SEPTEMBER 2026
ELIZABETH SEGER
DARCY WARD

AI-Assurance Ecosystems: Building the Infrastructure to Enable Confident Adoption of Safe AI

TONY BLAIR
INSTITUTE FOR
GLOBAL CHANGE

Contents

64 Contributors and Acknowledgements

Executive Summary

Artificial intelligence is becoming a foundational technology for economic growth, public-service transformation and national competitiveness. Governments and businesses are seeking to deploy AI across increasingly consequential domains – from health care and finance to critical infrastructure, national security and government operations. But intensifying concerns about AI safety, liability and governance are already slowing business investment, procurement and deployment.

Governments and businesses that successfully and safely integrate and scale AI across their operations and economies will strengthen productivity, competitiveness and long-term economic performance in a rapidly AI-driven world. Those that do not keep pace risk falling behind as AI adoption accelerates elsewhere, with slower productivity growth, weaker competitiveness and diminished long-term prosperity.

This creates a defining policy challenge for governments. They must enable the widespread adoption of AI while ensuring that evolving risks are effectively managed and public confidence is maintained. This is not a choice between innovation and governance; it is about building the governance infrastructure that allows both to thrive.

AI-assurance ecosystems are a critical part of the path forward.

AI assurance is the process of testing, evaluating and monitoring AI systems and communicating the resulting evidence to provide justified confidence in their trustworthiness.

An AI-assurance ecosystem embeds this process within the broader infrastructure needed for it to function effectively, bringing together standards, accreditation, incentive structures and supporting institutions into a coherent system that improves AI quality, demonstrates trustworthiness and embeds accountability. Rather than relying on any single governance

mechanism, it combines multiple, mutually reinforcing layers of oversight with standards and assurance practices that can evolve alongside the technology.

Getting this right will help establish the governance infrastructure to support innovation while building the trust and confidence required for widespread AI adoption. By creating clear expectations, generating credible evidence about the safety and reliability of AI systems, and embedding accountability throughout the AI lifecycle, well-functioning assurance ecosystems reduce uncertainty for developers, deployers, investors, regulators and citizens. This, in turn, accelerates AI adoption, strengthens public trust, reduces compliance burdens and creates the confidence needed for governments and businesses to deploy AI at scale.

AI assurance also presents a significant economic opportunity. Beyond improving AI itself, assurance is becoming a major industry in its own right – spanning testing, evaluation, auditing, certification, accreditation and specialist advisory services. Countries that build strong assurance ecosystems early will be well placed to deploy AI more effectively, shaping international standards and capturing a growing global market for AI-assurance services.

Governments' role is to create the conditions that allow AI-assurance ecosystems to emerge and mature. They can do this by setting clear expectations, creating demand for assurance, investing in enabling infrastructure where markets alone will fall short and coordinating the institutions for assurance ecosystems to function effectively.

Recommendations

Political leaders should focus on three priorities.

1. MAKE AI ASSURANCE A STRATEGIC NATIONAL PRIORITY

AI assurance should be recognised as a strategic national priority because it provides the governance infrastructure needed to deploy AI safely, accelerate adoption, strengthen public trust and capture the economic opportunities of an AI-enabled economy. In this way it is also central to AI sovereignty – enabling countries to make informed choices about which technologies to procure, how they are deployed and what risks are acceptable.

Governments should align the policy levers already at their disposal to translate today's broad demand for trustworthy AI into consistent demand for high-quality AI assurance, creating the sustained demand and clear expectations needed for assurance markets to emerge and giving organisations the confidence to develop, procure and deploy AI more widely.

- **Use regulation** to establish risk-proportionate AI-assurance obligations for frontier models and AI deployed in high-risk contexts, while allowing recognised standards, expert bodies and independent assurance providers to determine how those obligations are met.
- **Use public procurement** to establish clear expectations for trustworthy AI by specifying the assurance evidence government expects suppliers to provide throughout an AI system's lifecycle, helping shape wider market expectations and accelerate the adoption of robust assurance practices.

2. ESTABLISH FOUNDATIONS FOR AI ASSURANCE TO FUNCTION

In addition to aligning existing policy levers, governments will need to build and adapt the legal and market foundations for AI-assurance ecosystems to function effectively over the long term. Clear liability frameworks and mature AI-insurance markets embed AI assurance within the normal operation of AI markets by aligning legal and commercial incentives with good risk management and meaningful assurance. In doing so, they reduce uncertainty for developers, deployers and investors while creating enduring incentives for safer and more trustworthy AI systems.

