Trust Imperative 5.0

Building Trust in Government Through Practical AI Assurance

2026 Global Report

Note: This photo was generated with AI

Executive Summary

Artificial intelligence (AI) has the potential to help governments improve service quality, achieve better outcomes, raise public sector productivity, and build greater trust in government. Across countries, governments have responded to advances in AI capabilities by introducing governance and assurance frameworks, principles, risk assessment tools, transparency requirements and registers, and procurement controls. Together, these measures are intended to support responsible AI adoption while helping organizations manage risks associated with AI. But how effectively do they work in practice? What works well, what isn't working, and what lessons can be drawn from the experience of real practitioners so far?

Across the ten countries analyzed for this report, we found that many government organizations are grappling with common challenges: how to classify risk, what evidence is sufficient, who owns the decision, and what "safe enough" means in different AI use cases. When the answers to these questions are unclear, AI assurance processes often lead to delays, rework, and duplication.

BCG research finds that while most governments have established foundational principles and frameworks for working with AI, putting them into practice is far more challenging. Many governments have named accountabilities, tiered risk categories, practical transparency tools, and strong governance structures. But risk thresholds are often too vague to apply consistently. Accountable roles exist, but not always with clear mandates or decision rights. Processes remain fragmented and duplicative, and internal capabilities at the points of deployment and usage can be uneven. As a result, many lower-risk productivity use cases are subject to over governance, while more complex questions involving deployment, platforms, and shared responsibility remain unresolved.

These challenges may grow even more acute because many current AI frameworks were designed before the emergence of today's advanced frontier models and agentic AI use cases. These tools are more capable and accurate, but they are also more autonomous and interconnected, with shared responsibility across model providers, platforms, partners, and agencies. This situation leads to friction across risk assessment, procurement, deployment, and oversight. It also shifts the governance question. Particularly for agentic systems, the issue is not just whether a model is technically safe, but what authority is being delegated, which responsibilities must remain with a human, and how accountability is maintained as systems act within defined limits.

Rather than adding another layer of policy complexity, governments should focus on making existing assurance more usable. The goal is to maximize public value while staying disciplined enough to manage risk and scale AI responsibly. The opportunity cost of delay is significant. BCG estimates that GenAI could unlock $1.75 trillion in annual productivity benefits by 2033 for governments globally, but realizing that value depends on adoption proceeding without high-profile failures that erode public trust and trigger regulatory backlash.

To navigate this transition effectively, governments can make risk triage more proportionate, clarify accountability, redesign assurance for generative AI (GenAI) and agentic systems, embed assurance into delivery, create reusable artifacts, strengthen capability, and measure whether assurance is helping manage both risk and opportunity. With this approach, assurance can become an enabler of responsible AI adoption at scale. Governments that get this balance right will be better positioned to adopt AI confidently while enhancing public trust in the process.

01.

Balancing the AI Opportunity and Risks

Effective assurance is intended to help governments capture the value of AI while maintaining public trust and accountability. In many government and public sector agencies, AI ethics principles have been published, procurement guidance has been updated, and various forms of impact assessment, transparency recording, or risk review have been introduced. However, the public sector is lagging other industry sectors in terms of AI adoption and AI use cases. Are governments becoming overly conservative in their assessment of risk, limiting productivity gains and delaying service improvements? And are current assurance frameworks striking the right balance between managing risk and enabling public value? Is this really a trade-off or can you have both?

This report examines the evolution of AI risk, assurance, and governance frameworks in government and whether they are effective and efficient in practice. It also highlights what is working well in AI assurance and where further refinement could improve outcomes.

Key Focus of this Report

01. What is the purpose of AI principles and assurance frameworks and are they achieving their intended objectives?

02. What has been the experience of people in government who are applying AI assurance frameworks?

03. What impact are these frameworks having on governments' adoption and utilization of AI?

04. What are the lessons learned? What has worked well and what hasn't?

05. What improvements could governments introduce to make AI principles and assurance frameworks more effective?

Note: This photo was generated with AI

02.

What Has Happened Already

About this Report

This report is based on publicly available documents and research on AI adoption, trust and assurance analysis, prior work in our Trust Imperative series, and interviews with practitioners and experts in ten countries: Australia, Canada, France, Germany, India, Italy, Japan, Singapore, the UK, and the US. We examined national, subnational, and organizational AI governance frameworks, policies, and recent legislative developments to understand how assurance is structured. To complement this work, we conducted approximately 20 interviews with officials, regulators, vendors, and subject matter experts across the surveyed countries. The interviews provided illustrative and qualitative perspectives. We analyzed different approaches across countries and sectors, as well as across public and private organizations, to identify differences and recurring patterns in areas such as operational friction, risk classification, accountability, and vendor interactions. We then iteratively refined our findings through peer and expert review, focusing on practical insights and lessons learned.

All of the governments that we looked at in detail already have some combination of AI principles, named accountable roles, governance processes and committees, risk-based assessments, transparency requirements, and procurement controls. (See "About this Report.") These components typically sit in a layered stack: principles and ethical commitments at the top; governance policies and frameworks in the middle; and operational tools such as risk assessments, registers, templates, and workflows at the bottom. (See Exhibit 1.) The main architecture is already present, widely recognized, and established in both policy and academic literature.[1]

The following sections summarize the major elements of each country's architecture for AI. (See the Appendix for a complete list of each country's applicable frameworks.)

Australia

In alignment with its tiers of government, Australia has a national AI governance and assurance framework as well as state-level frameworks. The federal framework includes the AI Ethics Principles (2019), the National Framework for the Assurance of AI in Government (2024), the APS AI Plan (2025), the Responsible AI Policy update, the AI Impact Assessment Tool, and procurement guidance.[2] Each state or territory also has AI risk assurance frameworks of varying sophistication, with New South Wales and Western Australia offering the most comprehensive and mature assurance-style models.

Canada

Canada has a clear federal baseline, with variations in operational delivery at the departmental and provincial levels. The federal backbone is the Directive on Automated Decision-Making and the associated Algorithmic Impact Assessment, which has been updated several times since 2019.[3] Federally, the broader model remains policy-led rather than regulatory, with

built-in flexibility in case departments justify adopting an alternative path. The country's recent GenAI guidance has also been intentionally enabling in tone, permitting use while adding practical guardrails around data handling, transparency, and accountability.[4] In 2024, Ontario began moving from principles toward more defined AI governance, supported by the Enhancing Digital Security and Trust Act.[5]

France

France's approach to AI assurance reflects strong alignment with the EU-level direction. The EU AI Act anchors the framework, but France is also preparing a national plan to assign oversight responsibilities, with the Directorate-General for Competition, Consumer Affairs and Fraud Control, and the Directorate-General for Enterprise playing key coordinating roles, particularly in monitoring AI systems in the market.[6]

Germany

In Germany, the Federal Ministry of the Interior's AI Guidelines for the Federal Administration set common guardrails for AI use.[7] In 2025, the federal administration launched an AI Opportunity Market and a transparency database to surface existing and planned use cases.[8] The BSI (Federal Office for Information Security) published criteria catalogues for integrating externally provided GenAI models into federal applications and for assessing AI cloud services.[9] The German data protection authorities have issued AI and data protection guidance.[10] Germany's digital sovereignty and government cloud agenda focus on reducing reliance on individual providers, strengthening data protection and security, and defining clear conditions for how systems are deployed.[11]

