



AI-Ready Data Workflow

This workflow outlines the full lifecycle of preparing data for AI systems—from raw collection to validated model-ready inputs. Use this framework to standardize processes, improve data quality, and ensure AI success.

1. Raw Data Intake

- Collect data from all relevant sources (CRM, ERP, spreadsheets, APIs, logs).
- Verify access permissions and ownership.
- Check for sensitive data or compliance flags.

2. Data Cleaning

- Remove duplicates and corrupt records.
- Fix missing values and normalize formatting.
- Correct category labels and validate timestamps.

3. Data Structuring

- Convert data to structured formats (tabular, JSON, etc.).
- Standardize column names and schema.
- Organize data into training, validation, and test sets.

4. Compliance & Privacy Checks

- Identify and mask/remove PII.
- Confirm consent for user data.
- Verify alignment with GDPR, CCPA, HIPAA, or industry regulations.

5. Feature Engineering

- Select key features required for model training.

- Create new derived features if needed.
- Normalize or scale values for model readiness.

6. AI Model Input Preparation

- Format data according to model requirements.
- Ensure no leakage between training and testing sets.
- Validate data types and encoding format.

7. Output Validation

- Run initial tests to confirm model behavior.
- Check prediction consistency and accuracy.
- Document any anomalies for refinement.

8. Continuous Improvement Loop

- Monitor data drift over time.
- Update datasets based on new insights or errors.
- Maintain version control for datasets and transformations.