Part 1: Short Answer Questions (30 points)

1. Problem Definition (6 points)

Hypothetical AI problem: Predicting property price trends in a local real estate market.

Objectives:

- Forecast future property prices based on historical data.
- Identify factors influencing price fluctuations.
- Provide pricing recommendations for buyers and sellers.

Stakeholders:

- Real estate agents.
- Property investors.

Key Performance Indicator (KPI):

Mean Absolute Error (MAE) between predicted and actual property prices to measure prediction accuracy.

2. Data Collection & Preprocessing (8 points)

Data sources:

- Zillow API for property listings and price history.
- Public property records from local government databases.

Potential bias:

Data may be biased toward properties listed online, excluding off-market or private sales, which can skew price trends.

Preprocessing steps:

- Handle missing values by imputing or removing incomplete records.
- Normalize numerical features like property size and price.
- Encode categorical variables such as property type or neighborhood.

3. Model Development (8 points)

Model choice:

Random Forest, because it handles nonlinear relationships well, is robust to outliers, and interpretable for real estate features.

Data splitting:

Split data into 70% training, 15% validation, and 15% test sets to train the model, tune hyperparameters, and evaluate performance fairly.

Hyperparameters to tune:

- Number of trees (to balance bias and variance).
- Maximum tree depth (to prevent overfitting).

4. Evaluation & Deployment (8 points)

Evaluation metrics:

- Mean Absolute Error (MAE) to measure average prediction error in price units.
- R-squared (R^2) to assess how well the model explains price variability.

Concept drift:

Changes in market conditions or economic factors over time that affect property prices differently from training data. Monitor it by regularly comparing prediction errors on new data and retraining the model as needed.

Technical challenge:

Scalability—handling large volumes of real-time property data and delivering fast predictions to users without delays.

Part 2: Case Study Application (40 points)

Scenario: A hospital wants an AI system to predict patient readmission risk within 30 days of discharge.

Problem Scope (5 points)

- **Problem:** To accurately predict which patients are at high risk of readmission within 30 days of discharge, enabling targeted interventions to improve patient outcomes and reduce healthcare costs.
- **Objectives:**
    - Reduce 30-day patient readmission rates.
    - Improve patient care coordination and post-discharge planning.
    - Optimize hospital resource allocation.
- **Stakeholders:**
    - **Patients:** Benefit from improved health outcomes and reduced readmissions.
    - **Hospital Administration:** Aims to reduce costs, improve quality metrics, and optimize resource use.
    - **Clinicians (Doctors, Nurses):** Use predictions to guide care decisions and interventions.
    - **Data Scientists/AI Developers:** Responsible for building and maintaining the system.
    - **Regulators:** Ensure compliance with healthcare data privacy and safety standards.

Data Strategy (10 points)

- **Proposed Data Sources:**
    - **Electronic Health Records (EHRs):** Patient demographics, diagnoses (ICD-10 codes), procedures (CPT codes), medications, lab results, vital signs, discharge summaries, clinical notes.

- **Demographics:** Age, gender, socioeconomic status, insurance type, residential address (for social determinants of health).
- **Past Readmission History:** Previous hospitalizations and readmission events.
- **Social Determinants of Health (SDOH) Data:** Information on housing stability, food security, access to transportation, and social support (if available and ethically sourced).

- **Ethical Concerns:**

1. **Patient Privacy and Data Security (HIPAA Compliance):** Ensuring that sensitive patient health information (PHI) is protected from unauthorized access, use, or disclosure. This includes anonymization/de-identification techniques.

2. **Algorithmic Bias and Fairness:** The model might inadvertently discriminate against certain demographic groups (e.g., based on race, socioeconomic status) if the training data is unrepresentative or contains historical biases, leading to unequal access to interventions.