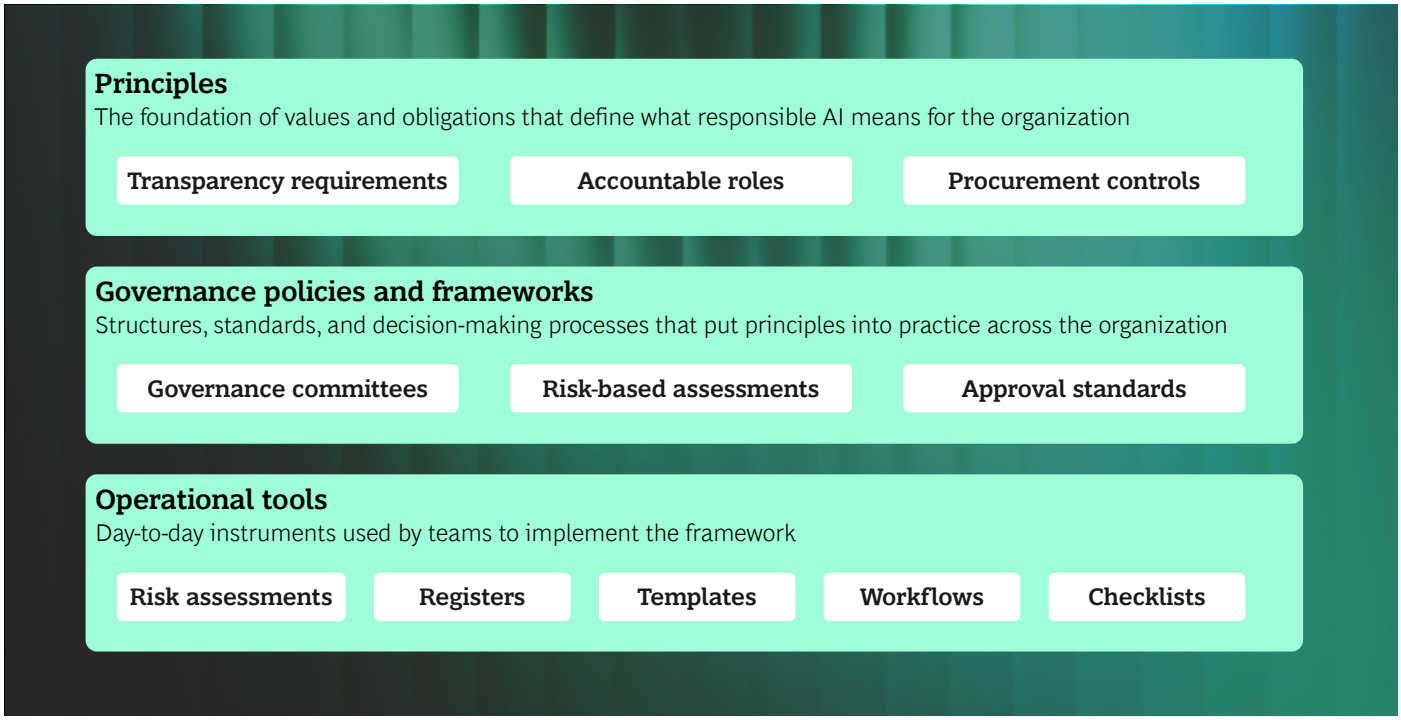
India

India combines a national model with state-level implementation adapted to specific contexts. At the national level, the IndiaAI Mission under MeitY is the primary vehicle for India's AI agenda, including governance guidelines, computing capacity, platform development, and ecosystem coordination. The National Institution for Transforming India (NITI Aayog) contributed to earlier responsible AI policy thinking.[12] The IndiaAI Governance Guidelines and a January 2026 techno-legal white paper are key documents, while the February 2026 IT (Intermediary Guidelines) Amendment Rules add a layer for platform and content governance.[13] Several states are setting up their own AI missions, reflecting a model of central coordination alongside state-level execution.

Italy

Italy's AI governance is a multilayered model in which national strategy, public administration guidance, and

Exhibit 1 - A Three-layer Architecture Is Widely Recognized in Policy and Practice



domestic law expand on the wider EU regime. Italy's AI Strategy defines actions across research, public administration, enterprise, and skills. AgID's guidance on AI in public administration focuses on navigating the current legislative landscape and anticipating future regulatory change.[14] The Department of Public Administration's training program, Adottare l'Intelligenza Artificiale nella PA, addresses capability-building.[15] Other relevant guidance appears in the EU AI Act, the Digital Operational Resilience Act, sectoral controls, and Italy's Law 132/2025.[16]

Japan

In Japan's model, national guidance sets direction and ministries apply it in context. The system emphasizes proportionate oversight, with streamlined pathways for lower-risk cases and structured escalation for higher-risk matters. National guidance, such as the AI Guidelines for Business, the Hiroshima Process principles, and associated government frameworks, establishes the overall direction.[17] In practice, ministries can use risk-based tools and models developed by local chief AI officers for lower-risk cases, while escalating higher-risk matters for additional review and advice.

Singapore

Singapore takes a centrally coordinated, tool-oriented approach to AI assurance. The government published its Model AI Governance Framework in 2019 and updated it in 2020 specifically to translate principles into practical organizational guidance.[18] The nation added A.I. Verify in 2022 as a testing framework and toolkit, published a Model AI Governance Framework for Generative AI in 2024, and launched a Model AI Governance Framework for Agentic AI in 2026.[19] Singapore pairs centralized AI policy with practical testing, self-assessment, and deployment guidance.

UK

The UK has adopted a pro-innovation model that relies on existing regulators and practical transparency tools, combining pragmatic integration with existing legal frameworks rather than creating a wholly separate AI regime. Frameworks such as the Algorithmic Transparency Recording Standard (ATRS) and the Data (Use and Access) Act 2025 are examples of this approach.[20] The ATRS hub publishes searchable records of algorithmic tools used across government, and its newer mandatory scope demonstrates how the UK uses transparency as an implementation lever to support high-level AI principles.

US

In the US, activity at the federal and state levels has created many mechanisms for assurance. The federal level is governed by frameworks such as the National Institute of Standards and Technology (NIST) AI Risk Management Framework, Office of Management and Budget memoranda on AI governance and inventories, and in defense and security settings, the Department of Defense Responsible AI Strategy and Implementation Pathway.[21] States such as California, Colorado, New Jersey, New York, and Texas have introduced their own operational and legal frameworks.[22]

Three Broad Categories of AI Risk Management

From these country examples, three broad patterns emerge. The first is the federal or devolved model, in which national baselines exist but practices vary across departments, states, or provinces. Australia, Canada, the US, and India all exhibit versions of this. In these settings, individual agencies often have significant autonomy in operationalizing national expectations. Although that structure can offer flexibility, it can also create variability in role design, threshold setting, and execution maturity.

The second category of AI risk management is the centrally coordinated, tool-oriented model. In these settings, the state is more explicit about how to translate principles into tooling, testing, and deployment practices. Singapore offers the clearest example of this approach, while Japan combines guidance, structured escalation pathways, and ministry-level implementation. Central guidance can make expectations easier to apply, but it depends on having guidance and tooling evolve alongside advances in AI capabilities.

The third category is the layered regulatory model, in which AI governance sits within a broader stack of legal, regulatory, and sector-specific obligations. Germany, Italy, and France illustrate this model most clearly. The UK follows its own transparency-led model grounded in existing regulators rather than establishing a standalone AI regulator. Often these governments have well-developed formal structures in place, illustrating how implementation can become more nuanced as multiple intersecting layers of law, compliance, and operational guidance interact.

Overall, we found that most governments have the core building blocks of AI governance in place at a high level, but challenges emerge in implementation at the operational level.

Note: This photo was edited using AI

03.

What Works and What Does Not

In our research, we found that some elements of AI assurance are working well but others are not. This section examines each of these elements in more detail.

Principles, Frameworks, and Governance

What Works: Principles, Frameworks, and Governance Are Now Widely in Place

Most governments already have well-established AI principles, assurance frameworks, and governance mechanisms. Across countries, the objective is broadly consistent: unlock the economic and societal value of AI adoption while preserving public trust, maintaining accountability, and reducing the possibility of legal, operational, and reputational harm.

For example, Canada's 2019 Directive on Automated Decision-Making and Algorithmic Impact Assessment created a federal backbone for administrative decision systems.[23] Singapore updated its Model Governance Framework in 2020, added A.I. Verify in 2022, and adopted separate agentic AI guidance in 2024 and 2026.[24] Germany has issued federal AI guidelines, BSI criteria catalogues, and a transparency database.[25] Australia's federal framework includes AI ethics principles, an assurance framework, an impact assessment tool, and procurement guidance.[26] Even where implementation is uneven, the existence of this architecture is meaningful. Most governments know broadly what good governance should include.

