhancements would likely improve both accuracy and fairness of the model.

2. **Diagram (5 points):**
   - Sketch a flowchart of the AI Development Workflow, labeling all stages.