Part 1: Short Answer Questions (30 points)

1. Problem Definition (6 points)

Hypothetical AI Problem: Predicting employee turnover risk in a technology company

3 Objectives:

* Identify employees at high risk of leaving within the next 6 months with 80% accuracy
* Reduce overall employee turnover rate by 25% through targeted retention interventions
* Provide actionable insights to HR managers for proactive employee engagement strategies

2 Stakeholders:

* Human Resources managers who need early warning systems to implement retention strategies
* Department supervisors who require insights to improve team management and employee satisfaction

1 Key Performance Indicator (KPI): Percentage reduction in actual employee turnover rate compared to baseline (measured quarterly over a 12-month period)

1. Data Collection & Preprocessing (8 points)

2 Data Sources:

* Human Resources Information System (HRIS): Employee demographics, job history, performance reviews, salary information, benefits usage, training records
* Employee engagement surveys: Job satisfaction scores, work-life balance ratings, career development feedback, manager relationship assessments

1 Potential Bias: Selection bias in survey data - Employees who are already disengaged or planning to leave may be less likely to participate in engagement surveys, leading to underrepresentation of at-risk employees in the training data. This could result in the model missing important patterns that indicate turnover risk.

3 Preprocessing Steps:

* Handle missing data using forward-fill for time-series features like performance scores and median imputation for numerical features like salary
* Feature engineering to create meaningful predictors such as tenure ratios, promotion frequency, and engagement trend indicators over time
* Normalize continuous variables like salary and performance scores using standardization to ensure equal weight in model training

1. Model Development (8 points)

Model Choice: Gradient Boosting (XGBoost)

Justification: XGBoost is ideal for this employee turnover prediction problem because it handles mixed data types effectively (numerical salary data and categorical job roles), provides feature importance rankings to identify key turnover predictors, performs well with structured HR data, and is robust to missing values common in employee datasets. It also captures complex interactions between factors like job satisfaction and tenure.

Data Splitting Strategy:

* Training set (70%): Used to train the model on historical employee data
* Validation set (15%): Used for hyperparameter tuning and model selection
* Test set (15%): Used for final unbiased performance evaluation
* Implement temporal splitting where older employee records are used for training and more recent data for testing to simulate real-world deployment conditions

2 Hyperparameters to Tune:

* Learning rate: Controls how much each tree contributes to the final prediction; lower rates generally improve accuracy but require more trees and longer training time
* Maximum depth: Controls tree complexity and overfitting; deeper trees can capture more complex patterns but may memorize training data rather than generalize to new employees

1. Evaluation & Deployment (8 points)

2 Evaluation Metrics:

* Recall (Sensitivity): Critical because missing an employee who actually leaves is more costly than false alarms; we want to identify as many potential leavers as possible to enable intervention
* Precision: Important to avoid overwhelming HR managers with false positives; ensures retention efforts are focused on employees who truly need intervention

Concept Drift: Concept drift occurs when the underlying relationship between employee characteristics and turnover likelihood changes over time. For example, economic conditions, company culture changes, or generational shifts in work preferences could alter what factors predict employee turnover.

Monitoring approach: Track model performance metrics monthly, compare feature distributions between training and current employee data, and retrain the model quarterly using recent data while maintaining performance benchmarks.

1 Technical Challenge: Real-time data integration - The system must continuously integrate data from multiple HR systems with different update frequencies and formats. Employee data changes frequently (promotions, role changes, performance updates), requiring the system to handle real-time data streaming while maintaining data consistency and ensuring predictions reflect the most current employee status for accurate turnover risk assessment.

Part 2: Case Study Application (40 points)

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

1. Problem Scope (5 points)

Problem Definition: Develop an AI system to predict which patients have high risk of readmission within 30 days of discharge to enable proactive interventions, improve patient outcomes, and reduce healthcare costs.

