AI Curator — Data Wrangling Checklist

Data Collection

Checklist Item	Check
Identify data sources (databases, APIs, spreadsheets, SaaS tools, web scraping)	☐
Gather initial raw data	☐
Verify data access permissions & security requirements	☐
Document data sources & collection methods	☐

Data Cleaning

Checklist Item	Check
Remove duplicate records	☐
Handle missing values (impute, remove, or flag)	☐
Fix inconsistent formats (dates, units, casing)	☐
Standardize column names & data types	☐
Identify and handle outliers appropriately	☐
Remove irrelevant or redundant features	☐

Data Transformation

Checklist Item	Check
Normalize or scale numeric features	☐
Encode categorical variables (one-hot, label encoding)	☐
Process text data (lowercasing, tokenization, stemming)	☐
Create calculated or engineered features	☐
Ensure all variables are in correct formats	☐

Data Integration

Checklist Item	Check
Merge datasets from multiple sources	☐
Resolve conflicting values between datasets	☐
Validate matching keys / identifiers	☐

Data Enrichment

Checklist Item	Check
Add external datasets (market, demographics, etc.)	☐
Create new features from contextual data	☐
Use AI/statistics to augment or enrich data	☐
Ensure enriched data is current & relevant	☐

Data Validation

Checklist Item	Check
Check for logical errors & contradictions	☐

Ensure data follows business rules	☐
Verify integrity (checksums, cross-checks)	☐
Confirm transformations didn't introduce errors	☐

Data Formatting & Delivery

Checklist Item	Check
Convert data to required formats (CSV, JSON, SQL, etc.)	☐
Format for readability & consistency	☐
Ensure column names & metadata follow standards	☐
Save final dataset securely with versioning	☐