- **Establish clear liability frameworks** that allocate responsibility to those best placed to manage AI risks, encourage organisations to demonstrate reasonable care through robust AI assurance, and provide the legal certainty needed for assurance markets to develop.
- **Support the development of AI-insurance markets**, recognising their potential to become one of the strongest commercial drivers of meaningful AI assurance. By requiring insurance in appropriate deployment contexts, governments can help insurers create continuous incentives for robust testing, evaluation and risk management while accelerating the development of specialist AI-insurance products.

3. GROW THE AI-ASSURANCE ECOSYSTEM TO REALISE BENEFITS AT SCALE

The previous recommendations establish the direction and the foundations. This is where they become reality. Growing a mature AI-assurance ecosystem enables organisations to access the expertise, institutions and infrastructure needed to deliver assurance at scale, allowing countries to realise the full benefits of AI through safer deployment, faster adoption, stronger public trust and new economic opportunities.

- **Make strategic decisions about where to build domestic capability**, where to partner internationally and where to rely on established assurance markets, concentrating investment where it creates the greatest strategic value while avoiding unnecessary duplication and strengthening international interoperability.
- **Develop the technical capability and human capacity needed to engage with AI assurance**. On the technical side, this means the tooling, infrastructure and appropriate access to AI systems that assurance providers need to evaluate them rigorously. On the human-capacity side, this can be achieved through assurance-provider professionalisation, dedicated talent pipelines and stronger links between frontier-AI expertise, academia and independent assurance providers, ensuring the ecosystem continues to evolve alongside advances in AI.

- **Help stakeholders identify and access competent, trustworthy assurance providers** through appropriate accreditation and professional certification, public directories, and collaborative initiatives that strengthen assurance markets, making high-quality assurance easier to procure and increasing confidence in AI-assurance outcomes.

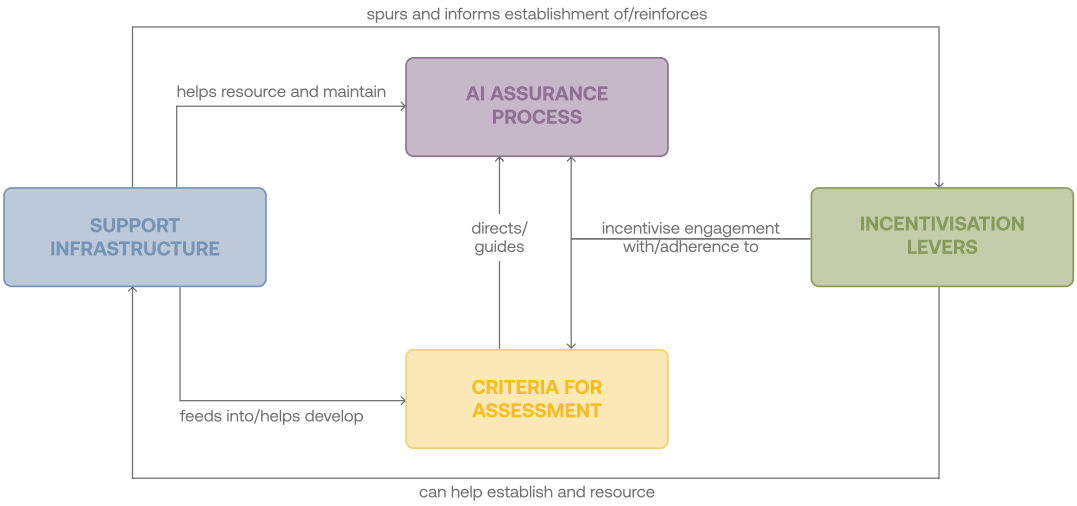
01

What Is AI Assurance?

Assurance is the set of processes, standards, checks and accountability mechanisms used to provide justified confidence that a product or service is safe, reliable and fit for its intended purpose, and to ensure that organisation and service providers behave responsibly and engage in adequate risk-management processes.

FIGURE 1

AI-assurance ecosystem overview map



Source: TBI

Assurance is commonplace across consumer goods and service sectors, from finance, pharmaceuticals and aviation to children's toys and kitchenware. Indeed, assurance processes are so common that their presence is often assumed. Consumers may not know the exact processes

involved, but for the most part, whether someone is buying a new frying pan, hopping on a flight, or popping a painkiller, they do so without worry, confident in the assumption that if a product or service is on the market or publicly advertised, someone, somewhere is holding the relevant industries to minimum safety standards.

As AI becomes increasingly powerful, evolving new application and risks, and being integrated throughout daily life – home, work, school, finance, health – it follows that assurance mechanisms should be employed for AI as well.

