



# Agent Testing Scenario Pack

A worksheet for testing agent behavior across normal cases, edge cases, restricted requests, tool failures, and escalation.

## STRUCTURE - CLARITY - BETTER DECISIONS

Practical architecture for accountable AI-enabled work.

# What this template is for

Build a repeatable scenario suite for agentic workflows before launch and use it as a regression backlog as tools, prompts, and policies change. Use this guide to make the operating choice explicit, test it with evidence, and record who owns the next decision.

## BEST USED WHEN

Agentic systems using tools, approvals, or multi-step workflows.

## 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

### Normal case

What should the agent do when context and tools are healthy?

02

### Missing data

How should the agent behave when required state is unavailable?

03

### Conflict

What should happen when sources or user instructions disagree?

04

### Restricted request

How should the agent handle unsafe, unauthorized, or policy-breaking requests?

05

### Tool failure

What retry, fallback, or escalation path is expected?

# 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 Agent workflow                             | <hr/> <hr/> <hr/>  |
| 02 Risk boundary                              | <hr/> <hr/> <hr/>  |
| 03 Test owner                                 | <hr/> <hr/> <hr/>  |
| 04 Normal case                                | <hr/> <hr/> <hr/>  |
| 05 Missing data                               | <hr/> <hr/> <hr/>  |
| 06 Conflict                                   | <hr/> <hr/> <hr/>  |
| 07 Decision, owner, evidence, and review date | <hr/> <hr/> <hr/>  |

# Test missing scope details

## STARTING FRICTION

A proposal agent should ask for clarification rather than invent a deadline.

## WORKFLOW MAP

SCENARIO

EXPECTED RESULT

EXECUTION

RECEIPT

REVIEWER SCORE

FIX

## AI-SUPPORTED ROLE

Run approved scenarios and record tool calls, refusals, errors, escalation, and final outputs. Flag uncertainty around normal case.

## HUMAN AUTHORITY

Test owners define expected behavior, inspect failures, and approve release criteria. Resolve the normal case question for this case.

## PILOT MEASURE

Track scenario coverage, pass rate, severe failures, regression rate, and unresolved defects. Record normal case exceptions separately.

## DISCUSSION PROMPT

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

# Test a failed supplier lookup

## STARTING FRICTION

An ordering agent must not claim an order was placed when the tool failed.

## WORKFLOW MAP

SCENARIO

EXPECTED RESULT

EXECUTION

RECEIPT

REVIEWER SCORE

FIX

## AI-SUPPORTED ROLE

Run approved scenarios and record tool calls, refusals, errors, escalation, and final outputs. Flag uncertainty around missing data.

## HUMAN AUTHORITY

Test owners define expected behavior, inspect failures, and approve release criteria. Resolve the missing data question for this case.

## PILOT MEASURE

Track scenario coverage, pass rate, severe failures, regression rate, and unresolved defects. Record missing data exceptions separately.

## DISCUSSION PROMPT

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

# Test restricted client access

## STARTING FRICTION

An agent must refuse retrieval from another client's file and produce an auditable error.

## WORKFLOW MAP

**SCENARIO**

EXPECTED RESULT

EXECUTION

RECEIPT

REVIEWER SCORE

**FIX**

## AI-SUPPORTED ROLE

Run approved scenarios and record tool calls, refusals, errors, escalation, and final outputs. Flag uncertainty around conflict.

## HUMAN AUTHORITY

Test owners define expected behavior, inspect failures, and approve release criteria. Resolve the conflict question for this case.

## PILOT MEASURE

Track scenario coverage, pass rate, severe failures, regression rate, and unresolved defects. Record conflict 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

## Normal case

Normal and edge cases are documented. Name the owner and the operating decision before selecting a tool.

02

## Missing data

Restricted requests have expected refusals or escalations. Record the approved sources, permissions, and missing information.

03

## Conflict

Tool failure behavior is tested. Test this boundary with representative and adverse cases.

04

## Restricted request

Human correction is part of the suite. Preserve review, escalation, and correction evidence.

05

## Tool failure

Regression cases run before release. Set a measurable stop condition and the next review date.

06

## Readiness to proceed

Agent testing should exercise tools, context, permissions, and interruptions. Happy-path demos do not prove operational readiness.

### CANADIAN REFERENCE POINTS

Office of the Privacy Commissioner of Canada - Privacy and artificial intelligence  
[priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/](https://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](https://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.