

# 1-Month Preparation Plan for an AI Hackathon

Participating in an AI hackathon requires structured planning to maximize efficiency and innovation. Below is a detailed 1-month timeline covering problem selection, data sourcing, model development, team coordination, and risk mitigation.

## Week 1: Problem Selection & Quantitative Formulation

### DAY 1-2: IDEATION & PROBLEM DEFINITION

- **Brainstorming:** Identify real-world problems (e.g., healthcare, finance, climate).
- **Feasibility Check:** Ensure the problem is AI-solvable (classification, regression, NLP, CV).
- **Quantitative Formulation:** Define success metrics (accuracy, F1-score, RMSE).

### DAY 3-4: LITERATURE REVIEW & BENCHMARKING

- Research existing solutions (Kaggle, arXiv, GitHub).
- Identify gaps where innovation is possible.
- Finalize problem statement with measurable objectives.

### DAY 5-7: INITIAL DATA SCOUTING

- Identify potential datasets (Kaggle, UCI, APIs, web scraping).
- Check data availability, licensing, and biases.
- Draft a preliminary data preprocessing plan.

**Deliverable:** Clear problem statement, success metrics, and initial dataset sources.

## Week 2: Data Sourcing & Preprocessing

### DAY 8-10: DATA COLLECTION & CLEANING

- Acquire datasets (APIs, public repositories, synthetic data if needed).
- Handle missing values (imputation, deletion).
- Remove duplicates and outliers.

### DAY 11-12: FEATURE ENGINEERING

- Normalize/standardize numerical data.
- Encode categorical variables (One-Hot, Label Encoding).
- Extract meaningful features (time-based, text embeddings).

### DAY 13-14: EXPLORATORY DATA ANALYSIS (EDA)

- Visualize distributions (histograms, box plots).
- Detect correlations (heatmaps, pair plots).
- Identify class imbalances (SMOTE, undersampling).

**Deliverable:** Cleaned dataset, EDA report, feature engineering pipeline.

## Week 3: Model Selection & Training

### DAY 15-17: BASELINE MODELS

- Start with simple models (Linear Regression, Decision Trees).
- Evaluate performance (cross-validation, confusion matrix).
- Compare against benchmarks.

### DAY 18-20: ADVANCED MODEL EXPERIMENTATION

- Test deep learning (CNNs, RNNs, Transformers) if applicable.
- Hyperparameter tuning (Grid Search, Random Search).
- Optimize for speed vs. accuracy trade-offs.

**DAY 21: COMPUTE & INFRASTRUCTURE SETUP**

- Choose cloud platforms (Google Colab, AWS, Azure).
- Optimize GPU usage for training.
- Implement version control (Git, DVC).

**Deliverable:** Trained models with performance metrics, code repository.

# Week 4: Refinement, Presentation & Risk Mitigation

**DAY 22-24: MODEL OPTIMIZATION & VALIDATION**

- Ensemble methods (Bagging, Boosting).
- Interpretability (SHAP, LIME).
- Final validation on unseen data.

**DAY 25-26: PRESENTATION & DOCUMENTATION**

- Prepare slides with problem, approach, results.
- Record a demo video if required.
- Document code (README, comments).

**DAY 27-28: CONTINGENCY PLANNING**

- Backup datasets and models.
- Prepare alternative approaches if primary fails.
- Conduct a mock presentation for feedback.

**Deliverable:** Final model, presentation deck, backup plan.

## Team Roles & Milestones

Role	Responsibilities
Team Lead	Project management, timeline tracking
Data Engineer	Data collection, cleaning, EDA
ML Engineer	Model training, optimization
Frontend/UX	Demo visualization (Streamlit, Gradio)
Documentation	Slides, reports, GitHub README

## Key Milestones:

1. WEEK 1: PROBLEM FINALIZED.
2. WEEK 2: CLEAN DATASET READY.
3. WEEK 3: BASELINE & ADVANCED MODELS TRAINED.
4. WEEK 4: FINAL SUBMISSION PREPARED.

## **Risk Mitigation Strategies**

- Data Issues: Have backup datasets.
- Model Failure: Keep simpler models as fallback.
- Compute Limits: Use free-tier cloud GPUs.
- Team Conflicts: Clear role division, daily standups.

By following this structured approach, the team can efficiently tackle the hackathon while minimizing last-minute hurdles.