

Here is the generated checklist:

Data Management Checklist

Data Management Requirements Capture

- Have all data management requirements been captured?
☐ [Yes/No]
- Are all data management requirements documented and available for review?
☐ [Yes/No]

Data Collection

- Has data collection been performed for each individual data point?
☐ [Yes/No]
- Is the collected data relevant to the ML model's objectives?
☐ [Yes/No]

Data Labelling (only in supervised learning)

- Has data labelling been performed for all training, validation, and test datasets?
☐ [Yes/No]
- Are the labels accurate and consistent across all datasets?
☐ [Yes/No]

Data Preparation

- Has pre-processing, data transformation, and feature engineering been applied to the data?
☐ [Yes/No]
- Is the prepared data relevant and suitable for the ML model's objectives?
☐ [Yes/No]

Identification of Data Sets

- Have the various datasets (training, validation, test) been identified and documented?
☐ [Yes/No]
- Are the datasets accurate and consistent across all phases of the project lifecycle?

☐ [Yes/No]

Data Set Verification

- Has data set verification been performed for accuracy, completeness, and representativeness?

☐ [Yes/No]

- Are the verified datasets suitable for the ML model's objectives?

☐ [Yes/No]

Independence Requirements between Data Sets

- Have independence requirements been met between the training, validation, and test datasets?

☐ [Yes/No]

- Are the independent datasets suitable for the ML model's objectives?

☐ [Yes/No]

Identification and Elimination of Unwanted Bias

- Has unwanted bias inherent to the datasets been identified and eliminated?

☐ [Yes/No]

- Are the bias-free datasets suitable for the ML model's objectives?

☐ [Yes/No]

Note: This checklist is intended to be used by both the author developing the AI-Machine Learning software and the reviewer reviewing it.