What Is Not Working: Processes Are Too Fragmented and Duplicative

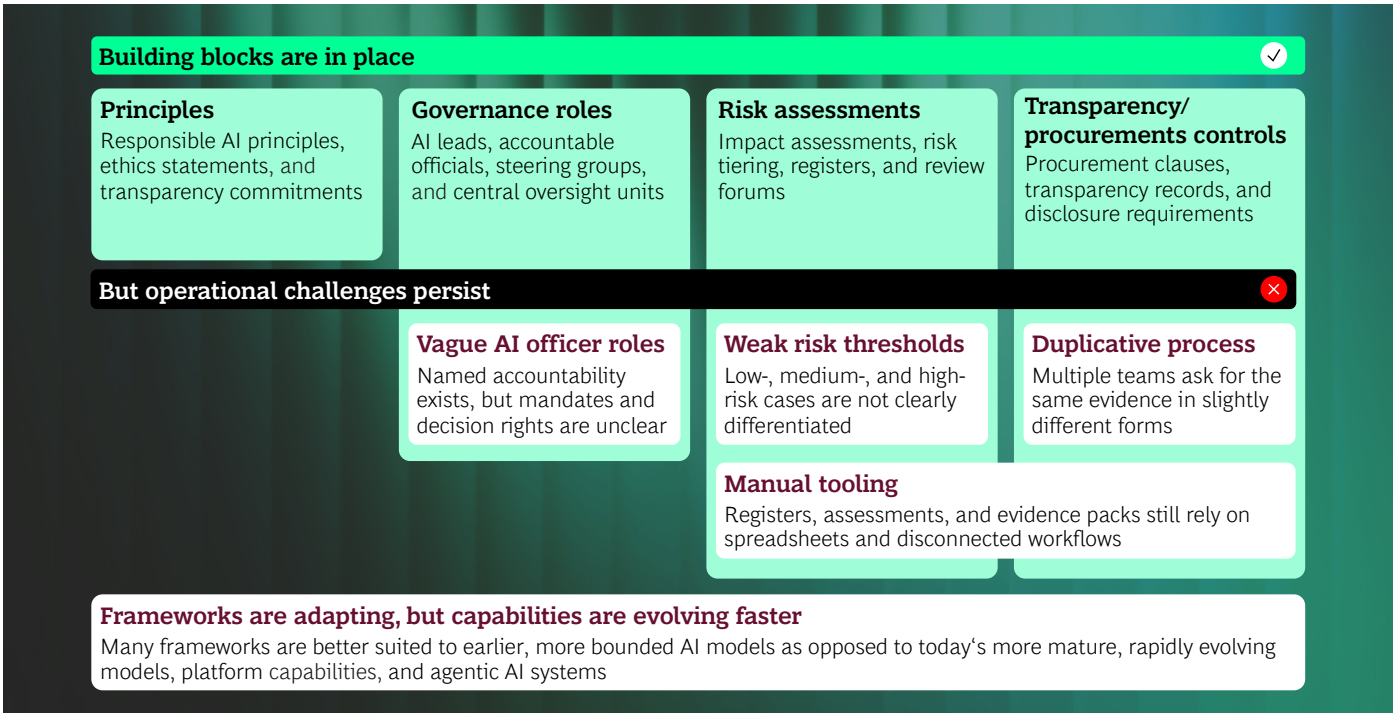
Even when the governance logic is sound, implementation may be too complex. Serial handoffs, duplicate review requests, and unclear sequencing create delays even for low-risk and high-priority use cases. (See Exhibit 2.)

One public sector team described piloting a GenAI assistant to triage enquiries and help staff draft responses. The prototype itself performed well. But as the team moved toward production, it had to navigate overlapping privacy, legal, procurement, and executive approvals, with no single integrated view of what evidence was needed, when it was needed, or who had authority to make final decisions.

One government agency identified 71 crossover points where different teams asked for the same or slightly different information related to data, project management, finance, ethics, and broader assurance.

> *There were 71 crossover points where different teams ask for the same, or slightly nuanced, information.*
>
> *-Senior government official, Australia*

Exhibit 2 - The Building Blocks of AI Assurance Are in Place, But Implementation Is Still a Challenge



Source: BCG analysis.

According to our interviewees, senior leaders generally agree that organizations should use AI responsibly, but the challenge emerges when they must decide whether a specific deployment meets that standard. At that point, the practical questions become much sharper: What is mandatory? What is recommended? Who owns the decision? What evidence is sufficient? And how much uncertainty is acceptable? When the established guidelines provide clear answers to those questions, leaders tend to have greater confidence in their teams and processes. When they do not, organizations often default to demanding additional reviews, more documentation, and further escalation.

Several interviewees raised shadow AI as an emerging concern when formal pathways are too difficult or slow to navigate. But the larger theme emerging from our interviews was the number of worthwhile use cases that never move forward at all. In these instances, organizations abandon potentially valuable use cases simply because the governance process seems too hard to navigate. For example, in Ontario, interviewees described very basic, low-risk use cases being governed so heavily that experimentation itself became difficult. In another government, teams had a hard time moving beyond proof of concept and into support, adoption, and business ownership.

> *We're really, really good at doing proofs of concept. We're not great after that.*
>
> *-Government practitioner, Australia*

> *We are exquisitely governing very basic, low-risk AI use cases to within an inch of their lives.*
>
> *-Senior government official, Germany*

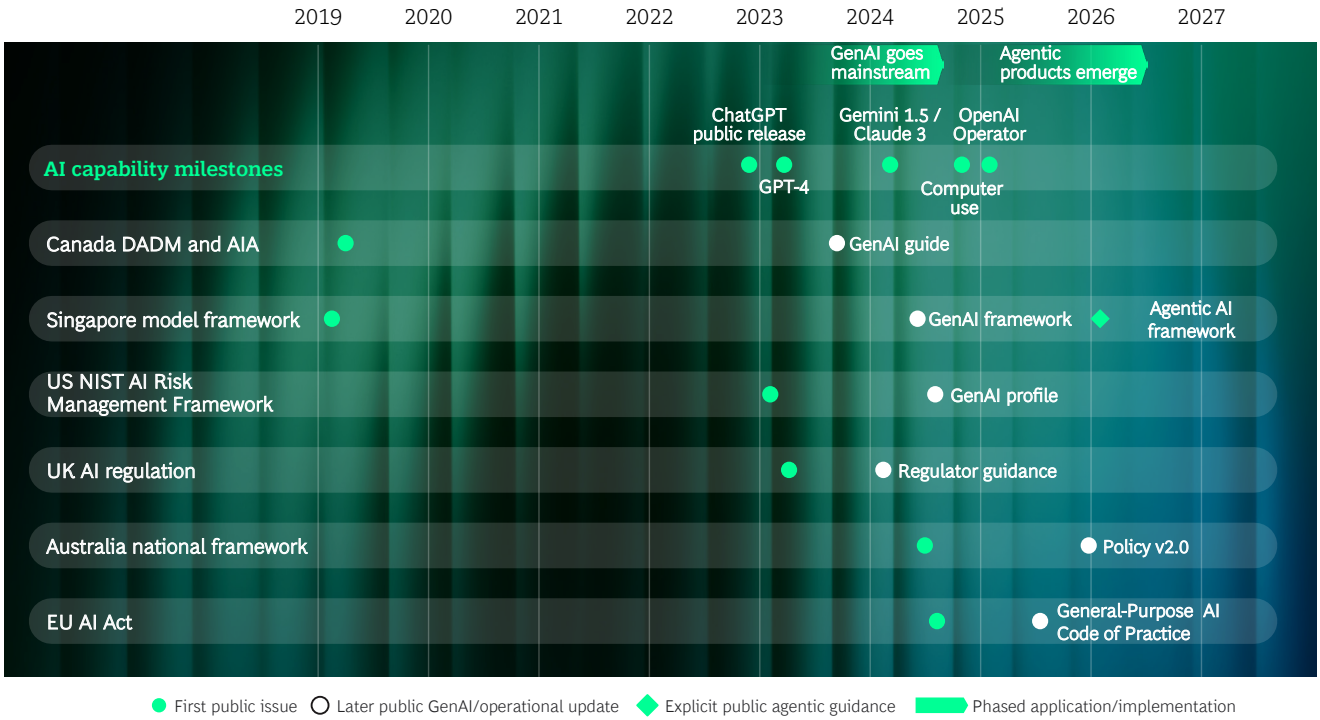
What Is Not Working: The Frameworks Are Evolving, but Model Capabilities Are Evolving Faster

Many current assurance frameworks were originally developed for earlier generations of AI. While several countries have already updated their frameworks or guidance in response to GenAI, model capabilities continue to evolve at a pace that is difficult for frameworks, guidance, and operational processes to match. (See Exhibit 3.) Even where frameworks have been updated, they are still struggling to cope with embedded AI within enterprise tools and platforms, and agentic AI that introduces additional questions around delegated decision-making and actions.[27]

Agentic AI is particularly challenging because agentic systems can chain tasks together, operate semi-autonomously within limits, and alter the distribution of work between human and machine. Those capabilities raise new questions: What authority does the system have to exercise its own judgment? What decisions remain with humans? What rules apply to delegated work? What escalation paths exist if something goes wrong? As a result, governing agentic AI starts to look more like a combination of IT governance and workforce or HR governance.

> *It's almost like you're building a governance framework for humans, but they're not humans.*
>
> *-Former federal government executive, Canada*

Exhibit 3 - AI Capability Is Moving from GenAI to Agents Faster Than Public Governance Frameworks Are Adapting



Source: BCG analysis.
Note: Only selected AI assurance frameworks shown. AIA = algorithmic impact assessment; DADM = Directive on Automated Decision-Making; NIST = National Institute of Standards and Technology.

Collaborating with vendors can create friction, too. In practice, responsibility is often split across model providers, platform providers, implementation partners, and the agency using the system. But participants frequently direct assurance questions to the wrong party or ask them in bespoke ways each time, leading to repeated due diligence, unclear ownership, and requests for information that no single participant can fully answer.

Risk Triage

What Works: Simpler Triage and Tiered Risk Thinking

Risk tiers are most useful when they are specific enough to be applied consistently. Clarity on this point gives users a way to distinguish genuinely high-risk systems from lower-risk use cases.

In New South Wales, the government shifted from a specialist-heavy process to a short inherent-risk triage supported by structured questions. The previous process could take around 40 hours to complete and required advanced expertise, whereas the redesigned approach could provide an initial directional assessment in less than 30 minutes, with deeper tooling available as needed. The new process also made reassessment more feasible, because the initial classification became faster and easier to revisit as the use case evolved during development or after deployment.

What Is Not Working: Unclear Guidelines for Risk Triage

Most frameworks recognize that some AI use cases should be treated differently than others. In practice, however, thresholds are often too vague for confident use, and reviewers who are uncertain about what AI can and cannot do tend to classify GenAI use cases as high risk by default.

One interviewee described a use case that had been classified as medium risk under a national threshold process but was treated as high risk under the agency's internal framework, pushing it into a much heavier assurance pathway. At another agency, whose customer cohort and policy environment were especially sensitive, almost any AI use case could be interpreted as inherently high risk in the absence of a triage model specific enough to distinguish meaningfully between different types of use.

As a result, lower-risk productivity use cases end up traveling the same route as far more consequential systems. For this reason, interviewees repeatedly advocated for clearer exemplars of what use cases count as low, medium, or high risk.

Triage does not need to be complex to be effective. It needs to be specific enough that teams can tell what can proceed locally, what needs review, and what genuinely requires deeper scrutiny.

> *A risk classification system with exemplars would be extremely helpful.*
>
> *-Senior government official, Australia*

Named Accountability

What Works: Named Accountability Creates a Starting Point

Increasingly, countries and agencies are appointing AI leads, accountable officials, or chief AI officers because AI governance needs clear ownership. Without being assigned to a named role, ownership of AI governance can become diffuse, leaving decisions to committees and ad hoc escalation.

In Japan, a ministry-level chief AI officer gives ministries a visible focal point for AI governance. Likewise, New South Wales and the Australian federal system are moving toward more explicit accountability roles, including a Chief AI Officer and AI accountable official. Such roles give agencies a clear place to anchor responsibility rather than leaving AI to sit ambiguously among IT, legal, risk, and delivery functions.

What Is Not Working: The Roles Exist but the Job Description Does Not

Named accountability is useful in principle, but in practice the role definitions may be too loosely defined. AI officers or other official positions may exist without the mandate, decision rights, or capabilities they need to guide decisions effectively. In those cases, the role can become just an additional reporting layer rather than a practical source of direction and accountability.

For example, agencies may appoint a very senior AI official without clearly defining the role, its decision boundaries, or what effective judgment should look like in practice. The result can be accountability on paper with inadequate operational capability in reality. For this reason, even though the appointment of AI officers is potentially a strong building block in AI governance, it is among the areas most in need of further refinement.

Transparency Tools, Practical Artifacts, and Supplier Evidence

What Works: Transparency Tools and Practical Artifacts Can Improve Consistency

Transparency artifacts can make AI governance more concrete. The UK's Algorithmic Transparency Recording Standard, for example, gives departments a practical structure for documenting purpose, data, and oversight, permitting visibility of many public uses in one place.[28] Similarly, Canada's broader AI registry surfaces uses that fall outside the narrower decision system framework, thereby revealing where departments are duplicating effort.[29] Singapore's A.I. Verify pairs a principle statement with testing tools, templates, and a repeatable method for checking claims.[30]

These tools turn principles into things that teams can actually work with: templates, testing methods, self-assessment approaches, transparency records, and consistently reviewable evidence. Such measures can increase trust in AI. Across the countries surveyed, detailed technical documentation, independent testing before launch, and regulator certification are among the most reliable and effective ways for governments to demonstrate that AI is safe and fair.

In these countries, the broader AI registry improves visibility across use cases and highlights opportunities to consolidate common tools, reducing the need for each department to build its own version. Visibility can help create a clearer picture of what is happening across government.

What Is Not Working: Unclear Ownership Across the Assurance Stack

Our interviews uncovered an ongoing challenge in the ways that governments collaborate with vendors and private sector partners. Assurance approaches originally designed for direct procurement of software or standalone models are now being applied to solutions that are platform-based, cloud-delivered, continuously updated, and governed through shared responsibility across model providers, platform providers, implementation partners, and government agencies.

In one instance, a platform provider was asked to supply model training methodology and evidence of sector-specific legal compliance for a third-party foundation model that it had not trained. In another, evaluators expected to receive deterministic outputs from a GenAI assistant. In a third, a customer sought guarantees that platform capabilities would remain unchanged for the duration of the contract, despite the fact that continuous releases were fundamental to the service model. These are not unreasonable assurance concerns, but they become sources of friction when directed to the wrong party or when the logic of traditional automation is applied to a probabilistic system.

A similar mismatch involves the way some governments assess cloud-based AI services and decide what evidence is required. Some review processes still reflect an on-premises audit model, requiring component-level evidence or bespoke mappings into internal governance tools, even when standard certifications, architecture summaries, evaluation results, change logs, and provider evidence packs would be more practical and informative. In some cases, providers are also asked to make deployment-level ethical or regulatory judgments that, in practice, should be the responsibility of the government agency using the system.

Practical artifacts are helpful only when participants direct their evidence expectations to the right party and align them with the way systems actually deliver modern AI products. The problem is not vendor involvement itself, but clarity over who is responsible for what evidence, what decisions, and what ongoing controls across the stack.

Technical Capability and Confidence

What Works: AI Safety Institutes and Technical Assurance Centers

AI safety institutes and adjacent technical assurance centers do not replace agency-level governance, but they can strengthen capabilities that many departments might otherwise struggle to build independently, such as rigorous technical testing, shared evaluation methods, and more consistent approaches to advanced-model risk.

The UK's AI Security Institute is explicitly focused on giving government a scientific understanding of advanced AI risks and developing and testing mitigations.[31] In the US, NIST's Center for AI Standards and Innovation focuses on testing, standards, and collaborative research, and has launched an AI Agent Standards Initiative aimed at governance of secure and interoperable agents.[32] Japan's AI Safety Institute has similarly positioned itself as a hub for AI safety methods, guides, and coordination across government, industry, and academia.[33] These countries are investing not only in policy and principles, but also in technical assurance capability.

What Is Not Working: Capability, Literacy, and Tooling at the Point of Use

Even when the rules are clear, the people applying them may not share the same understanding of AI risk. Interviewees described privacy teams, cyber teams, procurement teams, and delivery leads reaching different conclusions about the same use case. Having a framework is not enough if the people applying it don't share the same mental model of risk or have the tools they need to move evidence through the life cycle.

Weak tools exacerbate this problem. In one country, early registry began as an Excel list, with other assurance tools distributed as files, limiting visibility into which reviewers had completed their assessments and what use cases were emerging. When a program design splits registers, self-assessments, and evidence packs across spreadsheets, emails, and separate governance systems, evidence cannot move with the use case, forcing teams to recreate it.

Note: This photo was generated with AI

04.

How We Might Make It Better

Through our research, we have identified nine ways to make existing assurance models practical enough to capture AI opportunities while managing risk appropriately.

Make Risk Triage Easier

A more differentiated approach and well-defined pathways are needed to enable lower-risk applications to proceed more quickly and easily and ensure that deeper scrutiny is still conducted on higher-risk use cases. A practical triage model where low-, medium-, and high-risk categories are supported by practical examples and thresholds is easier to apply in practice. This typically involves providing specific examples that teams can recognize immediately and classify intuitively. For example, low risk could be an internal drafting assistant for policy teams, with no direct decision authority, where a human reviews outputs before use. Medium risk might be an internal case triage support tool that recommends next actions, but where a human makes the final decision and executes the action. High risk could be an external public-facing agent that provides answers to questions or inquiries.

In New South Wales, Australia's largest state government, the AI risk assessment process was redesigned around a simpler inherent-risk triage designed to give teams a faster initial view before launching a deeper review if needed. In Japan, authorities replaced blanket preapproval with a report-and-review model based on impact and sensitivity. In this system, lower-risk use cases (such as internal productivity tools) can proceed with internal documentation and notification; medium-risk use cases (such as internal decision-support tools) may require additional review within the ministry or consultation with central guidance; higher-risk use cases (particularly public-facing ones) are escalated for central review or advisory input.

The challenge is calibrating assurance to be stringent where stakes are high, and light where they are not.

Move from One-Time Approval to a Staged and Lifecycle Approach

Current assurance models often assume that risk assessments and decisions are linear and made upfront just once. This tends to be unrealistic in practice because not all uncertainty can be resolved upfront and AI systems by their nature evolve over time. Practitioners we interviewed indicated that forcing false certainty too early leads to delays, tick-and-flick, perfunctory answers, or abandonment of use cases. It is unrealistic to expect every uncertainty to be resolved before a system is even trialed, where risk often becomes clearer only after controlled experimentation and usage. For example, Japan's emphasis on trial environments and proof-of-concept verification points offers a positive example of how assurance can be managed, as does the broader shift toward reassessment when material changes occur.

