



## ETHICAL AI · GOVERNANCE

# Accountability before automation

Five safeguards for public institutions adopting artificial intelligence

**Dr Simon Aseno, MPH**

President and CEO  
Praxis Global Institute

## ABOUT THIS PUBLICATION

## Publication details

**Dr Simon Aseno, MPH**

President and CEO, Praxis Global Institute

Dr Aseno leads Praxis Global Institute's work connecting research, technology, education and advisory to strengthen people, systems and institutions.

**Corresponding author.** For enquiries about this brief, email [simon@praxisglobal.ac](mailto:simon@praxisglobal.ac).

**ABOUT THIS BRIEF**

Praxis Policy Briefs distil evidence on a current institutional challenge into practical recommendations for leaders, boards and policymakers. Each brief is short enough to read before a meeting, cites every source so that readers can check it, and ends with actions an institution can take within one budget cycle.

This brief examines what happens to accountability when public institutions adopt artificial intelligence. It draws on three kinds of evidence: court and tribunal decisions and the public reporting of them; published research on how organisations are adopting generative AI; and the four international governance frameworks most often cited by governments, regulators and funders, from UNESCO, the US National Institute of Standards and Technology (NIST), the World Health Organization (WHO) and the European Union. Sources were checked against original or authoritative secondary sources as at August 2026 and are listed in full at the end.

**SUGGESTED CITATION**

Aseno, S. B. (2026). *Accountability before automation: Five safeguards for public institutions adopting artificial intelligence* (Praxis Policy Brief No. 1, PGI-PB-2026-01). Praxis Global Institute. <https://doi.org/10.5281/zenodo.23307277>

**RIGHTS AND USE**

© 2026 Simon Aseno. Published by Praxis Global Institute under a Creative Commons Attribution 4.0 International licence (CC BY 4.0): you may share and adapt this brief, including for commercial purposes, with attribution.

**DISCLAIMER**

This brief is for information and does not constitute legal advice. Laws, dates and provisions cited were checked against the sources listed and may change. Views expressed are the author's.

**Permanent record and citation**

Scan to open this brief's permanent record, download the latest version and export the citation.

**DOI: 10.5281/zenodo.23307277**

**SERIES**

Praxis Policy Brief Series · No. 1

**PUBLICATION ID**

PGI-PB-2026-01 · Version 1.0

**PUBLISHED**

August 2026 · North Carolina, USA · Praxis Global Institute

**ENQUIRIES**

[simon@praxisglobal.ac](mailto:simon@praxisglobal.ac) · [praxisglobal.ac/contact](https://praxisglobal.ac/contact)

## SUMMARY

## Abstract

Public institutions are adopting generative artificial intelligence (AI) to draft documents, summarise evidence, answer enquiries and support decisions. Responsibility for the results does not transfer to the software. Standard terms for widely used AI services disclaim warranties and cap the provider's liability, and courts and tribunals have held the professionals and organisations that relied on AI accountable for its errors.

This brief reviews three widely reported cases from the United States, Canada and Australia, evidence that adoption is running ahead of results, and four international frameworks that now shape expectations: the UNESCO Recommendation on the Ethics of AI, the NIST AI Risk Management Framework, WHO guidance on large multi-modal models and the EU AI Act. It finds that all four converge on human oversight, transparency, risk management and clear accountability.

It recommends five safeguards that public institutions should put in place before deployment: a named owner for every AI use; a risk-tiered register; human review of consequential outputs; contracts that allocate liability; and capability building with continuous monitoring. A checklist and a 90-day plan help leaders turn these into practice.

## AT A GLANCE

## Key messages

- 1. Responsibility does not transfer to software.** Standard terms for widely used AI services disclaim warranties and cap the provider's liability. When an AI-assisted output causes harm, the institution that relied on it remains answerable.
- 2. Courts and tribunals already say so.** In 2023 a US federal court sanctioned lawyers who filed fictitious, AI-generated case law; in 2024 a Canadian tribunal held an airline liable for its chatbot's wrong advice.
- 3. Adoption is running ahead of results.** A 2025 MIT study found that about 95% of corporate generative AI pilots delivered no measurable financial return.
- 4. The global rulebook is converging.** UNESCO (2021), the US NIST framework (2023), WHO (2024) and the EU AI Act (2024) all call for human oversight, transparency, risk management and clear accountability.
- 5. Five safeguards close most of the gap:** a named owner for every AI use; a risk-tiered register; human review of consequential outputs; contracts that allocate liability; and capability building with monitoring.

## INSIDE

## Contents

|    |                                                    |    |
|----|----------------------------------------------------|----|
| —  | Abstract and key messages                          | ii |
| 01 | Introduction                                       | 1  |
| 02 | The accountability gap, and what recent cases show | 2  |
| 03 | A converging global rulebook                       | 3  |
| 04 | Five safeguards before deployment                  | 4  |
| 05 | Implementation checklist                           | 6  |
| 06 | Conclusion: from principle to practice             | 7  |
| —  | References                                         | 7  |

## WHO THIS BRIEF IS FOR

- **Ministers and senior officials** setting policy and standards for the use of AI across government.
- **Heads of public agencies, universities and hospitals** who answer for AI-assisted services to citizens, students and patients.
- **Boards and audit and risk committees** overseeing how AI risk is identified, owned and reported.
- **Procurement, legal, data, digital and information-governance teams** who write the contracts, policies and registers that make accountability work.
- **Development partners and funders** who expect responsible AI in the programmes they support.

## ABOUT PRAXIS GLOBAL INSTITUTE

Praxis Global Institute's mission is to deliver practical training, applied research, ethical AI and technology solutions, and strategic advisory that improve performance, strengthen systems and empower professionals.

Its work is organised in four connected areas: **Research & Data**, which builds evidence and data infrastructure for better decisions; **Technology & Innovation**, which designs ethical AI and digital systems around real workflows; **Education & Training**, which develops professional capability that carries into practice; and **Advisory & Partnerships**, which supports strategy and institutional change.

Learn more at [praxisglobal.ac](https://praxisglobal.ac).

## HOW TO USE THIS BRIEF

Read the abstract and key messages for the argument in two minutes. Sections 02 and 03 give the evidence; Section 04 sets out the five safeguards; Section 05 gives a checklist for leadership and board meetings and a 90-day plan to get started; and Section 06 concludes.

# Introduction

Institutions that once asked whether to use AI now ask how to use it well. This brief asks a prior question: when AI-assisted work goes wrong, who answers for it?

Artificial intelligence (AI) has moved from pilot projects into everyday work in public institutions. Generative tools draft letters, reports and policy notes; summarise consultations and research; answer questions from citizens, students and patients; and help staff prepare decisions about services. The appeal is easy to see: faster turnaround, wider reach and relief for teams that are stretched thin.

Debate about these tools is usually framed around efficiency or risk. This brief frames it around **accountability**: the ability to say who is responsible for an output, to explain how it was produced, and to correct it when it is wrong. Accountability is what allows mistakes to be found and fixed, and what allows citizens to trust that a decision made about them can be explained and challenged.

The argument is simple. Responsibility for AI-assisted work stays with the institution that uses it, whatever its contracts or its tools say. Institutions should therefore build accountability into AI adoption *before* deployment, rather than discover its absence after an incident.

Section 02 sets out the accountability gap and three cases in which professionals and organisations were held responsible for AI errors. Section 03 shows how four international frameworks converge on the same expectations. Section 04 recommends five safeguards; Section 05 turns them into a checklist and a 90-day plan; and Section 06 concludes.

## Scope

The brief is written for public institutions, but its safeguards apply equally to universities, hospitals, non-profit organisations and companies that deliver public

services. It addresses generative AI and AI-assisted decision support. It does not cover military or national-security uses of AI, or the technical evaluation of AI models.

## Approach

The brief is based on a review of publicly available sources: court and tribunal decisions and authoritative reporting of them, published research on AI adoption, and the texts and official summaries of international frameworks and legislation. Each factual statement is referenced, and sources are listed in full at the end.

### KEY TERMS

#### Generative AI

AI systems that produce new text, images or other content in response to prompts, such as chatbots and writing assistants.

#### AI-assisted output

Any document, answer, summary or recommendation produced wholly or partly with an AI tool.

#### Accountability

A named person or body is answerable for an output or decision, can explain it and can put it right.

#### Deployer

Under the EU AI Act, an organisation or public authority that uses an AI system under its own authority, as distinct from the provider that develops it.

## The accountability gap, and what recent cases show

AI can draft the report, answer the citizen and summarise the evidence. It cannot be held to account for any of it. People and institutions can, and are.

**Table 1. Who was held responsible when AI-assisted work went wrong**

| Case                                                                                    | Year | What happened                                                                                                                                    | Outcome                                                                                                                                      |
|-----------------------------------------------------------------------------------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Mata v. Avianca</b><br>(US District Court, S.D.N.Y.)                                 | 2023 | Lawyers filed a brief citing court decisions that ChatGPT had invented.                                                                          | The lawyers and their firm were sanctioned and fined US\$5,000.                                                                              |
| <b>Moffatt v. Air Canada</b><br>(BC Civil Resolution Tribunal)                          | 2024 | The airline's website chatbot gave a bereaved passenger wrong information about bereavement fares.                                               | The airline was held liable for negligent misrepresentation; its argument that the chatbot was responsible for its own actions was rejected. |
| <b>Deloitte report for Australia's Department of Employment and Workplace Relations</b> | 2025 | An AU\$440,000 review contained fabricated references and a made-up quote from a judge; a revised version disclosed use of a generative AI tool. | The firm agreed to repay the final instalment of its fee.                                                                                    |

### The accountability gap

Public institutions are adopting generative AI to draft documents, summarise evidence, answer enquiries and support decisions. The promise is real. The risk is less visible. When an AI-assisted output is wrong, someone must answer for it, and it will not be the software.

Consider the terms on which many institutions use these tools. OpenAI's terms of use, effective 1 January 2026, state that its services are provided "as is" and cap the company's aggregate liability at the greater of the amount paid for the service in the previous twelve months or US\$100. Enterprise agreements can be negotiated differently, but the default position is clear: the vendor disclaims the outcome and the user carries it.

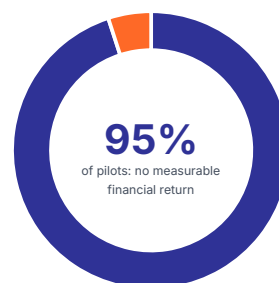
**AI changes who does the work. It does not change who answers for it.**

### What recent cases show

Three widely reported cases from three jurisdictions point the same way (Table 1). In each, the AI tool was not the party held responsible. The professionals and organisations that relied on it were.

The Deloitte case also shows that accountability works when it is designed in: a researcher spotted the errors, the press reported them, the commissioning department

acted and the firm repaid part of its fee. Without a named, answerable party, none of that would have been possible.



**Figure 1.** Share of corporate generative AI pilots reporting no measurable financial return. Source: MIT NANDA, *The GenAI Divide: State of AI in Business 2025* (150 interviews, a survey of 350 employees and 300 public deployments).

Results are also lagging behind adoption. The MIT study attributed most failures not to model quality but to a "learning gap" between tools and the way organisations actually work, and found that tools bought in from specialised vendors and partners succeeded roughly twice as often as those built entirely in-house.

### Why this matters for public institutions

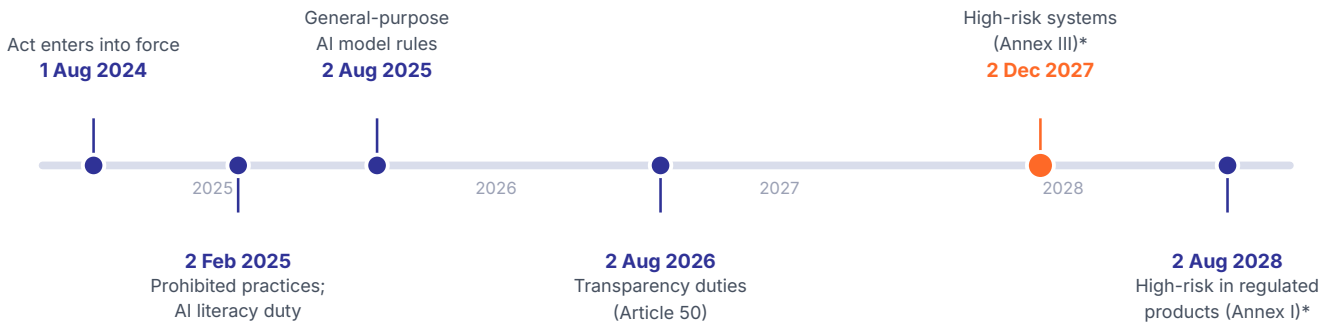
For a public body the stakes are higher than for a firm. Its decisions affect people who often cannot choose another provider, its errors are paid for with public money, and its legitimacy rests on being able to explain what it did

and why. An institution that cannot say who approved an AI-assisted decision cannot defend it to a court, an auditor or the person affected.

03

A converging global rulebook

Different institutions, one direction: human oversight, transparency, risk management and accountability.



**Figure 2.** EU AI Act: key application dates. \*High-risk dates as amended by Regulation (EU) 2026/1744 (Digital Omnibus), in force 27 July 2026. Sources: Regulation (EU) 2024/1689; EU legislative summaries (see references).

**Table 2.** Four frameworks, one set of expectations

| Instrument                                       | Status                       | Human oversight                  | Transparency                     | Risk management             | Accountability                         |
|--------------------------------------------------|------------------------------|----------------------------------|----------------------------------|-----------------------------|----------------------------------------|
| UNESCO Recommendation on the Ethics of AI (2021) | Normative, 193 member states | Humans keep final responsibility | Explainability and disclosure    | Impact assessment           | Oversight, audit, redress              |
| NIST AI RMF 1.0 (2023)                           | Voluntary guidance           | Roles defined under Govern       | Transparent, explainable systems | Map, Measure, Manage        | Accountable as a trustworthiness trait |
| WHO guidance on large multi-modal models (2024)  | Guidance for health          | Human oversight of clinical use  | Disclosure to users              | Pre-approval by a regulator | Post-release audits, published results |
| EU AI Act (2024)                                 | Binding law, phased          | Required for high-risk systems   | Article 50 disclosure duties     | Risk-management system      | Provider and deployer obligations      |

Four instruments now shape how institutions everywhere are expected to govern AI, whether or not they are legally bound by them. Their language differs; their core expectations do not (Table 2).

**UNESCO's Recommendation on the Ethics of Artificial Intelligence**, adopted by its 193 member states in November 2021, was the first global normative

framework on AI ethics. It places human rights, human oversight and accountability at the centre and asks states to report on implementation.

**The NIST AI Risk Management Framework 1.0**, released by the US National Institute of Standards and Technology in January 2023, is voluntary guidance organised around four functions: *Govern, Map, Measure and Manage*.

**The World Health Organization's guidance on large multi-modal models**, published in January 2024, sets out more than 40 recommendations for governments, developers and providers, including mandatory post-release auditing and impact assessments of AI used in health.

**The EU AI Act** (Regulation (EU) 2024/1689) entered into force on 1 August 2024 and applies in stages (Figure 2, above). Its prohibitions and AI-literacy measures have applied since February 2025, and rules for general-purpose AI models since August 2025. Following the 2026 Digital Omnibus amendments (Regulation (EU) 2026/1744), obligations for high-risk systems, which include uses in education, employment and access to essential public services, now apply from December 2027, and from August 2028 for AI built into regulated products.

Suppliers serving several markets increasingly build to the strictest common standard, so institutions outside Europe that adopt these safeguards now will find procurement, audit and partnership with international funders easier later.

## 04

## Five safeguards before deployment

Practical steps any public institution can take within one budget cycle, whatever its size.



Safeguards are numbered as in the text; their order of application follows the life of an AI use.

**Figure 3.** The five safeguards across the life of an AI use. Human review (safeguard 3) is the point at which most of the errors in Table 1 would have been caught.

## 1 Name an accountable owner for every AI use

Each tool and use case should have a senior owner who answers for its outputs, approves its scope and can switch it off. Record the owner in the register (safeguard 2). Accountability that belongs to "the system" belongs to no one.

## 2 Keep a risk-tiered register of AI uses

List every AI use, its purpose, data, users and the people it affects. Classify each as low, medium or high risk, using the EU AI Act's high-risk areas (education, employment, essential services, law enforcement) as a reference point. High-risk uses need an impact assessment before go-live; this mirrors the NIST *Map* function.

## 3 Require human review of consequential outputs

No AI-generated content should leave the institution, or shape a decision about a person, without review by a qualified person who has checked facts, figures, quotations and citations against original sources. Two of the three cases in Table 1 involved documents a careful source check would have caught. The third shows that public-facing chatbots need tested, approved knowledge sources and a clear route to a person.

## 4 Contract for accountability

Procurement should not accept "as is" terms for consequential uses. Seek performance commitments, data-protection terms, audit and information rights, incident-notification duties, a clear allocation of liability and an exit plan, and ask suppliers to disclose when they use generative AI in work they deliver to you.

## 5 Build capability and monitor continuously

Train staff to use, question and verify AI outputs; the EU AI Act asks organisations to take measures to support AI literacy. Keep an incident log, review high-risk uses at least annually, and publish summaries of audits, as WHO recommends for health.

### WHAT THE SAFEGUARDS HAVE IN COMMON

None requires new technology or a large budget. Each makes responsibility visible: who owns a use, how risky it is, who checked an output, what the supplier has promised, and whether staff know how to question what the tool produces.

## Five questions for boards

1. Which AI tools are in use across the institution today, and who owns each one?
2. Which of those uses could affect a decision about a person, and have they been assessed?
3. What do our supplier contracts say about liability, audit and incidents?
4. How do we know that AI-assisted documents leaving the institution have been checked?
5. When did management last report AI incidents and reviews to this board?

A board that cannot answer these questions has an accountability gap, whatever its AI policy says on paper.

## Implementation checklist

Use this page in a leadership or board meeting. Each "no" is a gap to close before the next AI deployment.

| Stage  | Question                                                                                                | Safeguard      | Yes / No |
|--------|---------------------------------------------------------------------------------------------------------|----------------|----------|
| Before | Is there a named senior owner for this AI use?                                                          | 1 · Owner      |          |
|        | Is it recorded in the AI register with a risk tier, and has a high-risk use had an impact assessment?   | 2 · Register   |          |
|        | Do supplier terms cover performance, data protection, audit, incident notification, liability and exit? | 4 · Contract   |          |
| During | Is every consequential output reviewed by a qualified person, with sources checked?                     | 3 · Review     |          |
|        | Have users been trained to question and verify outputs, and do they know how to report problems?        | 5 · Capability |          |
|        | Are people told when they are dealing with an AI system?                                                | 3 · Review     |          |
| After  | Is there an incident log, and are incidents reviewed by the owner?                                      | 5 · Monitor    |          |
|        | Is each high-risk use reviewed at least annually, with a summary published?                             | 5 · Monitor    |          |

### Getting started: the first 90 days

**Table 3.** A 90-day plan to put the safeguards in place

| Days  | Actions                                                                                                                                                             | Result                             |
|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| 1–30  | Ask each department to list the AI tools it uses; appoint a senior owner for each use; nominate one executive to lead AI governance.                                | A first register with named owners |
| 31–60 | Classify each use by risk; complete impact assessments for high-risk uses; set a rule that consequential outputs are reviewed by a qualified person before release. | Risk tiers and a review procedure  |
| 61–90 | Review supplier contracts against safeguard 4; train users; open an incident log; report the register and gaps to the board.                                        | A baseline the board can monitor   |

## Conclusion: from principle to practice

Accountability is cheapest to build before deployment and most expensive to discover after an incident.

The institutions that benefit most from AI will not be those that adopt it fastest, but those that can show, at any moment, who is responsible for what it produces. The five safeguards in this brief are deliberately modest. Together they make that responsibility visible, and they align any institution with the direction of international law and guidance.

### Three common objections

**"We are too small for this."** The safeguards scale down. A small agency's register can be a single spreadsheet, and its owner can be the head of the unit using the tool. What matters is that someone is named.

**"Our vendor is responsible for the tool."** The vendor is responsible for what it promised in its contract, and default terms promise very little. The institution remains answerable to the people its decisions affect.

**"These laws do not apply to us."** Perhaps not yet. But suppliers, funders and partners increasingly expect the same standards, and the principles in Table 2 describe good governance wherever an institution operates.

### How Praxis can help

Praxis Global Institute works with public institutions to put these safeguards in place: designing AI registers and review workflows through our [Technology & Innovation](#) practice, training leaders and staff through [Education & Training](#), and advising boards and ministries through

[Advisory & Partnerships](#). For enquiries about this brief, email Dr Aseno at [simon@praxisglobal.ac](mailto:simon@praxisglobal.ac), or [contact Praxis](#).

### References

1. OpenAI (2026). Terms of Use, effective 1 January 2026. [openai.com/policies/row-terms-of-use](https://openai.com/policies/row-terms-of-use)
2. Ambrogi, R. (2023). Court imposes sanctions on lawyers who filed bogus cases after relying on ChatGPT. LawSites. [lawnext.com](https://lawnext.com)
3. American Bar Association (2024). BC Tribunal confirms companies remain liable for information provided by AI chatbot. Business Law Today. [americanbar.org](https://americanbar.org)
4. Associated Press (2025). Deloitte to partially refund Australian government for report with apparent AI-generated errors, 7 October 2025. [news4jax.com](https://news4jax.com)
5. MIT NANDA (2025). The GenAI Divide: State of AI in Business 2025, as reported by Computing. [computing.co.uk](https://computing.co.uk)
6. UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. [unesco.org](https://unesco.org)
7. National Institute of Standards and Technology (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. [doi.org/10.6028/NIST.AI.100-1](https://doi.org/10.6028/NIST.AI.100-1)
8. World Health Organization (2024). WHO releases AI ethics and governance guidance for large multi-modal models, 18 January 2024. [who.int](https://who.int)
9. European Union (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). [eur-lex.europa.eu](https://eur-lex.europa.eu)
10. Gibson Dunn (2026). EU AI Act omnibus agreement: postponed high-risk deadlines and other key changes (Regulation (EU) 2026/1744, Digital Omnibus on AI). [gibsondunn.com](https://gibsondunn.com)



## Ideas into action

Praxis Global Institute delivers practical training, applied research, ethical AI and technology solutions, and strategic advisory that improve performance, strengthen systems and empower professionals.

---

### Research & Data

Evidence and data infrastructure for better decisions.

---

### Technology & Innovation

Ethical AI and digital systems built around real workflows.

---

### Education & Training

Professional capability that carries into practice.

---

### Advisory & Partnerships

Strategy and partnership for institutional change.

Praxis Policy Briefs are a numbered series of citable publications, each with a permanent digital object identifier (DOI).



DOI 10.5281/zenodo.23307277 · PGI-PB-2026-01