



AI Outcome Measurement Scorecard

A worksheet for defining what better means before an AI workflow becomes a business case.

STRUCTURE - CLARITY - BETTER DECISIONS

Practical architecture for accountable AI-enabled work.

What this template is for

Map baseline metrics, success criteria, quality signals, review burden, customer impact, cost-to-serve, and metric ownership before celebrating output volume. Use this guide to make the operating choice explicit, test it with evidence, and record who owns the next decision.

BEST USED WHEN

Teams proving operational impact after an AI workflow launches.

WHO SHOULD BE IN THE ROOM

Include the workflow owner, a person who performs the work, a reviewer with authority to approve or stop it, and the relevant data or tool owner.

CANADIAN CONTEXT

Confirm applicable federal, provincial, territorial, sector, professional, contractual, accessibility, and bilingual-service requirements. Use this workbook as operational preparation, not legal advice.

A five-step working rhythm

01

Time saved

Which cycle time, handoff, or preparation step should improve?

02

Quality improved

Which error, completeness, decision, or customer outcome should improve?

03

Review burden

Does AI reduce, shift, or increase human review work?

04

Cost-to-serve

How does model, tool, and operator cost compare with workflow value?

05

Confidence signal

What would prove the system is helping rather than creating hidden rework?

Build the working record

Complete one record for one workflow. Write enough to support a decision, not a generic technology wish list.

PROMPT	YOUR WORKING NOTES
01 Workflow measured	
02 Current baseline	
03 Business owner	
04 Time saved	
05 Quality improved	
06 Review burden	
07 Decision, owner, evidence, and review date	

Measure proposal preparation

STARTING FRICTION

Drafting time falls, but staff spend more time correcting unsupported claims.

WORKFLOW MAP



AI-SUPPORTED ROLE

Compile baseline and pilot signals while separating output volume from real business improvement. Flag uncertainty around time saved.

HUMAN AUTHORITY

The business owner interprets tradeoffs and decides whether to tune, scale, pause, or stop. Resolve the time saved question for this case.

PILOT MEASURE

Compare cycle time, quality, rework, cost, and reviewer effort before and after the pilot. Record time saved exceptions separately.

DISCUSSION PROMPT

What evidence would the reviewer need to trust, correct, reject, or escalate this output?

Measure stock decisions

STARTING FRICTION

Order suggestions look faster; the owner needs to see spoilage, stock-outs, and emergency purchases.

WORKFLOW MAP



AI-SUPPORTED ROLE

Compile baseline and pilot signals while separating output volume from real business improvement. Flag uncertainty around quality improved.

HUMAN AUTHORITY

The business owner interprets tradeoffs and decides whether to tune, scale, pause, or stop. Resolve the quality improved question for this case.

PILOT MEASURE

Compare cycle time, quality, rework, cost, and reviewer effort before and after the pilot. Record quality improved exceptions separately.

DISCUSSION PROMPT

What evidence would the reviewer need to trust, correct, reject, or escalate this output?

Measure client-review preparation

STARTING FRICTION

More summaries are produced, but the firm must test omissions and professional review time.

WORKFLOW MAP



AI-SUPPORTED ROLE

Compile baseline and pilot signals while separating output volume from real business improvement. Flag uncertainty around review burden.

HUMAN AUTHORITY

The business owner interprets tradeoffs and decides whether to tune, scale, pause, or stop. Resolve the review burden question for this case.

PILOT MEASURE

Compare cycle time, quality, rework, cost, and reviewer effort before and after the pilot. Record review burden exceptions separately.

DISCUSSION PROMPT

What evidence would the reviewer need to trust, correct, reject, or escalate this output?

Before the pilot grows

Use these prompts to test whether the proposed workflow can operate responsibly in its Canadian business context.

01

Time saved

The workflow has a measurable baseline. Name the owner and the operating decision before selecting a tool.

02

Quality improved

Success criteria are tied to business outcomes. Record the approved sources, permissions, and missing information.

03

Review burden

Quality and review burden are measured together. Test this boundary with representative and adverse cases.

04

Cost-to-serve

Metric owners are named. Preserve review, escalation, and correction evidence.

05

Confidence signal

Before/after comparison has a review cadence. Set a measurable stop condition and the next review date.

06

Readiness to proceed

AI value is operational value. A workflow that produces more output but weakens quality, confidence, or review capacity has not improved.

CANADIAN REFERENCE POINTS

Office of the Privacy Commissioner of Canada - Privacy and artificial intelligence
priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/

Government of Canada - Artificial Intelligence guidance, principles, and policy
canada.ca/en/services/science/innovation/artificial-intelligence.html

General operational guidance only. Confirm applicable legal, privacy, cybersecurity, professional, contractual, and regulatory requirements with qualified advisors.