

Here is the generated checklist in Markdown format:

Data Management Checklist

- Did you capture data management requirements? ☐☐
 - Is the objective to manage data and deliver data sets for the W-shaped process?
 - Are the individual data point objectives identified, as well as those for the data sets used later in the project lifecycle?
- Have you collected the necessary data? ☐☐
 - Is the data relevant to the AI/ML application?
 - Are there any missing or irrelevant data points?
- Has the data been labelled (only for supervised learning)? ☐☐
 - Is the labelling accurate and consistent?
 - Are there any errors or inconsistencies in the labelling process?
- Have you prepared the data (pre-processing, transformation, feature engineering)? ☐☐
 - Are the data sets clean and free of errors?
 - Are there any unnecessary features or variables that can be removed?
- Have you identified the various data sets used in the learning phase? ☐☐
 - Are the training, validation, and test data sets clearly defined?
 - Are there any duplicate or redundant data sets?
- Has the data been verified (accuracy, completeness, representativeness)? ☐☐
 - Is the data accurate and consistent with the ML requirements?
 - Are there any errors or inconsistencies in the verification process?
- Are the data sets independent? ☐☐
 - Are there any correlations or dependencies between the data sets?
 - Can the data sets be used separately without affecting each other?
- Has unwanted bias been identified and eliminated from the data sets? ☐☐
 - Is the data representative of the target population or scenario?
 - Are there any biases or unfair advantages in the data that need to be addressed?

Let me know if you'd like me to generate anything else! 😊