Objectives:

* Identify high-risk patients with 85% accuracy before discharge
* Reduce 30-day readmission rates by 20% through targeted interventions
* Provide actionable insights to healthcare teams for discharge planning
* Optimize resource allocation for post-discharge care coordination

Stakeholders:

* Primary: Physicians, nurses, discharge planners, case managers
* Secondary: Hospital administrators, insurance providers, patients and families
* Regulatory: Healthcare quality oversight bodies, compliance officers

1. Data Strategy (10 points)

Data Sources:

* Electronic Health Records (EHRs): Medical history, diagnoses, medications, vital signs, lab results, procedure codes
* Administrative data: Length of stay, admission type, insurance information, discharge disposition
* Demographics: Age, gender, socioeconomic status, geographic location
* Clinical notes: Discharge summaries, nursing assessments (processed via natural language processing)
* Pharmacy records: Medication adherence, prescription complexity, drug interactions
* Social determinants: Housing stability, transportation access, caregiver support systems

2 Ethical Concerns:

1. Patient Privacy and Data Security: Electronic health records contain highly sensitive personal information protected under HIPAA regulations. Risk of data breaches or unauthorized access could expose confidential patient information and violate privacy rights.
2. Algorithmic Bias and Health Equity: Training data may reflect historical healthcare disparities, potentially leading to biased predictions that systematically underestimate readmission risk for certain demographic groups, particularly racial minorities or socioeconomically disadvantaged patients.

Preprocessing Pipeline:

1. Data Integration: Merge data from multiple hospital systems and standardize medical coding formats
2. Missing Data Handling: Use clinical decision rules for critical values and multiple imputation for lab results based on patient similarity
3. Feature Engineering:
   * Create comorbidity indices (Charlson Comorbidity Index, Elixhauser scores)
   * Calculate medication complexity scores and polypharmacy indicators
   * Generate time-based features (days since last admission, length of stay patterns)
   * Develop social risk scores combining housing, transportation, and support factors
4. Temporal Alignment: Ensure all features reflect patient status at the time of discharge
5. Normalization: Standardize continuous variables like lab values and vital signs using z-score normalization
6. Categorical Encoding: Use one-hot encoding for diagnoses, procedures, and medications
7. Model Development (10 points)

Model Selection: Random Forest

Justification: Random Forest is well-suited for this healthcare prediction problem because it handles mixed data types effectively (numerical lab values and categorical diagnoses), provides interpretable feature importance rankings that clinicians can understand, performs well with structured medical data, is robust to missing values common in healthcare datasets, and reduces overfitting through ensemble averaging.

Hypothetical Confusion Matrix: Predicted No Readmission | Readmission Actual No Readmission 820 | 80 Actual Readmission 60 | 40

Calculations:

* Precision = 40/(40+80) = 0.33 (33%)
* Recall = 40/(40+60) = 0.40 (40%)
* Accuracy = (820+40)/(820+80+60+40) = 0.86 (86%)

1. Deployment (10 points)

Integration Steps:

1. API Development: Create secure RESTful API endpoints for real-time prediction requests
2. EHR Integration: Embed prediction module directly into existing discharge workflow systems
3. Clinical Decision Support: Implement risk score alerts and recommendations in physician dashboards
4. User Interface Design: Create intuitive displays showing risk scores with contributing factors and suggested interventions
5. Automated Scoring: Set up batch processing for daily risk assessment of all discharged patients
6. Performance Monitoring: Establish real-time tracking of prediction accuracy and system performance

HIPAA Compliance:

* Data Encryption: Implement end-to-end encryption for all data transmission and storage using AES-256 standards
* Access Controls: Deploy role-based authentication with multi-factor authentication and comprehensive audit logging
* Minimum Necessary Principle: Restrict data access to only information required for specific predictions
* Business Associate Agreements: Ensure all third-party vendors and cloud providers sign appropriate BAAs
* Regular Security Audits: Conduct quarterly penetration testing and vulnerability assessments
* Staff Training: Provide mandatory ongoing HIPAA compliance education for all system users

1. Optimization (5 points)

Method to Address Overfitting: Cross-Validation with Early Stopping

Implement stratified k-fold cross-validation (k=5) combined with early stopping mechanisms:

* Use stratified sampling to ensure balanced representation of readmission cases across all folds
* Monitor validation performance during training and stop when validation error begins increasing
* Apply bootstrap aggregating within each fold to further reduce variance
* Use out-of-bag error estimation to assess model generalization without separate validation set
* This approach prevents the model from memorizing specific patient patterns while ensuring robust performance across diverse patient populations

Part 3: Critical Thinking (20 points)

1. Ethics & Bias (10 points)

How Biased Training Data Affects Patient Outcomes: Biased training data could lead to systematic underestimation of readmission risk for certain patient populations. For example, if training data contains fewer examples from minority communities or rural areas, the model may fail to recognize their specific risk factors such as transportation barriers or cultural factors affecting medication adherence. This could result in minority patients receiving inadequate post-discharge support, widening existing health disparities, reducing trust in healthcare systems among affected communities, and potentially creating legal liability for discriminatory healthcare practices.

