ASSESSMENT AREA 01 / 05

Data flows

Can your ML pipeline access the operational records it needs?

ML use case / prediction target: _____

Organization / reviewer / date: _____

Check each item when the assessment is documented, not when the issue is necessarily resolved. Record open findings below.

Operational records assessment

- [ ] Identify ERP, CRM and legacy systems containing relevant records.
- [ ] Determine how the required records can be accessed.
- [ ] Document records unavailable through existing integrations.

Manual CSV exports

- [ ] Identify datasets transferred through CSV exports.
- [ ] Determine which exports depend on manual employee actions.
- [ ] Document extraction steps that need changing or automation.

Custom integration scripts

- [ ] Identify scripts used to extract operational records.
- [ ] Map the source systems and destinations connected by each script.
- [ ] Document integration dependencies that could interrupt data delivery.

EXPECTED EVIDENCE

Source-system inventory and access barriers, distinguishing accessible records from records requiring integration work.

Findings / affected datasets / open questions

ASSESSMENT AREA 02 / 05

ML data quality

Which data quality issues must be resolved before ML data ingestion?

Source records

- [ ] Identify records with missing values or fields.
- [ ] Document differences in formats across operational sources.
- [ ] Record suspected inaccuracies requiring source-system validation.
- [ ] Identify issues with an established correction method.

Duplicate records

- [ ] Determine whether duplicates have been identified in source datasets.
- [ ] Identify existing deduplication steps.
- [ ] Document datasets requiring further duplicate analysis.
- [ ] Record unresolved questions about which records should be retained.

Source data cleaning

- [ ] Document cleaning tasks for existing source records.
- [ ] Identify preprocessing activities required before ingestion.
- [ ] Separate source-data corrections from ingestion pipeline development.
- [ ] Record dependencies between cleaning and downstream processing.

EXPECTED EVIDENCE

Data-quality issue register linked to datasets and source systems; separate source corrections, preprocessing and ingestion work.

Findings / affected datasets / open questions

ASSESSMENT AREA 03 / 05

Historical records

Do historical records contain the evidence for your prediction target?

Prediction target suitability

- [ ] Define the operational outcome the model is intended to predict.
- [ ] Identify historical records associated with that outcome.
- [ ] Map records to their original ERP, CRM or other operational systems.
- [ ] Document information required for the intended prediction.

Historical coverage

- [ ] Identify the earliest and latest records for each relevant source.
- [ ] Document periods for which records cannot be retrieved.
- [ ] Distinguish inaccessible records from records never collected.
- [ ] Identify whether recovering records requires extra integration work.

Recorded historical outcomes

- [ ] Identify where actual historical outcomes are recorded.
- [ ] Determine whether outcome records are accessible.
- [ ] Document missing outcome information.
- [ ] Record unresolved questions about linking inputs to actual outcomes.

EXPECTED EVIDENCE

Historical coverage and outcome mapping, with gaps documented. Availability alone does not establish suitability for training or validation.

Findings / affected datasets / open questions

ASSESSMENT AREA 04 / 05

Data infrastructure

Can existing infrastructure support repeatable ML data delivery?

Data transfers

- [ ] Identify operational databases supplying records for the ML use case.
- [ ] Document how records are extracted from each source.
- [ ] Map intermediate files, scripts, applications and destinations.
- [ ] Identify transfer paths lacking a documented integration mechanism.

Manual handoffs and custom scripts

- [ ] Identify extraction tasks requiring manual employee actions.
- [ ] Document scheduled exports and associated source systems.
- [ ] Inventory custom scripts used to transfer operational records.
- [ ] Identify dependencies on applications, scripts or manual handoffs.
- [ ] Record transfers requiring further technical evaluation.

Preprocessing dependencies

- [ ] Identify preprocessing required before ingestion.
- [ ] Document the component or workflow responsible for each step.
- [ ] Map dependencies between extraction, cleaning and ingestion.
- [ ] Identify preparation activities without an assigned component.
- [ ] Record infrastructure changes requiring technical evaluation.

EXPECTED EVIDENCE

End-to-end extraction and ingestion workflow map, including manual dependencies and unassigned processing steps.

Findings / affected datasets / open questions

ASSESSMENT AREA 05 / 05

Access & governance

Who controls access to and maintenance of operational ML data?

Dataset owners

- [ ] Identify the responsible owner or team for each relevant dataset.
- [ ] Document the business systems containing the datasets.
- [ ] Identify who can authorize access to required records.
- [ ] Record datasets with unclear ownership.

Permissions and privacy controls

- [ ] Identify existing ERP, CRM and database access permissions.
- [ ] Document which teams or systems have authorized access.
- [ ] Identify datasets subject to privacy restrictions.
- [ ] Record unresolved authorization requirements for the ML workflow.
- [ ] Identify access decisions requiring data-owner or security-team review.

Maintenance responsibilities

- [ ] Identify the team maintaining each operational data source.
- [ ] Identify who maintains cleaning and preprocessing workflows.
- [ ] Record unassigned responsibilities and maintenance questions that must be addressed before the pilot decision.

EXPECTED EVIDENCE

Dataset ownership, access and maintenance register showing confirmed responsibilities and unresolved authorization decisions.

Findings / affected datasets / open questions

ASSESSMENT AREA 06 / DECISION RECORD

Consolidate your findings

Classify each assessed area using the outcomes below. Do not interpret a completed checklist as proof of overall ML readiness.

R | Documented remediation need

A specific barrier is identified. Define and evaluate corrective work.

V | Further verification required

A readiness condition remains unconfirmed. Obtain missing evidence before approval.

N | No barrier identified

No issue was found in this area. Record the finding without inferring overall readiness.

Assessment area	R	V	N
Data flows	☐	☐	☐
ML data quality	☐	☐	☐
Historical records	☐	☐	☐
Data infrastructure	☐	☐	☐
Access & governance	☐	☐	☐

PRIORITY ACTION REGISTER

Legacy-system access barriers

Data cleaning / infrastructure integration work

Unverified historical data / governance requirements

Next step: scope a technical assessment

Prioritize remediation and missing evidence for the intended ML use case before approving a pilot.

Explore our AI readiness assessment >