AI assurance increasingly needs to function as an ongoing capability. Prompts and guardrails must be managed, updates governed, drift monitored, evidence kept current as systems change, and human supervision maintained.

Potential Actions

Moving toward staged and lifecycle-based assurance allows early-stage systems to proceed under defined controls, then reassess when the user base expands, new data sources are added, models or prompts change materially, or monitoring reveals new issues.

Staged assurance also helps organizations build competence and confidence over time. Early controlled deployment through pilots and proofs of concept allows teams to test not only the technology itself, but also the clarity of roles, escalation paths, and monitoring arrangements, and the controls' practical usability before the system scales.

What This Could Look Like in Practice

The approach has four stages:

Stage 1. Sandbox/pilot or proof of concept with limited users and tightly controlled data

Stage 2. Controlled deployment with specifically defined use cases and monitoring in place

Stage 3. Scale with broader rollout and stronger oversight

Stage 4. Ongoing monitoring with clear triggers for re-assessment

Streamline and Integrate Review and Approval Workflows

The experience of many delivery teams seeking approvals often involves navigating a labyrinth of ethics, risk, legal, cyber, privacy, procurement, and other requirements which are overlapping, duplicative, and sequential. In many cases, AI assurance and risk has been overlayed on top of existing policies, processes, committees, and governance structures. The end-to-end process is often not clear or well-defined, with multiple hand-offs and sign-offs. Each approval requires different assessments, paperwork, and evidence to be presented.

Potential Actions

Best practice is to have a clear pathway or workflow which begins at the earliest stage of discovery and treats assurance artifacts as living documents that move with the use case through delivery so the same evidence can support architecture, privacy, cyber, procurement, and governance rather than being recreated at each stage.

This assurance workflow can be supported with a digital workflow management tool and assurance register so that the status of any use case can be tracked and reported on, relevant documents are readily available at every stage, and there is an established audit trail for decisions and sign-offs.

What This Could Look Like in Practice

- Risk classification done early during the discovery phase
- Evidence created and added as part of each stage of design, build, and testing
- Reusable artifacts such as architecture summaries, data handling notes, and evaluation outputs
- Same artifacts can be used across risk, legal, privacy, data protection, procurement, and cyber

Reduce Duplication with Reusable Artifacts for Common Patterns

Reusable artifacts are standard documents and templates that can be used repeatedly across similar use cases instead of being recreated for each review. The UK's Algorithmic Transparency Recording Standard provides a repeatable transparency record that helps teams document purpose, data, oversight, and explainability in a consistent way rather than reinventing the same material for each use case.[34] Canada's registry approach improves visibility across uses, making it easier to identify duplication, compare similar deployments, and consolidate common tools rather than having every department build its own version.[35] Singapore's A.I. Verify shows how testing and self-assessment can be made more concrete by linking broad governance principles to repeatable checks and technical validation.[36]

The private sector practice offers reusable artifacts such as case cards, architecture summaries, change logs, and standard evidence sets. This matters not only within agencies but also in collaboration with vendors and implementation partners, because it allows the same evidence to move across procurement, design review, release approval, and ongoing monitoring. It can be useful to consolidate mandatory-versus-advisory guidance in one authoritative place and to standardize the artifacts most teams repeatedly need, such as architecture summaries, data-handling notes, evaluation results, change logs, and supplier documentation. This can reduce the amount of work spent recreating the same evidence for multiple use cases or audiences.

Potential Actions

Reusable evidence packs for common AI patterns, supported by short AI use cards, can reduce duplication substantially. Instead of every team writing a fresh assurance narrative for an "internal summarization assistant" or "meeting note generator," agencies could start from a standard pack and adapt only the elements specific to the use case. The same principle could apply to suppliers and implementation partners who could provide standard artifacts that agencies tailor to their own context rather than treating each review as a bespoke exercise.

For example, AWS AI Service Cards offer a single place to find intended use cases and limitations, responsible AI design choices, and performance optimization best practices for AI service and models.[37] Agencies still need

to assess performance within their own context, but vendor-supplied artifacts like these can make assurance faster and clearer when they align with agency requirements. Reviewers also benefit from seeing a familiar structure and evidence format each time.

What This Could Look Like in Practice

- Each evidence could contain:
  - Intended use
  - Standard risks
  - Minimum controls
  - Data handling notes
  - Architecture summary
  - Human oversight model
  - Monitoring approach
  - Vendor, model, and implementation partner information
  - Known limitations
- Each AI use card should say:
  - What this pattern is for
  - What it is not for
  - Where it is appropriate
  - What minimum evidence is needed
  - What the standard escalation triggers are

Provide Clearer Accountabilities and Roles

Naming specific roles, such as an AI lead, accountable official, or Chief AI Officer is necessary but not sufficient. Several interviewees described situations involving the creation of accountability roles without clear job descriptions, decision rights, or defined ways of working with legal, cyber, privacy, procurement, and delivery teams.

When the roles are too vague, agencies can end up appointing someone who is senior enough to sign off but too far removed from delivery to guide the decisions effectively. In some cases, central roles such as Japan's ministry-level Chief AI Officer model provide a useful focal point for adoption and oversight.[38] If decision making becomes too centralized, however, it can become disconnected from day-to-day delivery, reinforcing the need to pair senior accountability with clear ownership at the use-case level.

Named accountability is most effective when tied to clear decision rights, capability expectations, and escalation paths. It is important to clearly define what these roles are responsible for, what authority they hold, what capabilities they require, and how they operate with business and control functions in practice. This means defining who sponsors adoption, who owns the use case, who can approve release, who accepts residual risk, and who is responsible for monitoring after go-live. The goal is to ensure that accountability is distributed clearly enough for practical decisions to be made and reviewed.

Importantly, a named AI accountability role is best understood not as a single point of responsibility for all AI risk across the organization, but as a governance function: one responsible for ensuring that the right accountability sits with the right party at each point in the AI lifecycle, and that shared responsibility across model providers, platforms, implementation partners, and the deploying agency is clearly allocated and actively managed. The absence of a clear job description compounds this challenge: without it, there is no basis for defining what that allocation should look like in practice, or for ensuring each party fulfills their responsibilities accordingly.

A more precisely defined operating model could distinguish between a small number of defined roles:

- **AI Champion/Chief AI Officer.** Provides enterprise-level sponsorship, sets direction, and resolves escalations where needed
- **Business or Service Owner**. Owns the use case, the expected value, and the decision to proceed within agreed risk boundaries
- **Accountable Official.** Accepts residual risk for higher-impact cases and confirms that the required evidence is in place before release
- **Technology or Product Owner.** Implements controls, maintains the system, and manages changes, monitoring, and incident response after go-live
- **Control Functions (Legal, Cybersecurity, Privacy, Procurement, Risk).** Review and challenge the use case within their domains, but do not replace the accountable owner

This creates a clearer decision pathway in practice. For example, a lower-risk internal productivity tool might be approved by the business owner within a defined pathway once the required evidence is complete, while a higher-impact public-facing system might require escalation to an accountable official or governance forum. This arrangement enables clearer accountability, fewer unnecessary escalations, and more consistent decisions in practice.

Create More Differentiated Approaches to Assurance

Many of the current AI assurance frameworks are designed to handle specific, singular, pre-defined point solutions; however, they are not well-suited for assessing the risks associated with general purpose applications or platforms. In some cases, government

Note: This photo was generated with AI

agencies have struggled to assess embedded AI tools such as copilots and assistants because the frameworks assume specific, narrowly defined uses, whereas enterprise platforms are general purpose tools and might support a broad and undefined range of uses and applications. Similarly, as part of the AI risk assessment process, platform providers are sometimes asked for model training about foundational models that they did not build. And evaluators have been seeking a level of accuracy or deterministic output from probabilistic systems.

Assurance frameworks need to adapt and reflect the layered responsibilities involved in developing and deploying AI systems. Frameworks can distinguish between first-party model development, third-party foundation models, and platform orchestration, while clarifying the roles of vendors and implementation partners across the assurance chain. In practice, this also means directing requests to the right party: model providers for model-level information, platform providers for configuration and change management, implementation partners for integration design, and the deploying agency for use-case accountability.

What is new, however, is that existing AI assurance frameworks have largely focused on the capabilities and safety properties of large language models in isolation. While model-level assurance remains necessary, it is insufficient for the complex, multiparty ecosystems through which AI is deployed in practice. In agency deployments, the orchestration layer (the system prompt, the exposed context, the permitted actions, and the triggered tasks built around the model) is often more determinative of risk than the model itself. A well-aligned model operating within a poorly governed orchestration layer can still produce harmful, non-compliant, or out-of-scope outcomes. Current frameworks do not

adequately assign ownership of this layer to the parties who design and operate it.