Strategy to Mitigate Bias: Implement Fairness-Aware Model Training with Demographic Parity Constraints by adding algorithmic fairness techniques that ensure equitable treatment across demographic groups. This includes adding fairness constraints during model training to equalize prediction accuracy across racial and ethnic groups, using adversarial debiasing methods to remove demographic information from internal model representations, conducting regular bias audits by evaluating model performance across different patient subgroups, and establishing feedback loops with diverse clinical stakeholders to identify and address potential sources of bias in real-world deployment.

1. Trade-offs (10 points)

Interpretability vs. Accuracy Trade-off: In healthcare, there exists significant tension between using highly accurate but complex models (such as deep neural networks) versus simpler, more interpretable models (like logistic regression). Healthcare providers need to understand why a model makes specific predictions to trust its recommendations and explain decisions to patients and families. However, simpler models may miss complex medical patterns that could be life-saving. The optimal approach often involves using moderately complex models like Random Forest or Gradient Boosting that provide good predictive accuracy while still offering feature importance explanations that clinicians can understand and act upon.

Impact of Limited Computational Resources: Limited computational resources would significantly impact model choice and deployment strategy. The hospital would need to choose simpler, more efficient models such as logistic regression or basic Random Forest instead of computationally intensive deep learning approaches. They would need to reduce feature complexity and dimensionality through careful feature selection, implement efficient batch processing rather than real-time predictions for all patients, consider cloud-based solutions for model training while keeping inference local to reduce ongoing costs, prioritize model efficiency over marginal accuracy improvements, and potentially explore federated learning approaches to share computational costs with other healthcare institutions.

Part 4: Reflection & Workflow Diagram (10 points)

1. Reflection (5 points)

Most Challenging Part: The most challenging aspect of this workflow was balancing the technical machine learning requirements with the complex ethical, regulatory, and clinical constraints inherent in healthcare applications. Unlike other domains, healthcare AI must navigate stringent privacy regulations like HIPAA, address potential life-or-death consequences of prediction errors, ensure clinical interpretability for medical professionals, and avoid perpetuating existing healthcare disparities. The challenge lies in maintaining high predictive performance while ensuring the system is fair, transparent, and clinically actionable across diverse patient populations.

Improvements with More Time/Resources:

* Multi-institutional validation: Test the model across different hospital systems and geographic regions to ensure broad generalizability
* Longitudinal outcome tracking: Implement long-term follow-up studies to validate that interventions based on predictions actually improve patient outcomes
* Advanced natural language processing: Develop sophisticated NLP capabilities to extract richer insights from clinical notes and discharge summaries
* Clinician feedback integration: Create systematic mechanisms for collecting and incorporating feedback from healthcare providers to continuously improve model performance
* Causal inference analysis: Move beyond predictive modeling to understand causal relationships between interventions and readmission outcomes

1. Diagram (5 points)

AI Development Workflow for Hospital Readmission Prediction

1. BUSINESS UNDERSTANDING ↓ [Define readmission problem, identify stakeholders, establish success metrics] ↓
2. DATA UNDERSTANDING ↓ [Explore EHR data, assess data quality, identify available sources] ↓
3. DATA PREPARATION ↓ [Clean data, integrate sources, engineer features, handle missing values] ↓
4. MODELING ↓ [Select algorithm, train model, tune hyperparameters, validate performance] ↓
5. EVALUATION ↓ [Test model performance, validate with clinicians, assess fairness] ↓
6. DEPLOYMENT ↓ [Integrate with EHR systems, ensure HIPAA compliance, monitor performance] ↓
7. MONITORING & MAINTENANCE ↓ [Track model drift, retrain periodically, gather clinical feedback] ↓ [Continuous Feedback Loop back to Business Understanding]

Key Considerations Throughout:

* Ethics and Fairness: Continuously assessed at every stage of development
* Regulatory Compliance: HIPAA and FDA requirements integrated throughout the process
* Clinical Validation: Healthcare professionals involved in evaluation and deployment phases
* Patient Safety: Primary consideration in all decision-making processes
* Continuous Improvement: Regular model updates based on new data and clinical outcomes