Problem Scope (5 points)

- **Problem**: Hospital readmissions within 30 days of discharge burden healthcare systems and may indicate poor care outcomes.
- **Objective**: Build an AI model to predict patient readmission risk, enabling proactive interventions.
- **Stakeholders**: Hospital administrators, physicians, patients, IT staff, and regulatory bodies.

Data Strategy (10 points)

Proposed Data Sources

- **Electronic Health Records (EHRs)**: Clinical history, discharge summaries, diagnoses, medications.
- **Demographics**: Age, gender, socio-economic status.
- **Utilization Data**: Previous admissions, outpatient visits, emergency visits.
- **Social Determinants of Health (SDOH)**: Housing, employment, access to care.

Ethical Concerns

1. **Patient Privacy**: Data must be de-identified and securely stored to avoid breaches.
2. **Bias & Fairness**: Ensure the model doesn't unfairly penalize groups based on race, income, or geography.

🔧 **Preprocessing Pipeline**

1. **Data Cleaning**: Handle missing values, remove duplicates.
2. **Normalization**: Scale numerical features (e.g., age, lab results).
3. **Categorical Encoding**: One-hot encoding for diagnosis codes and discharge types.
4. **Feature Engineering**:
    - Compute **length of stay**, **number of prior admissions**, **time since last visit**.
    - Extract **textual features** from discharge summaries using NLP (e.g., TF-IDF or embeddings).
5. **Split Dataset**: Train/test (e.g., 80/20), stratified by readmission status.

Model Development (10 points)

Model Choice

- **Gradient Boosting (e.g., XGBoost)**: Handles mixed data types well, offers high performance and interpretability (via SHAP).

Hypothetical Confusion Matrix

	Predicted: No Readmit	**Predicted: Readmit**
Actual: No	800	200
Actual: Yes	100	400

- **Precision** (Readmit) = 400 / (400 + 200) = **0.67**
- **Recall** (Readmit) = 400 / (400 + 100) = **0.80**

Deployment (10 points)

🛠 Integration Steps

1. **API Development**: Wrap model in RESTful API for hospital systems.
2. **EHR Integration**: Connect to clinical dashboards (e.g., Epic, Cerner).
3. **Alerting Mechanism**: Notify staff of high-risk patients' post-discharge.
4. **Feedback Loop**: Monitor model performance and retrain periodically.

Compliance

- **HIPAA Alignment**:
    - Encrypt data at rest and in transit.
    - Role-based access control.
    - Log audit trails for model usage.
- Conduct regular **risk assessments** and **Data Protection Impact Assessments (DPIA)**.

Optimization (5 points)

Overfitting Mitigation

- **Early Stopping**: Halt training when validation loss starts increasing, preventing the model from memorizing noise.