

# Identity and Purpose Checklist

## Data Management (Derived Requirements)

### 1. Data Management Requirements Capture:

- Has the data management process been properly captured? ✓
- Are all necessary data points identified? ✓

### 4. Data Collection:

- Is the data collected from a single source or multiple sources? ✓
- Are the data collection methods documented? ✓

### 7. Data Labelling (only in supervised learning):

- Is the labelling process manual or automatic? ✓
- Are the labels accurate and consistent? ✓

### 10. Data Preparation (pre-processing, data transformation, feature engineering):

- Has the data been pre-processed for quality and consistency? ✓
- Are the features engineered to improve model performance? ✓

### 13. Identification of Data Sets:

- Have all necessary data sets (training, validation, test) been identified? ✓
- Are the data sets properly separated and stored? ✓

### 16. Data Set Verification:

- Has the accuracy, completeness, and representativeness of each data set been verified? ✓
- Are the data sets free from unwanted bias? ✓

### 19. Independence Requirements between Data Sets:

- Have independence requirements between data sets been met? ✓
- Are there any dependencies or correlations between data sets that need to be addressed? ✓

## AI/ML Constituent ODD Definition

### 1. Nominal Data:

- Has nominal data been properly defined for each data set? ✓
- Are the nominal values consistent across all data sets? ✓

### 4. Edge Cases, Corner Cases, and Infeasible Corner Cases Data:

- Have edge cases, corner cases, and infeasible corner cases data been identified and addressed? ✓
- Are these cases properly represented in the training data? ✓

## **7. Detection and Removal of Inliers:**

- Has the model been trained to detect and remove inliers? ✓
- Are there any inliers present in the test data that need to be addressed? ✓

## **10. Detection and Management of Novelty:**

- Has the model been trained to detect novelty? ✓
- Are there any novelty present in the test data that need to be addressed? ✓

# **DQRs (Data Quality Requirements)**

## **1. Data Relevance:**

- Is the data relevant to support the intended use? ✓

## **3. Data Origin:**

- Can the origin of the data be determined? ✓

## **5. Annotation Process:**

- Are the annotation process requirements met? ✓

## **7. Format, Accuracy, and Resolution of Data:**

- Is the format, accuracy, and resolution of the data properly defined? ✓

## **9. Traceability of Data:**

- Can the data be traced back to its origin throughout the pipeline? ✓

## **11. Data Integrity:**

- Has the data been corrupted during storage, processing, or transmission? ✓

## **13. Completeness and Representativeness of Data Sets:**

- Are the data sets complete and representative of the intended use? ✓

Note: The above checklist is generated based on the provided text and is meant to be a comprehensive review of the requirements for AI/ML software development and review.