AI Development Workflow Assignment

Part 3: Critical Thinking (20 points)

Ethics & Bias (10 points)

Biased training data—such as underrepresentation of certain age groups, ethnicities, or income levels—can lead to unequal prediction outcomes. For example, the model may incorrectly estimate readmission risk for minorities or low-income patients, potentially causing unfair treatment or missed interventions.

To mitigate this bias:

- • Implement stratified sampling to balance representation.
- • Perform fairness audits using tools like IBM AI Fairness 360.
- • Remove or balance sensitive features such as race, age, gender where appropriate.

Trade-offs (10 points)

In healthcare, there is a crucial trade-off between model interpretability and accuracy. Highly accurate models (e.g., deep neural networks) may be black-boxes, which is risky in clinical decision-making where explainability is essential. Simpler models (e.g., decision trees) offer interpretability but might lack predictive power.

With limited computational resources (e.g., in smaller hospitals):

- • Avoid computationally expensive models like deep learning.
- • Prefer lightweight models such as Logistic Regression or Decision Trees.
- • Use batch inference rather than real-time predictions to conserve resources.