A more robust approach distributes accountability explicitly across the delivery stack. (See Exhibit 4.) Frameworks can distinguish between first-party model development, third-party foundation models, and platform orchestration, while clarifying the roles of vendors and implementation partners across the assurance chain. In practice, this means directing requests to the right party: model providers for model-level information and capability disclosures; orchestration platforms for configurable guardrails, audit logging, and access controls; implementation partners for integration design and bounded deployment; and the deploying agency for use-case approval, delegated authority, and operational risk governance.

Agencies that can confidently take accountability for the use case are typically better positioned to work effectively with vendors because the boundaries of shared responsibility become clearer for all parties involved. Government agencies should remain accountable for use-case intent, delegated authority, human review gates, and operational risk, while suppliers and implementation partners provide model documentation, change notices, platform controls, integration evidence, and known limitations. Formulating more technology-informed questions can help avoid asking one party for evidence that only another part of the stack can realistically provide.

Vendors and implementation partners can provide the relevant product documentation and technical evidence for their part of the stack.

Update Assurance Frameworks for Agentic AI

Prompted or context-driven AI presents a different oversight challenge from goal-directed agentic or multi-agent systems operating across workflows. Unlike prompted AI, an agentic system can pursue objectives across multiple steps: selecting tools, retrieving data, calling external systems, and triggering actions without a human reviewing each one. Many governments have not yet updated their AI assurance and governance frameworks for the developments in agentic AI. Singapore is an exception; its model governance framework was followed by A.I. Verify, then by separate generative AI

Exhibit 4 - Agentic AI Requires Shared Responsibility and Structured Supervision

Stakeholder	Primary Accountability	Key Controls
Government Agency	AI scope, role definition, and risk governance	Use case approval, bounded permissions, oversight
LLM Provider	Model safety, alignment, and capability transparency	Model cards, capability disclosures, safety evaluations
AI Orchestration Platform	Platform-level guardrails and enforcement mechanisms	Configurable guardrails, audit logging, access controls
Implementation Partner	Faithful and controlled deployment of the defined use case	Bounded context, permitted actions, restricted task scope
Adjacent Applications	Secure and compliant integration with the AI system	API security, data handling, integration governance

Source: BCG and Salesforce analysis.

and agentic AI guidance as the technology evolved.[39] This guidance sets out how responsibility for governing these systems should be shared between the business and technical teams.

For agentic AI, the business should define what the agent can do and what remains with a human to do. The technical teams then enforce those boundaries through engineering guardrails. For example, consider an internal agent that drafts responses, retrieves applicable policies, and recommends next actions for staff. In a weak governance and assurance model, the business defines the desired outcome, security performs a checkpoint review, and the developer is left to make the agent "safe" through prompts and technical controls; however, this places too much responsibility on the implementation team.

A better approach would be for the business owner to define upfront what the agent may and may not do; what authority remains with the human; which conduct, delegation, HR, and legal rules apply; and when escalation is required. Technology then implements these constraints in the system through permissions, tool access, prompt restrictions, monitoring, and audit logs.

Where a third-party model or platform is involved, vendor documentation and change notices form part of the evidence base, but do not replace the deploying organization's accountability for the use case.

Once deployed into production, the agent is best treated as a form of delegated digital worker that requires bounded authority, continuous named human supervision, clear escalation rules, continuous monitoring, and the ability to intervene or revoke access when needed.

Even with an enhanced framework in place, some agentic use cases may still be inappropriate for government deployment, particularly where consequential decisions affect rights, eligibility, or access to services.

What This Could Look Like in Practice

Business owns

- Purpose of the use case
- What the agent may and may not do
- What decisions remain with a human
- Delegated authority
- Human review gates
- Legal, policy, and HR rules
- Escalation thresholds

Technology owns

- System configuration
- Permissions
- Tool access
- Prompt and guardrail design
- Monitoring
- Testing
- Rollback

Vendors/ platform providers support

- Model-level documentation and known limitations
- Release notes, change logs, and control information

Security/compliance/HR/legal/risk support

- Whether the delegated work is appropriate
- Whether conduct, workforce, and legal constraints are reflected
- Whether security, regulatory, and policy constraints are reflected
- Whether escalation and supervision rules are adequate
- Sign-off for prompts used in customer-facing or regulated contexts
- Definitions of what data can and cannot be used in prompts

Combine Broad AI Capability Uplift with Strong Central Technical Assurance

Delivery teams working on AI initiatives and AI risk reviewers often have different levels of competence and understanding of AI risk. Staff sometimes lack confidence using approved tools and require more training to build this confidence. In Germany, for example, the rollout of a new chatbot required several live webinars with staff to obtain a "GenAI license" before they felt confident and comfortable using the tool.

At the same time, some countries are strengthening their central technical capabilities through AI safety institutes and adjacent centers. The goal of these centers is to build practical AI literacy across risk, procurement, privacy, cyber, and leadership roles, while also providing shared technical assurance capability for testing, evaluation methods, and common safety approaches where agencies cannot build this on their own.

Together, these measures build the staff competence and confidence needed to utilize the defined decision pathways, rather than escalating and amplifying uncertainty through the system.

Potential Actions

Establish a dual capability model that combines

- Building out a minimum level of AI literacy and fluency across all staff and teams, so that lower- and medium-risk decisions can be made confidently
- Creating the shared technical assurance capability and expertise to support escalation and more complex evaluations

What This Could Look Like in Practice

- Mandatory AI literacy training and accreditations or certifications
- Simple and explicit guidance for
  - What AI can be used for (green light) and cannot be used for (red light or no-fly zones)
  - What "safe use" means in practice
- Center of Excellence or equivalent shared unit providing
  - Reusable assurance artifacts and evaluation templates
  - Guidance on emerging model capabilities and risk patterns
  - A single point of escalation for novel or ambiguous cases
- Central or shared teams providing model evaluation support as well as security and risk testing
- AI safety institutes or central units acting as extended technical and reusable capability hubs

Measure Both Value and Risk

Most AI assurance systems measure activity, such as forms completed, reviews undertaken, and artifacts produced. A better approach would assess both how well an organization is realizing the value from AI as well as how effectively and efficiently it is managing risk. In addition to tracking incidents, escalations, and compliance, it also involves tracking the time to decision and deployment, and subsequent AI adoption and usage.

Potential Actions

One practical approach is to track and report on the efficiency and effectiveness of risk management activities and opportunities captured and benefits realized.

What This Could Look Like in Practice

Management of risks

- Rate of first-pass approvals
- Incident or adverse event trends
- Number of escalations required
- Compliance quality monitoring
- AI monitoring drift alerts

Opportunity unlocked

- Time to decision
- Time to deploy and scale
- Number of low-risk and medium-risk use cases approved
- Number of AI systems deployed in production and in usage at-scale
- Number of active AI users
- Estimated hours saved, cost savings, or value achieved

05.

Responsible AI Adoption Builds Trust in Government

Many governments already have the high-level principles and frameworks for AI governance and assurance in place, but the operational layer beneath them is still a work-in-progress. Roles can be too vague, risk thresholds are not always well-defined or applied consistently, processes are often manual, and many frameworks are outdated given rapid developments in AI models and capabilities.

The focus now is to make AI assurance clearer, more practical, and better aligned to how contemporary AI is built and deployed. This means improving risk triage, strengthening lifecycle monitoring, clarifying decision pathways, increasing reuse of evidence, and defining shared responsibilities more clearly across agencies, vendors, and implementation partners. Ultimately, the goal is to maximize the opportunity to create public value while staying appropriately disciplined to manage risk and scale AI responsibly.

BCG's biannual survey of citizens around the world across 40+ countries examines attitudes towards the use of digital and AI in government, and the latest results for 2026 show that citizens' usage of AI and overall literacy are increasing. Across the surveyed countries, on average, self-reported AI proficiency beyond a basic level rose by roughly 9% from 2024 to 2026, and respondents using AI at least weekly rose by more than 25%.

As citizens become more active users and their familiarity with the technology grows, their expectations of government adopting and using AI are also rising. People who use the technology regularly and have greater expertise are less likely to be fearful of AI and more likely to believe that the benefits outweigh the risks. As AI becomes a more visible and familiar part of daily life in workplaces, in consumer products, and in how people search, write, and make decisions, citizens will form views about whether their governments are keeping up with the latest technologies in the ways they operate and deliver services. When governments fall too far behind societal and private-sector adoption, the gap can contribute to declining confidence in public institutions, alongside other drivers of trust such as competence, fairness, and responsiveness.

Effective AI assurance helps governments build both competence and confidence to move faster and more responsibly without compromising safety. When this happens, AI becomes more than a productivity or efficiency tool: it enables the public to derive greater benefit and value from AI and becomes a force multiplier that builds trust in government.

About the Authors

Miguel Carrasco is a Managing Director and Senior Partner in the Sydney office of Boston Consulting Group and is chair for the Centre for Public Impact, Australia & New Zealand. You may contact him by email at carrasco.miguel@bcg.com.

Daniel Selikowitz is a Managing Director and Partner in the firm's Sydney office. You may contact him by email at selikowitz.daniel@bcg.com.

Cindy Sia is a Platinion IT Architect in BCG's Melbourne office. You may contact her by email at sia.cindy@bcg.com.

Gisele Kapterian leads public sector strategy across Australia and New Zealand for Salesforce. You may contact her by email at gkapterian@salesforce.com.

Justin Tauber leads the agentic technology, trust and adoption practice for Salesforce Australia & New Zealand. You may contact him by email at jtauber@salesforce.com.

Acknowledgments

The authors would like to thank Keira Lowther, Anika Baset, and Elysa Neumann from the Centre for Public Impact (CPI) for their input and support. They would also like to thank Claudine Challita, Michael Bonaddio, and Akshay Koul at Salesforce for their writing and support.

They are also grateful to Yuna Moon, Rebecca Diepenheim, Pandi Velladurai, Debbie Spears, Faisal Faraz, Hannah Coatsolonia, and Amy Strong at BCG for their contribution to the writing, design, and production.

They further wish to thank Richard Sargeant, Kirsten Rulf, Steven Mills, Yoshihisa Niwa, Masahiro Nakagawa, Aparna Bharadwaj, Mario Gonsalves, Saibal Chakraborty, Eric Sullivan, Lacy Ketzner, Emily Aptaker, Steven Mills, Brooke Bollyky, Patrick Roche, Davide Di Domenico, Edoardo Palminasi, Stefano Masino, Benjamin Desalm, Anne Kleppe, Daniel Zepter, and Lionel Corre.

This report was jointly funded and developed by BCG and Salesforce, in collaboration with the Centre for Public Impact, and includes reflections based on the authors' direct experience of government-vendor dynamics in relation to AI risk and assurance.

Note: This photo was generated with AI

Appendix

This list summarizes key digital and AI assurance instruments released in Australia, Canada, the EU, Germany, India, Italy, Japan, Singapore, the UK, and the US between 2012 and 2026.

Jurisdiction	Instrument	Type	Status	Year released
Australia (national)	AI Impact Assessment Tool	Policy and self-assessment tool	Voluntary for industry; mandatory for APS and non-corporate Commonwealth entities	2025
Australia (national)	AI Policy Update	Principles	Voluntary for industry; mandatory for APS and non-corporate Commonwealth entities	2026
Australia (national)	AI Procurement Guidance	Guidance	Voluntary for industry; mandatory for Australian Capital Territory and non-corporate Commonwealth entities	2025
Australia (national)	APS AI Plan	Principles	Voluntary for industry; mandatory for Australian Capital Territory and non-corporate Commonwealth entities	2025
Australia (national)	Australia's AI Ethics Principles	Principles	Voluntary	2019
Australia (national)	National Framework for the Assurance of AI in Government	Framework	Voluntary for industry; mandatory for Australian Capital Territory and non-corporate Commonwealth entities	2024
Australia (territory: Australian Capital Territory)	ACT AI Assurance Framework	Policy and self-assessment framework	Applies to Australian Capital Territory government	2025
Australia (territory: Northern Territory)	NT Government AI Assurance Framework	Framework and self-assessment	Applies to Northern Territory government agencies and government-owned corporations	2025
Australia (state: New South Wales)	NSW AI Ethics Policy and AI Assessment Framework (AIAF)	Policy and self-assessment tool	Mandatory for New South Wales agencies	2022
Australia (state: New South Wales)	WHS Amendment Bill	Policy	Mandatory for New South Wales agencies	2026
Australia (state: Queensland)	Queensland Government AI governance policy	Policy	Applies across Queensland government entities	2024
Australia (state: South Australia)	SA Artificial Intelligence Ethics Policy (DTF/P9.1)	Policy	Voluntary	2019
Australia (state: Tasmania)	Guidance for the use of AI in Tasmanian Government	Guidance	Voluntary	2024
Australia (state: Victoria)	Responsible AI Framework	Framework and transparency record to fill	Mandatory for Victorian agencies	2024
Australia (state: Western Australia)	WA Government AI Policy and AI Assurance Framework	Policy	Mandatory for Western Australian agencies	2025
Canada (national)	Algorithmic Impact Assessment tool (AIA) and AIA public registry	Assessment tool and transparency mechanism	Mandatory	2019
Canada (national)	Directive on Automated Decision-Making (DADM)	Policy	Mandatory	2019
Canada (national)	Enhancing Digital Security and Trust Act	Policy	Mandatory	2024
Canada (province: Ontario)	Responsible use of Artificial Intelligence Directive	Policy	Applies to Ontario government organizations	2022
Canada (province: Alberta)	Alberta Data Ethics Framework	Policy	Voluntary	2025
European Union countries (supra-national)	EU AI Act (Regulation (EU) 2024/1689)	Law	Binding (EU)	2024

Jurisdiction	Instrument	Type	Status	Year released
Germany (national)	AI opportunity market and transparency database	Transparency mechanism	Binding	2025
Germany (national)	AI Playbook	Guidance	Voluntary	2025
Germany (national)	BSI Criteria Catalogues	Guidance	Voluntary	2021
Germany (national)	Federal Ministry of the Interior's AI Guidelines for the Federal Administration	Guidance	Voluntary	2022
Germany (national)	Leitlinien für den KI-Einsatz in der Bundesverwaltung	Guidance	Voluntary	2025
Germany (national)	Magenta Book	Guidance	Voluntary	2025
Germany (state: Bavaria)	KI-Leitfaden für Behörden (Bavaria)	Guidance	Voluntary	2025
India (national)	India AI Governance Guidelines	Guidance	Voluntary	2025
India (national)	Intermediary Guidelines	Guidance	Voluntary	2021
India (national)	IT (Intermediary Guidelines) Amendment Rules	Policy	Voluntary	2026
India (national)	NITI Aayog Responsible AI #AIForAll	Guidance	Voluntary	2021
India (national)	Strengthening AI Governance through Techno-Legal Framework	Guidance	Voluntary	2026
India (national)	White paper on strengthening AI governance	Guidance	Voluntary	2026
India (state: Karnataka)	Karnataka AI Governance Approach	Governance mechanism	Mandatory	Unknown
India (state: Telangana)	Telangana AI Framework	Framework	Voluntary	2020
Italy	Adottare l'Intelligenza Artificiale nella PA	Guidance	Policy	2025
Italy	AgID	Guidance	Policy	2012
Italy	National AI Law (Law 132/2025)	Law	Binding	2025
Japan (national)	AI Guidelines for Business	Voluntary	All organizations in Japan	2024
Japan (national)	Hiroshima Process	Principles	Voluntary	2023
Japan (national)	Planned amendment to personal information protection rules to facilitate AI development	Framework	Voluntary	2026
Singapore (national)	AI Verify	Testing framework and toolkit	Mandatory	2022
Singapore (national)	Model AI Governance Framework	Framework/guidance	Voluntary	2019
Singapore (national)	Model AI Governance Framework for Agentic AI	Framework/guidance	Voluntary	2026
Singapore (national)	Model AI Governance Framework for GenAI	Framework/guidance	Voluntary	2024
United Kingdom (national)	Algorithmic Transparency Recording Standard (ATRS)	Transparency mechanism	Mandatory	2021

Jurisdiction	Instrument	Type	Status	Year released
United Kingdom (national)	Data (Use and Access) 2025	Guidance	Voluntary	2025
United Kingdom (national)	Data and AI Ethics Framework	Guidance	Voluntary	2023
United States (federal: defense)	DoD Responsible AI Strategy & Implementation Pathway	Guidance	Mandatory with DoD	2022
United States (federal)	Memorandum on Accelerating Federal Use of AI through Innovation, Governance and Public Trust	Policy	Mandatory for US agencies	2025
United States (federal)	NIST AI Risk Management Framework	Framework	Voluntary	2023
United States (federal)	OMB Memorandum M-25-22	Policy	Mandatory for US agencies	2025
United States (state: California)	Executive Order N-12-23 (GenAI in state government)	Policy	State agencies under the authority of the governor of California	2023
United States (state: California)	GenAI guidelines for Public Sector Procurement, Uses and Training	Guidance	State agencies under the authority of the governor of California	2023
United States (state: Colorado)	SB24-205 (Colorado AI Act)	Governance guidance	Developers and deployers in the state of Colorado	2024
United States (state: Massachusetts)	EOTSS Enterprise Privacy Office Guidelines for the Use and Development of GenAI	Governance guidance	Commonwealth agencies in the state	2024
United States (state: Massachusetts)	OTSS Enterprise Use and Development of Generative AI Policy (2024)	Policy	Commonwealth agencies in the state	2024
United States (state: New Jersey)	NJOIT Joint Circular 25-OIT-001 - Guidance on Responsible Use of Generative AI (2025)	Law	State of New Jersey workforce	2025
United States (state: New York)	NYS-P24-001 Acceptable Use of AI (ITS policy) (2024)	Policy	New York state entities and workforce	2024
United States (state: New York)	State Technology Law §103-E – AI Inventory (2024)	Law	New York state entities and workforce	2024
United States (state: Texas)	HB 149 (Responsible AI Governance / AI Protection subtitle in Texas codes)	Policy	Workforce in the state of Texas	2025

Source: BCG analysis.

