



## AI Training Dataset Specification Template

Use this template to define the structure, quality expectations, and requirements for any dataset used in AI model training. Fill in each section to ensure clarity, consistency, and alignment with project goals.

### 1. Project Overview

- Project Name:
- AI Use Case:
- Business Objective:
- Primary Stakeholders:
- Target Model Type (classification, regression, NLP, etc.):

### 2. Problem Definition

- What problem is the AI solving?
- Expected outcomes:
- Success criteria:
- Constraints or limitations:

### 3. Dataset Description

- Dataset Name:
- Source(s):
- Data Type (text, images, tabular, audio, etc.):
- Dataset Size (rows/entries):
- File Formats:
- Update Frequency:

### 4. Input Features (X)

- List all required features:
- Feature types:

- Feature definitions:
- Handling missing values:
- Normalization/standardization requirements:

## 5. Output Labels (Y)

- Label description:
- Label type (categorical, numeric, multi-class, etc.):
- Label encoding requirements:
- Sample labeled examples:

## 6. Quality Requirements

- Minimum completeness threshold (%):
- Minimum accuracy/validity requirements:
- Noise tolerance:
- Data consistency rules:

## 7. Sample Rows

- Include 3–5 example rows here or attach a sample file.

## 8. Edge Cases

- Known anomalies or outliers:
- Rare but important patterns:
- Data exceptions the AI must learn to handle:

## 9. Exclusions

- Data to exclude from training:
- Reasons for exclusion:
- Privacy or compliance restrictions:

## 10. Security & Compliance

- PII present (Yes/No):
- De-identification needed:
- Compliance frameworks (GDPR, HIPAA, CCPA, etc.):
- Data access restrictions:

## 11. Version Control

- Dataset version:
- Change history:
- Owner/responsible team: