# AI Data Cleaning Checklist

## 1. Remove Duplicates
• Run duplicate detection across all tables.

• Verify that remaining records are unique.

• Check for near-duplicates (similar names, emails, IDs).

## 2. Normalize Formatting
• Ensure consistent capitalization (e.g., names, titles).

• Standardize date formats (MM/DD/YYYY or ISO).

• Unify phone number formatting.

## 3. Fix Missing Values
• Identify all missing data fields.

• Decide to fill, default, infer—or remove rows.

• Flag any critical missing values for review.

## 4. Standardize Labels
• Ensure category names are consistent (e.g., 'USA' vs 'United States').

• Combine similar or redundant categories.

• Confirm all tags follow the same naming convention.

## 5. Validate Timestamps
• Check time zone consistency.

• Verify chronological accuracy.

• Remove impossible or corrupt dates.

## 6. Remove Irrelevant Fields

• Identify non-essential columns.

• Remove or archive low-value fields.

• Ensure only relevant fields remain for AI use.

## 7. Privacy & Compliance Check

• Mask or remove PII where required.

• Confirm consent for sensitive data.

• Validate compliance with GDPR, CCPA, or industry requirements.