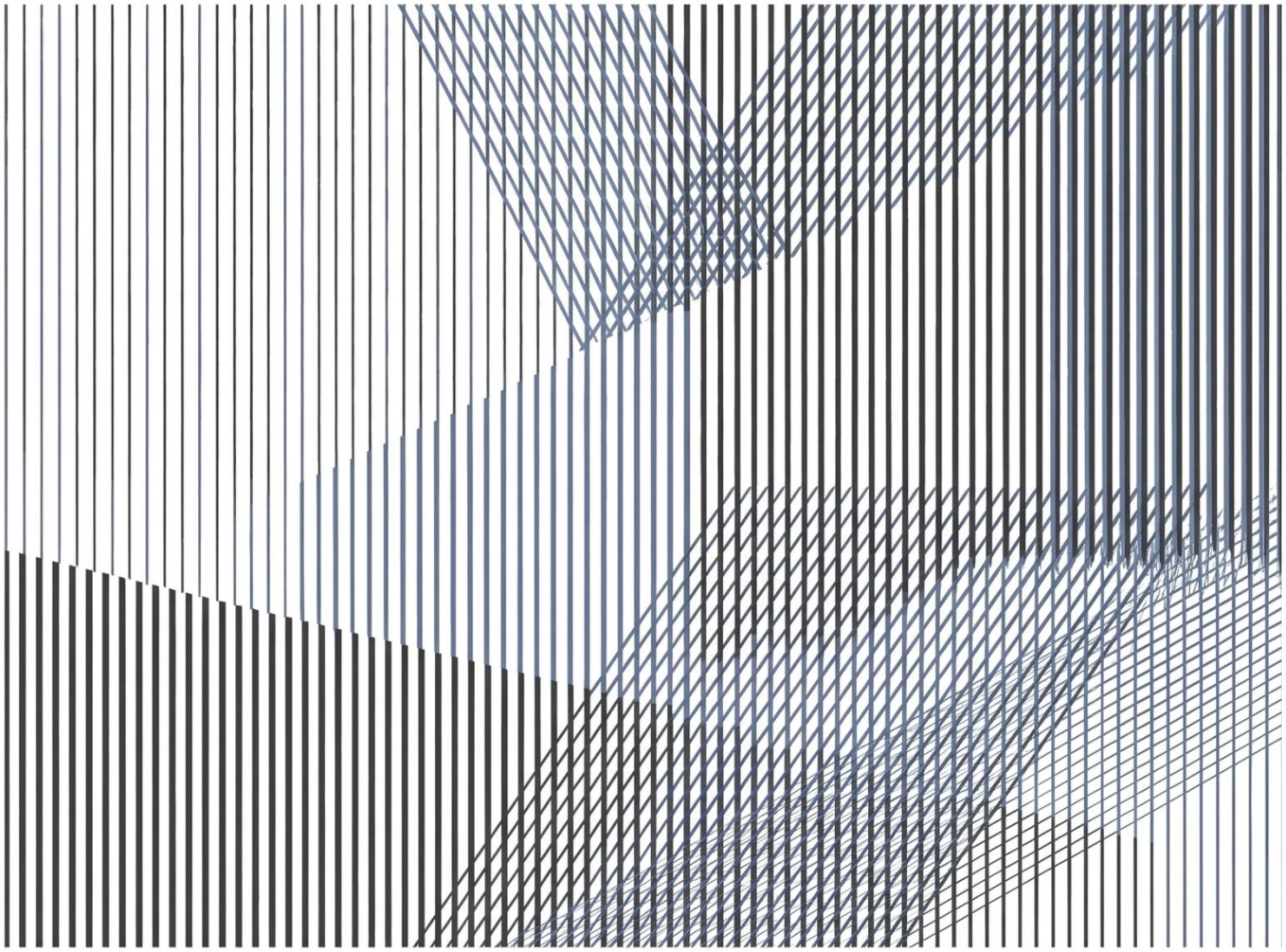
# Using AI in Public-Sector Audit

## A Guide for Audit Leaders



**C.L. Callaghan, M.Sc.**
Principal Consultant
November 2026

PolicySpark.ai

## ABOUT POLICYSPARK.AI

PolicySpark.ai is an independent consulting practice that helps public-sector organizations use evidence well. Its work covers evaluation, performance measurement, policy research and evidence synthesis for federal departments and agencies and for international development organizations. It combines established research and evaluation methods with AI-assisted tools for finding, organizing and checking evidence, so that every conclusion can be traced to its sources and judgment stays with the people accountable for it.

