Part 1: Short Answer Questions (30 points)

1. Problem Definition (6 points)

Hypothetical AI Problem:

Predicting crop disease outbreaks in Kenyan maize farms using satellite and sensor data.

Objectives:

• Detect early signs of crop disease from environmental and image data.

• Alert farmers and agricultural officers for timely intervention.

• Reduce crop losses and improve food security.

Stakeholders:

• Smallholder farmers in Kenya.

• Ministry of Agriculture and ICT Authority.

Key Performance Indicator (KPI):

• Disease detection accuracy (%) — proportion of correctly identified disease cases vs total cases.

2. Data Collection & preprocessing (8 points)

Data Sources:

• Satellite imagery (e.g., NDVI, temperature, humidity).

• IoT sensor data from soil and weather monitoring devices.

Potential Bias:

• Geographic bias: Data may be concentrated in well-funded regions, underrepresenting remote or low-resource farms.

Preprocessing Steps:

• Handle missing sensor readings using interpolation or imputation.

• Normalize environmental variables (e.g., temperature, moisture) for consistent scale.

• Augment image data (e.g., rotation, zoom) to improve model robustness.

3. Model Development (8 points)

Chosen Model:

Convolutional Neural Network (CNN) — ideal for analyzing spatial patterns in satellite and image data.

Data Splitting Strategy:

• 70% training, 15% validation, 15% test — stratified to maintain class balance (healthy vs diseased crops).

Hyperparameters to Tune:

• Learning rate: Controls how fast the model updates weights — critical for convergence.

• Number of convolutional layers: Affects model depth and ability to capture complex patterns.

4. Evaluation & Deployment (8 points)

Evaluation Metrics:

• Precision: Measures how many predicted disease cases were correct — important to avoid false alarms.

• Recall: Measures how many actual disease cases were detected — crucial for timely intervention.

Concept Drift:

• Occurs when the relationship between input data and target changes over time (e.g., new disease strains or climate shifts).

• Monitor using periodic re-evaluation on recent data and tracking drops in model accuracy.

Technical Challenge:

• Scalability: Deploying the model across diverse regions with limited connectivity and hardware — requires lightweight models and edge computing solutions.