- **Preprocessing Pipeline:**

1. **Data Cleaning:**
    - Handle missing values (e.g., imputation for lab results, dropping features with excessive missingness).
    - Correct inconsistencies and errors (e.g., typos in diagnoses, out-of-range values).
    - Remove duplicate records.
2. **Data Transformation:**
    - **Categorical Encoding:** One-hot encoding for nominal variables (e.g., gender, insurance type), ordinal encoding for ordinal variables (e.g., severity scores).
    - **Numerical Scaling:** Standardization or normalization for numerical features (e.g., lab values, age) to ensure features contribute equally to the model.
3. **Feature Engineering:**

        - **Comorbidity Scores:** Calculate a Charlson Comorbidity Index or Elixhauser Comorbidity Index based on patient diagnoses to quantify overall health burden.
        - **Length of Stay (LOS):** Calculate the duration of the current hospitalization.
        - **Medication Adherence Indicators:** Create features from prescription refill history or medication reconciliation at discharge.
        - **Frequency of Past Hospitalizations:** Number of admissions in the last 6 months or year.
        - **Time Since Last Admission:** Days since the most recent previous hospitalization.
        - **Discharge Disposition:** Categorical feature indicating where the patient is discharged to (e.g., home, skilled nursing facility, hospice).

Model Development (10 points)

- **Model Selection and Justification:**
    - **Model:** Gradient Boosting Machine (e.g., XGBoost or LightGBM).
    - **Justification:**
        - **High Performance:** GBMs are known for their strong predictive accuracy on tabular data, often outperforming simpler models.
        - **Handles Mixed Data Types:** Can naturally handle both numerical and categorical features.
        - **Feature Importance:** Provides insights into which factors are most influential in predicting readmission, aiding clinical interpretability.
        - **Robustness to Outliers:** Less sensitive to outliers compared to some other models.
        - **Scalability:** Efficient for large datasets, which are common in healthcare.
- **Confusion Matrix and Precision/Recall (Hypothetical Data):**
    - Let's assume our model predicts readmission for 100 patients.

- **Confusion Matrix:** | | Predicted Readmission | Predicted No Readmission | | :---------- | :---------- | :---------- | | **Actual Readmission** | True Positives (TP) = 20 | False Negatives (FN) = 5 | | **Actual No Readmission** | False Positives (FP) = 10 | True Negatives (TN) = 65 |
- **Calculation:**
  - **Precision:** Of all patients predicted to be readmitted, what proportion actually were? Precision=TP+FPTP=20+1020=3020≈0.67 (or 67%) *Interpretation:* When the model predicts readmission, it is correct 67% of the time.
  - **Recall:** Of all patients who actually were readmitted, what proportion did the model correctly identify? Recall=TP+FNTP=20+520=2520=0.80 (or 80%) *Interpretation:* The model identifies 80% of all actual readmissions.

Deployment (10 points)

- **Integration Steps:**
  1. **API Development:** Create a RESTful API endpoint for the trained model, allowing other hospital systems to send patient data and receive readmission risk predictions.
  2. **Data Ingestion Pipeline:** Establish a secure and automated pipeline to extract relevant patient data from EHRs in near real-time or batch, transform it into the features required by the model, and feed it to the API.
  3. **Integration with Clinical Workflow:**
     - Display predictions within the EHR system (e.g., as a risk score on a patient's dashboard).
     - Trigger alerts or notifications to care managers or discharge planners for high-risk patients.
     - Integrate with existing patient management or care coordination platforms.
  4. **Monitoring and Alerting:** Implement a system to continuously monitor model performance (e.g., drift in predictions, data quality issues) and alert relevant teams if performance degrades.

5. **User Interface (UI) Development:** Develop user-friendly dashboards for clinicians to view risk scores, contributing factors, and track intervention effectiveness.

- **Ensuring Compliance with Healthcare Regulations (e.g., HIPAA):**
    1. **Data De-identification/Anonymization:** Implement robust techniques to de-identify PHI before it is used for model training, testing, or even prediction where possible, adhering to HIPAA's Safe Harbor or Expert Determination methods.
    2. **Access Control and Encryption:** Implement strict role-based access control (RBAC) to the AI system and underlying data. Ensure all data (at rest and in transit) is encrypted using industry-standard protocols.
    3. **Audit Trails:** Maintain comprehensive audit logs of all data access, model predictions, and system changes to demonstrate accountability and compliance.
    4. **Data Use Agreements (DUAs) and Business Associate Agreements (BAAs):** Establish formal agreements with all third-party vendors or internal departments involved in data handling to ensure they comply with HIPAA regulations.
    5. **Regular Security Audits and Penetration Testing:** Conduct periodic assessments to identify and remediate vulnerabilities in the system.
    6. **Transparency and Explainability:** While not directly a HIPAA requirement, providing explainable AI (XAI) insights can help clinicians understand predictions, which is crucial for ethical use and demonstrating due diligence in patient care.

Optimization (5 points)

- **Method to Address Overfitting:**
    - **Regularization:** Apply L1 (Lasso) or L2 (Ridge) regularization during model training. This adds a penalty to the model's loss function based on the magnitude of the coefficients (L2) or the absolute values of the coefficients (L1), discouraging overly complex models that fit the training data too closely. For tree-based models like GBMs, regularization can be applied by controlling parameters like max_depth, min_child_weight, subsample, colsample_bytree, and lambda/alpha (L1/L2 regularization terms).

Part 3: Critical Thinking (20 points)

Ethics & Bias (10 points)

- **How might biased training data affect patient outcomes in the case study?**
  Biased training data, often reflecting historical disparities in healthcare access and quality, can lead to the AI model making inaccurate or unfair predictions for certain patient groups. For instance:
    - If the training data disproportionately represents healthier or more privileged populations, the model might underpredict readmission risk for underserved communities (e.g., based on race, socioeconomic status, or geographic location). This could lead to these high-risk patients not receiving necessary post-discharge interventions, exacerbating health inequities.
    - Conversely, it might overpredict risk for certain groups, leading to unnecessary interventions or stigmatization.
    - Errors in data collection or missing data for specific demographics could also introduce bias, making the model less effective or even harmful for those groups.

- **Suggest 1 strategy to mitigate this bias. Fairness-Aware Data Collection and Preprocessing:**
    - **Strategy:**
      Actively seek to collect more representative data from diverse patient populations, ensuring that all demographic groups are adequately represented in the training dataset. During preprocessing, techniques like **resampling (oversampling minority classes or undersampling majority classes)**, **reweighing (assigning different weights to data points based on their group membership)**, or **adversarial debiasing** can be used to reduce the impact of existing biases in the data before model training. Additionally, **feature selection should carefully consider proxy variables** that might inadvertently carry sensitive attribute information (e.g., zip code acting as a proxy for race or income).

Trade-offs (10 points)

- **Discuss the trade-off between model interpretability and accuracy in healthcare.**
    - **Interpretability:** Refers to the extent to which humans can understand the reasoning behind a model's predictions. Simple models like linear regression or decision trees are highly interpretable, as their decision rules are transparent.
    - **Accuracy:** Refers to how well the model's predictions match the actual outcomes. Complex models, such as deep neural networks or ensemble methods like Gradient Boosting Machines (GBMs), often achieve higher accuracy by capturing intricate non-linear relationships in the data.
    - **Trade-off in Healthcare:** In healthcare, there's a significant tension. While highly accurate models are desirable for critical predictions like readmission risk, clinicians often need to understand *why* a prediction was made to trust the system, justify interventions, and explain decisions to patients. A "black box" model, even if highly accurate, might be resisted by medical professionals due to a lack of transparency and accountability. For instance, if a model predicts high readmission risk, a clinician needs to know *which factors* (e.g., specific comorbidities, recent lab values, medication non-adherence) contributed most to that prediction to formulate an effective care plan. This need for interpretability can sometimes mean sacrificing a small degree of predictive accuracy by choosing a slightly less complex but more transparent model, or by employing explainable AI (XAI) techniques on complex models.

- **If the hospital has limited computational resources, how might this impact model choice?**
  Limited computational resources (e.g., less powerful servers, limited GPU access, restricted cloud computing budget) would significantly impact model choice in several ways:
    - **Preference for Simpler Models:** The hospital would likely need to opt for simpler, less computationally intensive models. Instead of deep learning models or large ensemble methods like complex GBMs, they might choose logistic regression, simpler decision trees, or smaller random forests. These models require less processing power and memory for training and inference.
    - **Reduced Feature Engineering Complexity:** Extensive feature engineering, especially involving complex transformations or interactions, can be

computationally demanding. The hospital might need to limit the number and complexity of features used.

- **Smaller Data Subsets:** Training on the full dataset might be infeasible. The hospital might need to sample smaller subsets of data for training or use techniques that allow for incremental learning.
- **Batch vs. Real-time Inference:** Real-time predictions require more immediate computational power. With limited resources, the hospital might be restricted to batch predictions, where data is processed periodically (e.g., overnight) rather than on demand.
- **Impact on Model Performance:** There's a risk that simpler models, chosen due to resource constraints, might not capture all the nuances in the data, potentially leading to lower predictive accuracy compared to what could be achieved with more powerful resources.

Part 4: Reflection & Workflow Diagram (10 points)

Reflection (5 points)

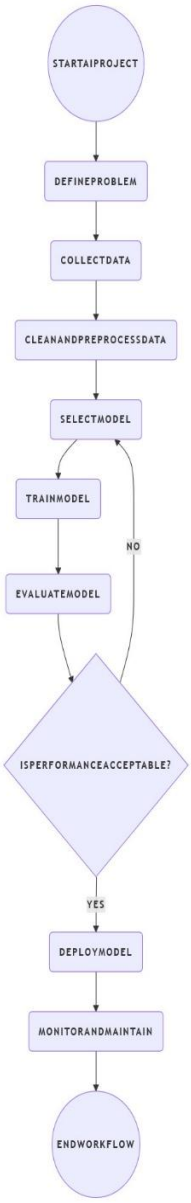
- **What was the most challenging part of the workflow? Why?**
  The most challenging part of this workflow would likely be Data Strategy, specifically addressing ethical concerns and ensuring data quality/representativeness.
    - **Why:**
      Healthcare data is inherently sensitive and complex. Ensuring patient privacy (HIPAA compliance) while simultaneously gathering enough diverse and high-quality data to build a robust and fair model is a significant hurdle.
      De-identification is crucial but can sometimes reduce data utility. Identifying and mitigating algorithmic bias requires deep understanding of both the data's limitations and the potential societal impact of the model's predictions.
      Furthermore, integrating data from disparate EHR systems, which often have varying formats and levels of completeness, adds substantial complexity to the preprocessing pipeline.
- **How would you improve your approach with more time/resources?**
  With more time and resources, I would significantly improve the approach by:
    - **Enhanced Data Governance and Collaboration:** Establish a dedicated data governance committee involving ethicists, legal experts, clinicians, and data scientists to continuously review data collection, usage, and model outputs for fairness and compliance.
    - **Advanced Bias Detection and Mitigation:** Implement more sophisticated fairness metrics and bias detection tools (e.g., Aequitas, Fairlearn) during model development. Explore advanced bias mitigation techniques like post-processing (e.g., equalized odds, demographic parity) to adjust model outputs for fairness without retraining.
    - **Robust Explainable AI (XAI):** Integrate more comprehensive XAI tools (e.g., SHAP, LIME) directly into the clinical workflow, allowing clinicians to query *why* a specific prediction was made for an individual patient, fostering trust and enabling better decision-making.
    - **Prospective Validation and A/B Testing:** Conduct rigorous prospective validation studies and A/B tests in a controlled clinical environment to assess the real-world

impact of the AI system on patient outcomes and resource utilization before full-scale deployment.

- **Continuous Learning and Feedback Loops:** Implement a robust system for continuous model retraining and updating based on new data and feedback from clinicians, ensuring the model remains accurate and relevant over time.

Diagram (5 points)

- **Sketch a flowchart of the AI Development Workflow, labeling all stages.**



Explanation of Stages:

- **Problem Definition & Scope:** Clearly define the clinical problem, objectives, and identify key stakeholders.
- **Data Collection & Ingestion:** Gather raw data from various sources (EHRs, demographics) and establish secure pipelines.
- **Data Preprocessing & Feature Engineering:** Clean, transform, and create new features from the raw data to prepare it for modeling.
- **Model Selection & Training:** Choose an appropriate AI model and train it using the prepared dataset.
- **Model Evaluation & Validation:** Assess the model's performance using metrics (precision, recall, etc.) on unseen data. Iterate on model architecture or parameters if performance is not satisfactory.
- **Bias & Ethical Review:** Continuously evaluate the model for fairness, bias, and adherence to ethical guidelines. This stage is iterative and informs all other stages.
- **Deployment & Integration:** Integrate the validated model into the hospital's existing IT infrastructure and clinical workflows.
- **Monitoring & Maintenance:** Continuously track the model's performance in a real-world setting, identify data drift or performance degradation, and manage updates.
- **Feedback Loops:** Crucial connections allowing insights from later stages (monitoring, user feedback) to inform earlier stages (data preprocessing, model re-training).