## ABOUT THE AUTHOR

Chris Callaghan is the founder and principal consultant at PolicySpark.ai. Trained in applied science and biology, Chris began in field ecology and ecosystem research. Over more than 25 years in federal service and consulting, Chris has worked in environmental management, public-sector audit, regulatory performance, program evaluation and science policy, bringing scientific, program and stakeholder evidence together into practical advice for decision-makers.

## HOW THIS GUIDE WAS PRODUCED

In keeping with the principles set out in this guide, AI tools assisted in preparing it, including organizing a source-linked research database, drafting working text, checking references and consistency, and supporting graphics and layout. The author defined the scope and questions, examined key sources, actively directed and edited every section, and takes sole responsibility for all judgments and conclusions. See the Production note at the end of this guide.

## SUGGESTED CITATION

Callaghan, C.L. (2026). *Using AI in Public-Sector Audit: A Guide for Audit Leaders*. PolicySpark.ai.

## DISCLAIMER

This guide presents the views of the author. It is intended to inform discussion and decisions within audit institutions and does not constitute professional audit, legal or other advice. PolicySpark.ai uses AI-assisted evidence-synthesis workflows in its own consulting practice; this guide examines source-linked evidence databases as one option alongside simpler approaches. The audit described in Section 3 and Annex A is fictional: the program, figures and outputs were constructed for this guide. Commercial and open-source tools are named as examples of a kind of tool; naming them is not an assessment or an endorsement.

© 2026 PolicySpark.ai. All rights reserved. Public-sector audit institutions may circulate this guide internally for non-commercial use with attribution.

policyspark.ai · chris@policyspark.ai

## Executive summary

THE MAIN MESSAGE

AI can strengthen public-sector audit by extending what auditors can investigate and explain, on two conditions. Institutions must be able to support the conclusions they draw from AI-assisted work, and auditors must keep the judgment needed to challenge those conclusions.

### Why this guide

Public-sector auditors are asked questions that cut across programs and organizations. The evidence needed to answer them sits in documents, data and systems that were never designed to be read together. With limited staff and time, audit teams examine only part of that evidence. A team that has found one credible explanation also tends to settle on it before fully considering what else could account for the facts.

Artificial intelligence (AI) could ease this constraint by helping teams examine more of the relevant evidence and pursue important questions further. However, AI does not change what an audit must deliver: conclusions that the institution can explain and support.

Audit institutions are already exploring AI, for tasks such as anomaly detection, document processing and risk assessment. Moving from pilot projects to routine use is proving difficult, with fragmented data, limited expertise and evolving governance among the obstacles.

This guide is written first for audit leaders who must decide how their institutions use AI, and also serves as a reference for engagement directors, methodology leads and auditors. It concentrates on performance audit, is anchored in Canadian assurance standards, statutory confidentiality and legislative independence, and draws widely on international audit practice and research. It addresses four questions:

| | | |
|---|---|---|
| 1 | What can AI contribute to audit work? | → Sections 2–3 |
| 2 | How should institutions choose where to use it and decide when its results can be relied on? | → Sections 4.1–4.2 |
| 3 | How can worthwhile use be sustained, and when should an application be extended, changed or stopped? | → Sections 4.3–4.7 |
| 4 | How can institutions keep developing the professional judgment that AI-assisted work depends on? | → Section 5 |

It draws on studies of AI in audit and related professional work, on the published plans and audit reports of audit institutions, and on a fictional performance audit that shows how the pieces fit together. Its focus is using AI in audit work rather than auditing AI systems in government, though governing internal AI use well strengthens credibility when auditing others. Exhibit 1.1 shows which sections fit each reader role so leaders can skim while audit teams use the full reference.

## What the guide finds

REACH → Sections 2 and 3

### AI's most promising contribution is reach.

Most audit applications reported so far aim at efficiency in existing work. The larger opportunity lies where audits currently stop short. AI can help auditors assemble dispersed evidence while keeping every passage and figure linked to its source. A team can then compare a whole program instead of a few files, select cases with better information and test its conclusions against more explanations where teams normally stop. The potential gain is a better-supported answer, not only a faster one, especially when teams carry what they learn from one step to the next. Wider coverage also carries a risk: in one financial-audit experiment, auditors who tested every transaction did less follow-up when outside evidence later suggested fraud.

RELIANCE → Sections 2.6 and 4.2

### How far AI can be relied on depends on the task, not only on the tool.

In experiments, AI assistance has improved speed and quality on some tasks and reduced accuracy on others. Output can be fluent, well cited and wrong, and it does not show what a search left out. The same tool can therefore suit one role and not another: screening leads tolerates more error than shaping audit scope or supporting a finding. Review is not a safeguard by itself, because reviewers can become less careful when they believe a machine prepared the work. After checking results, auditors still decide which errors could change a finding and document consequential checks on the engagement file for entity clearance and legislative scrutiny.

SUSTAINING → Sections 4.3–4.7

### Sustaining AI requires continuing investment and independent safeguards.

Once in use, an application depends on people who can maintain it, users who can interpret its results and a record of why it was approved. Recent audits of public agencies show the risks. The Australian Taxation Office lacked a central view of its AI use and baseline comparisons. In Ontario's government, staff accessed about 400 AI websites in five months (60% rated unsafe) and officials left gaps in supplier test reports unaddressed. Public audit offices must also protect confidential and Protected B (or equivalent) records under statutory secrecy, preserve legislative independence on cloud or shared infrastructure and match governance to office size. Whether an application works and whether it is worth its cost are separate questions, and evidence on retiring applications or granting autonomy remains thin.

JUDGMENT → Section 5

### Judgment has to be developed deliberately.

A well-reasoned working paper used to signal that its author understood the reasoning. With AI, a supervisor may receive a stronger product without knowing how much its author understands. Delegating work to AI could remove practice and feedback that build judgment, or free time for harder inquiry. Evidence points both ways. In one small experiment, participants who practised an unfamiliar task with AI scored lower on an unaided test. In one firm, an AI tool improved speed and quality for less experienced workers, while the most experienced saw small quality declines. No study reviewed here follows auditors over extended AI-assisted work, so the long-term effect on expertise remains unknown.

## What audit leaders can do

The guide recommends six connected actions. The first starts from the opportunity, the next four concern supporting the conclusions and the sixth concerns sustaining judgment.

| | |
|---|---|
| 01 | **Start from audit needs.** Identify where recent audits stopped short, assess each candidate application against what it would improve, and pilot it on real audit work. |
| 02 | **Keep evidence traceable to its source.** Make sure auditors can trace each conclusion to the evidence behind it. Where evidence is spread across many records, an evidence database that links every passage and figure to its source is one strong way to do so. |
| 03 | **Approve each tool for a specific task.** Match verification to what depends on the result, test for what the tool leaves out, document consequential checks on the audit file and check what reviewers actually detect. |
| 04 | **Plan for continuing support.** Budget for the people, data work and records each application needs, proportionate to office size. Keep a current view of what is in use, protect confidential audit records, preserve legislative independence and retain the expertise to assess suppliers' claims. |
| 05 | **Decide in advance how each application will be reviewed.** Set criteria and responsibilities for its future, consider the full range of responses from improving it to retiring it, set explicit limits on AI autonomy and decide what to disclose to audited bodies and legislatures. |
| 06 | **Develop people alongside the tools.** Assess the work and the person separately, examine what a task teaches before delegating it, and choose tool designs that require auditors to think, for example by asking for the auditor's judgment before showing the AI's. |

An institution that can show what its own use of AI contributes and how it is controlled is also better placed to ask the same of the bodies it audits.

### Where to begin

An institution does not need all six actions in place before it starts. A reasonable first step is a single application:

- chosen for an audit need;
- built on a defined body of evidence;
- approved for a specific task; and
- accompanied by a record of what it contributes and what verification it requires.

Writing assistance, AI search for scoping and uses that leave the auditor's own reading in place (such as summarizing minutes that auditors still read) are sensible early steps. Uses that extend what auditors can examine deserve a place in the office's plans from the start.

Arrangements can match office size. A small office may need only a short list of approved uses, a named lead, periodic review and peer sharing with other independent audit offices. Systems with wide autonomy can wait: no audit institution consulted by the OECD yet uses them in practice.

## What remains uncertain

Three questions cannot yet be answered from the available evidence:

– how well AI performs on audit work itself;
– which combination of checks works for a particular AI application in audit; and
– how extended use of AI affects audit expertise over a career.

An institution's own experience is therefore its most relevant evidence. By recording what each application contributes, what it costs and what their staff can do, institutions can answer these questions for their own work and help others answer them too.