Footnotes

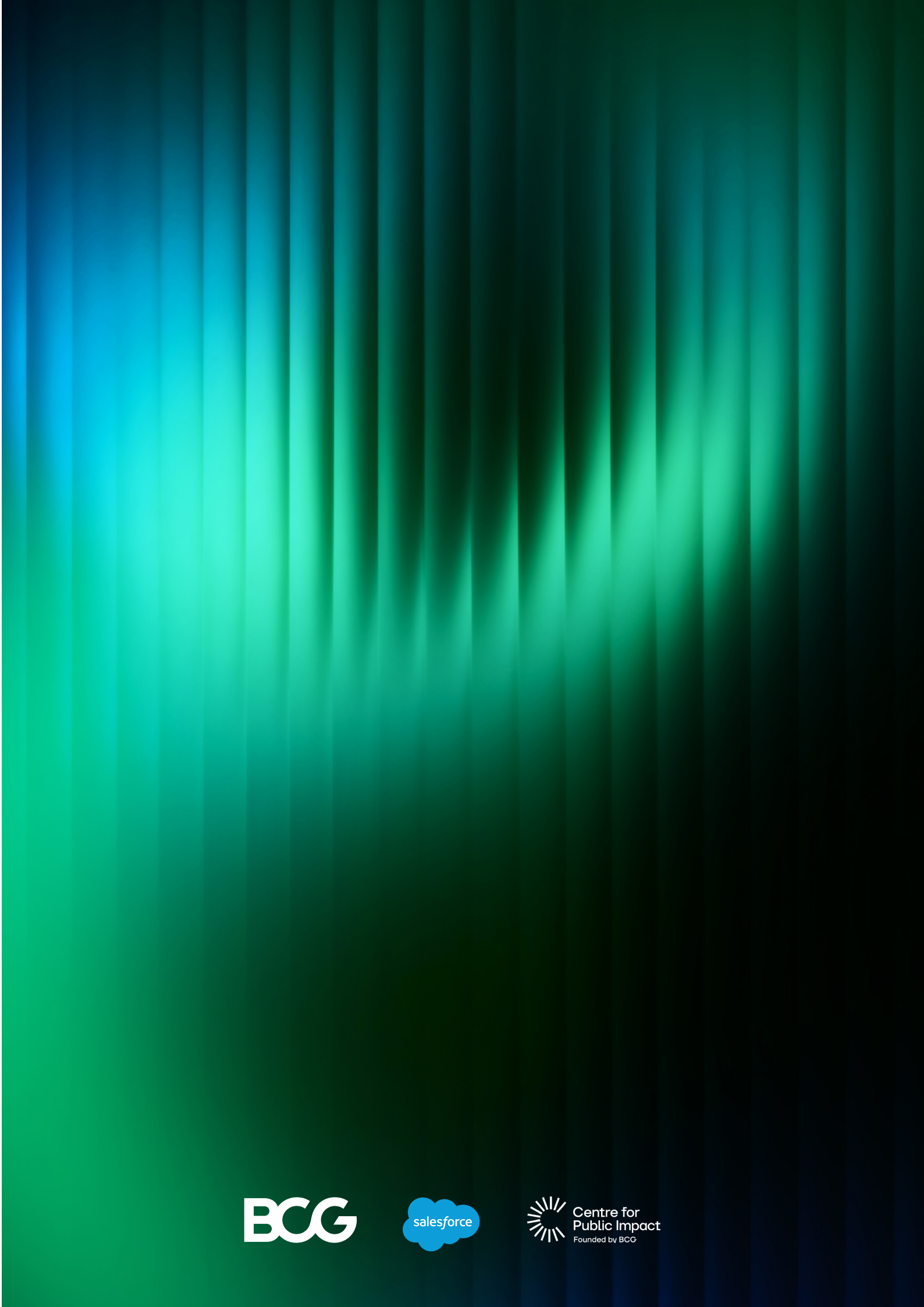
1. Oxford Handbook of AI Governance, 2022; Gregory, S. AI Governance Handbook, 2025.
2. Australian Government – AI Ethics Principles, 2019, updated 2025; National Framework for the Assurance of AI in Government.
3. Directive on Automated Decision-Making, 2019, updated 2026; Algorithmic Impact Assessment, 2019.
4. Guide on the usage of generative AI, 2025.
5. Enhancing Digital Security and Trust Act, 2024.
6. EU AI Act, 2024; Directorate-General for Competition, 2024.
7. Federal Ministry of the Interior's AI Guidelines for the Federal Administration, 2022.
8. AI Opportunity Market and transparency database, 2025.
9. BSI Criteria Catalogues, 2021.
10. AI and Data Protection Guidelines, 2025.
11. Digital Sovereignty, 2026.
12. NITI Aayog Responsible AI, 2021.
13. India AI Governance Guidelines, 2025; White Paper on strengthening AI governance, 2026; Intermediary Guidelines, 2021, updated 2026.
14. Italian Strategy for Artificial Intelligence, 2024.
15. Adottare l'Intelligenza Artificiale nella PA, 2025.
16. EU AI Act, 2024; Digital Operational Resilience Act, 2025; Italy Law 132/2025, 2025.
17. AI Guidelines for Business, 2024; Hiroshima Process Principles, 2023.
18. Singapore Model AI Governance Framework, 2019.
19. AI Verify, 2022; Model Governance Framework for Gen AI, 2024; Model Governance Framework for Agentic AI, 2026.
20. Algorithmic Transparency Records, 2021, updated 2025; Data (Use and Access) Act, 2025.
21. NIST AI Risk Management Framework, 2023; OMB Memoranda on AI Governance, 2025; Memorandum on Accelerating Federal Use of AI through Innovation, Governance, and Public Trust, 2025; US Department of Defense Responsible AI Strategy and Implementation Pathway, 2022.
22. GenAI guidelines for the State of California, 2024; Colorado Consumer Protections for AI, 2024; New Jersey Innovation Authority AI Task Force, 2023; NYS Acceptable Use of AI, 2024, updated 2025; The Texas Responsible AI Governance Act, 2025.
23. Directive on Automated Decision-Making, 2019, updated 2026; Algorithmic Impact Assessment, 2019.
24. Singapore Model AI Governance Framework, 2019; AI Verify, 2025; Model Governance Framework for Agentic AI, 2026.
25. Federal Ministry of the Interior's AI Guidelines for the Federal Administration, 2022.
26. Australian Government – AI Ethics Principles, 2019, updated 2025; National Framework for the Assurance of AI in Government, 2024; AI Impact assessment tool, 2024; AI procurement guidance, 2025.
27. Oxford Handbook of AI Governance, 2022; Gregory, S. AI Governance Handbook, 2025.
28. Algorithmic Transparency Records, 2021, updated 2025.
29. AI Registry, 2025.
30. AI Verify, 2022.
31. UK AI Security Institute, 2025.
32. NIST Center for AI Standards and Innovation, 2025; AI Agent Standards Initiative, 2026.
33. Japan AI Safety Institute, 2024.
34. Algorithmic Transparency Records, 2021, updated 2025.
35. Canada Transparency Registry, 2019.
36. AI Verify, 2022.
37. AWS AI Service Cards, 2022.
38. Japan CAIO Model, 2025.
39. AI Verify, 2022; Model Governance Framework for Gen AI, 2024; Model Governance Framework for Agentic AI, 2026.

Disclaimer

The purpose of this report is to provide general and preliminary information, and its contents should not be relied upon or construed as advice or similar. The contents of this report are disclosed in good faith, and subject to change without notice. The report contains BCG and Salesforce trademarks, confidential and proprietary information, and BCG and Salesforce retain all right, title and interest to its contents. The report does not contain a complete analysis of every material fact on the subject matter, and all warranties, representations and guarantees pertaining to the reliability, timelines, suitability, accuracy or completeness of its contents are expressly disclaimed. BCG and Salesforce, and their subsidiaries and affiliates, disclaim all liability relating to or arising from access, use or reliance on this report, including but not limited to direct, indirect, incidental, special or consequential losses arising from the information in this report, howsoever arising, including third party claims.

Artificial Intelligence Disclosure Statement

This report includes images and content generated with the assistance of artificial intelligence (AI) technology and tools. However, the final outputs, concepts and insights are based on human expertise, judgment, input, interpretation and decision-making. For further inquiries regarding the AI technologies employed in this report or for additional details on the methodology, please contact us.


BCG
salesforce
Centre for
Public Impact
Founded by BCG