1. Ethics & Bias (10 points)

How Biased Training Data Affects Patient Outcomes

Biased training data can lead to **unfair or harmful predictions**, particularly in healthcare where disparities already exist. Potential impacts include:

1.**Underdiagnosis of High-Risk Groups**

•If historical data underrepresents certain demographics (e.g., minority populations, low-income patients), the model may **fail to flag them as high-risk**, leading to inadequate care.
•**Example:** A model trained mostly on affluent patients may underestimate readmission risk for homeless individuals due to missing social determinant factors.

2.**Over-prediction for Certain Groups**

•If a demographic group has historically had higher readmissions due to systemic barriers (e.g., lack of transportation), the model may **unfairly label all similar patients as high-risk**, leading to unnecessary interventions.
•**Example:** A model associating "Medicaid patients" with higher readmissions might bias clinicians against them, even if individual cases don't warrant it.

3.**Reinforcing Existing Disparities**

•If biased predictions lead to **diverting resources away from some groups**, it could worsen health inequities.

Strategy to Mitigate Bias: Adversarial Debiasing

•**How it works:**

•Train the model to **predict readmission risk while simultaneously minimizing its ability to predict protected attributes** (e.g., race, insurance status).
•Uses an adversarial network to penalize the model for learning biased patterns.
•**Advantages:**

•Does not require removing useful features (e.g., ZIP code can still be used for social risk, but not for racial bias).
•Actively reduces discrimination rather than just detecting it.

Alternative Strategies:

•**Reweighting:** Assign higher importance to underrepresented groups during training.
•**Fairness Constraints:** Enforce statistical parity (equal readmission prediction rates across groups).

---

2. Trade-offs (10 points)

Trade-off: Model Interpretability vs. Accuracy

In healthcare, **interpretability is often prioritized over pure accuracy** because:

High-Accuracy, Low-Interpretability Models (e.g., Deep Learning)	Interpretable Models (e.g., Logistic Regression, Decision Trees)
✔ May achieve **slightly better AUC/accuracy** (e.g., 92% vs. 88%).	✔ Doctors can **understand and trust** predictions (e.g., "Patient flagged due to diabetes + prior admission").
✗ **Black-box nature** makes clinicians skeptical.	✗ May **sacrifice some predictive power** for simplicity.
✗ Hard to explain to regulators (e.g., FDA, HIPAA compliance).	✔ Easier to **debug and audit for bias**.

Best Compromise:

•Use **XGBoost/LightGBM** (balance of accuracy + interpretability via feature importance).
•Supplement with **SHAP values/LIME** to explain individual predictions.

Impact of Limited Computational Resources on Model Choice

If the hospital lacks high-performance GPUs/cloud infrastructure:

1.**Simpler Models Become Necessary**

•**Logistic Regression** or **Random Forests** may replace deep learning.
•Pros: Lower compute needs, easier to deploy on-premise servers.

•Cons: Potentially lower accuracy for complex patterns.

2.**Reduced Feature Complexity**

•Fewer features (e.g., drop NLP-extracted notes from discharge summaries).

•Use **PCA** or **feature selection** to reduce dimensionality.

3.**Batch Processing vs. Real-Time Predictions**

•If real-time inference is too costly, switch to **nightly batch predictions**.

4.**Edge Deployment**

•Run lightweight models (e.g., **TensorFlow Lite**) on local hospital servers instead of cloud APIs.

Example Workaround:

•Train an **XGBoost model** on a subset of high-impact features (e.g., comorbidities, prior admissions) instead of a full EHR dataset.