

Data quality checklist

Before using data in an AI system, check whether it is accurate, complete, consistent and up to date. This will help produce reliable results in AI outputs.

Not all data is equal. For example, names are more important than titles.

Accuracy

Check whether the data correctly reflects real people, objects or events.

- ☐ Verify the data against a trusted source
- ☐ Remove or correct values that are impossible or unrealistic

Example issues:

- ABN and company name do not match
- Person's age is outside a normal human range
- Postcode does not contain 4 digits.

Completeness

Check all the data required for AI task present and available.

- ☐ Verify all mandatory fields are populated. For example, with a user ID and timestamp
- ☐ Identify any missing data and check whether it is missing at random or if there is a pattern. For example, AI never sees data from a specific region.

Consistency

Check that the data is consistent across different systems and sources.

- ☐ Verify that the same information has the same status across your systems. For example, a customer has the same status in your CRM as in your billing database.
- ☐ Identify any missing data and check whether it is missing at random or if there is a pattern. For example, AI never sees data from a specific region.

Data rules

Check the data conforms to the right format, range and business rules.

- ☐ Validate email addresses are in a valid [name@domain.com](#) format
- ☐ Check your category tags match the allowed list of options. For example, 'MKtg' compared to 'Marketing'.

Timeliness

Check the data is available when needed and up to date for its intended use.

- ☐ Understand the delay between a real-world event and when that data is available in your AI system. For example, does your AI see stock levels in real time or with a two-hour delay
- ☐ Check whether the data is still relevant for your needs. For example, a customer's interests from 2022 still useful in 2026.

Uniqueness

Check that each record is recorded only once.

- ☐ Identify duplicate records where the same entity appears under different IDs.
- ☐ Check whether processing the same event twice would skew your AI results. For example, a transaction processed twice would distort your analysis.

Additional data trust checks for AI

The [United Kingdom Data Management Association](#) (DAMA UK) and emerging standards suggest these additional trust checks for AI.

Traceability (lineage)

Check you could prove which data source influenced a specific AI output.

- ☐ Ensure you can explain to customers or regulators which data the AI used to make a decision.

Ownership

Decide on who is accountable for every AI system in your business.

- ☐ Assign a specific person accountable for every AI system your organisation uses.
- ☐ Ensure the accountable people are familiar with the technology and understand its business implications.

Interpretability/reference

Check that your data uses clear definitions and reference material so that AI can understand what it means.

- ☐ Ensure all coded values have definitions. For example, what 'status 04' means.

Data coverage/bias

Check your data covers all the scenarios you expect the AI to handle.

- ☐ Verify your data includes all the groups or situations the AI will encounter. For example, your training data include all demographics your AI will serve.