Understanding the AI Development Workflow

Part 1: Short Answer Questions (30 Points)

Step 1: Problem Definition (6 Points)
Hypothetical AI Problem:
Developing an AI model to predict the risk of patient readmission within 30 days after discharge.

Objectives:
- Prevent avoidable hospital readmissions to enhance patient outcomes.
- Improve efficiency in hospital resource allocation.
- Personalize post-discharge care plans based on readmission risk.

Stakeholders:
- Hospital clinicians, care coordinators, and administrators.
- Patients and their families.

Key Performance Indicator (KPI):
Reduction in 30-day hospital readmission rate over a given timeframe.

Step 2: Data Collection & Preprocessing (8 Points)
Data Sources:
- Electronic Health Records (EHRs): Clinical notes, lab results, diagnosis history, discharge summaries.
- Patient Demographics & Social Determinants: Age, gender, location, insurance type, socioeconomic status.

Potential Bias:
Underrepresentation of minority or underserved populations may result in unfair predictions. Historical disparities in healthcare access and treatment may embed systemic bias into the training data.

Preprocessing Steps:
1. Handling Missing Data: Impute missing values using statistical or model-based techniques.
2. Normalizing Numerical Features: Scale variables such as age, lab values, and hospital stay duration.
3. Encoding Categorical Variables: Use one-hot encoding for nominal features; apply label encoding or embeddings for ordinal data.

Step 3: Model Development (8 Points)
Model Choice:
Gradient Boosting Machine (e.g., XGBoost) - chosen for its performance on structured data, robustness to missing values, and interpretability via feature importance.

Data Splitting Strategy:
- 70% Training Set
- 15% Validation Set (for tuning hyperparameters such as learning rate and tree depth)
- 15% Test Set (for evaluating generalization)

Hyperparameters:
- Learning Rate
- Max Tree Depth

Step 4: Evaluation & Deployment (8 Points)
Evaluation Metrics:

Precision = 80 / (80 + 15) = 0.842
Recall = 80 / (80 + 20) = 0.800

Concept Drift:
Definition: A shift in the data distribution over time that reduces model accuracy.
Monitoring: Use tools like Alibi Detect or River for drift detection and retrain the model.

Scalability:
- Use Docker containers
- Host on AWS/GCP
- RESTful integration with hospital EHR systems

Part 2: Case Study Application (40 Points)

Problem Scope:
Build an AI solution to predict the likelihood of hospital readmission within 30 days of discharge.

Objectives:
- Decrease readmission rates by at least 15%.
- Enable more accurate, patient-specific discharge planning.
- Support efficient use of hospital resources.

Data Strategy:
Sources:
- EHRs, Demographics, Admission/Discharge Records

Ethical Concerns:
1. Patient Privacy (HIPAA)
2. Algorithmic Bias

Preprocessing:
- Cleaning, Feature Engineering (comorbidities, readmissions), One-hot encoding & normalization

Model Development:
XGBoost selected for accuracy & interpretability.

Confusion Matrix:

Actual/Predicted	Readmit	No Readmit
Readmit	80	20
No Readmit	15	85

Metrics:
- Precision = 0.842
- Recall = 0.800

Deployment:
- Docker, Flask/FastAPI, EHR Integration via REST API

Compliance:
- Encryption, RBAC, HIPAA audits

Optimization:
- Cross-validation, L1/L2 Regularization, Early stopping

Part 3: Critical Thinking (20 Points)

Ethics & Bias:
Impact:
Bias in historical data may underpredict risks for minority groups, resulting in unfair treatment.

Strategies:
- Bias Audits
- Fairness-aware Algorithms
- Diverse Datasets
- Stakeholder Involvement

Trade-offs:
Interpretability vs. Accuracy:
- Simple models = transparency
- Complex models = accuracy + SHAP explanations

Resource Constraints:
- Model compression
- Lightweight models
- Edge deployment

Part 4: Reflection & Workflow Diagram (10 Points)

Reflection:
Most challenging part was preprocessing and bias mitigation.
Improvements would include better datasets, collaboration with clinicians, and access to compute resources.

Workflow:
Problem Definition -> Data Collection -> Preprocessing -> Model Development -> Evaluation -> Deployment -> Monitoring