🧠 Part 3: Critical Thinking (20 points)

1. Ethics & Bias (10 points)

How might biased training data affect patient outcomes?
If the training data overrepresents certain demographics (e.g., older patients, specific races, or insured individuals), the model could learn patterns that favor those groups while underperforming for others. This can result in:

- Underdiagnosis or overdiagnosis for underrepresented groups
- Unfair resource allocation
- Worsening of existing healthcare inequalities

1 Strategy to Mitigate Bias:
☑ **Balanced Sampling & Fairness Audits**

- Use techniques like SMOTE or stratified sampling to balance the dataset across sensitive features (e.g., race, gender, insurance status)
- Conduct fairness audits by comparing model performance across subgroups (e.g., AUC for males vs females)
- Regularly retrain the model with updated, diversified data

2. Trade-offs (10 points)

Interpretability vs Accuracy in Healthcare

- **High Accuracy Models** (e.g., deep neural nets): Better predictive power but hard to explain → risky in healthcare where **transparency** is critical.
- **Interpretable Models** (e.g., decision trees, logistic regression): Easier for doctors to trust but may lack precision.

🧠 *In healthcare, interpretability often wins.* A slightly less accurate model that can be explained is more ethical and safer.

What if the hospital has limited computational resources?

- Avoid heavy models like large deep learning architectures
- Use lightweight models like **Logistic Regression, Decision Trees, or Random Forests** with fewer trees
- Optimize features beforehand (dimensionality reduction, feature selection)
- Deploy using efficient frameworks like **ONNX** or **TensorFlow Lite**