As summarised by the Partnership on AI, the former UK Department for Science, Innovation & Technology defines AI assurance as "the process of measuring, evaluating, and communicating the trustworthiness of AI systems, components, and development practices".[1]

The goals of AI assurance are straightforward:

- **To ensure that AI systems being developed and placed on the market are safe, secure and trustworthy**. This is done by setting standards and best practice, providing guidance to support developers and deployers in meeting those standards, and incentivising behaviour throughout the AI lifecycle that yields safer and more trustworthy technologies.
- **To provide mechanisms to support downstream stakeholders in making well-informed decisions** about which AI models and AI-enabled systems they choose to use and rely on. This is done by communicating information about assurance processes engagement and outcomes, for example, through transparent reporting or certifications.
- **To support AI innovation and adoption**, particularly in high-impact and high-risk domains, by reducing uncertainty, building confidence, streamlining compliance, and providing clear pathways for the responsible development, procurement and deployment of AI systems.

How AI-assurance environments are taking shape to deliver on the above goals is more complicated. New actors are emerging in the AI-assurance market, including services for model testing and evaluation, third-party

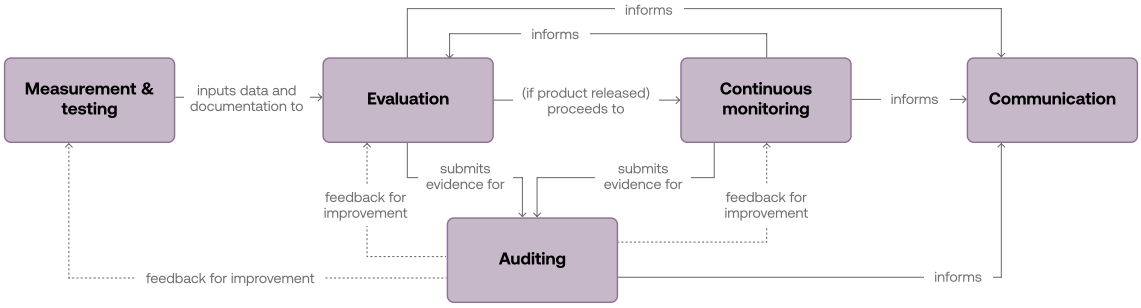
auditors, insurance providers, professional networks and accreditation bodies. Mechanisms for incentivising and enforcing assurance are also being tailored to the unique characteristics of AI (for instance, challenges related to explainability, emergent capabilities, and in some cases, potential for catastrophic harm).[2]

The AI-Assurance Process Sits at the Core

At the heart of AI assurance sits the AI-assurance process: the core set of activities involved in testing and evaluating AI capabilities and risks, monitoring AI system performance post-deployment, auditing AI evaluation results and company risk-management processes, and communicating with downstream stakeholders to convey appropriate levels of trust.

FIGURE 2

Core AI-assurance process activities



Source: TBI

Testing and measurement generate evidence; evaluation and audit determine whether systems, and the people and processes around them, meet relevant expectations; monitoring checks whether this remains true as systems and operating contexts change; and communication enables others to make informed decisions based on findings. Assurance findings may be communicated through evaluation reports, audit opinions, certifications, ratings, disclosures or incident reports.

Critically these activities should operate throughout the AI lifecycle, with assurance evidence informing system design and development, as well as decisions about procurement, deployment and ongoing operation, rather than functioning primarily as a retrospective compliance exercise.

While not detailed in the diagram to maintain readability, within the AI-assurance process there are a variety of entities (objects of assurance) that are assessed by a variety of actors (assurance providers).

Objects of assurance are the entities (technologies, people, organisations) being assessed for safety, reliability, proficiency and trustworthiness:

- **AI models:** Evaluated for capabilities, risks and vulnerabilities, and whether they meet relevant technical or legal requirements.
- **AI systems and applications:** Examined to check the complete system performs in its intended context, including reliability, security, risks and compliance before and after deployment. This system-level assessment is more meaningful than model-level assurance alone: trustworthiness depends on how models interact with data, software, safeguards, human operators and the environment in which they are deployed.
- **AI developers and deployers:** Assessed for implementation of effective governance, risk-management and compliance processes throughout the AI lifecycle.
- **Assurance providers:** Assessed for their competency, consistency, and independence in providing AI-assurance services by certification and accreditation bodies.

Assurance providers are the people and organisations that conduct the assessments and help improve risk-mitigation processes. In practice, single providers will often perform a combination of the functions described below:

- **Technical testers and evaluators:** Measure how an AI system behaves, where it might fail and whether its safeguards work as intended. Many testers and evaluators will also work to help AI developers and deployers act on evaluation findings to make improvements – including strengthening safeguards, addressing identified vulnerabilities and

improving risk-management practices. Testers and evaluators may be internal teams maintained by AI developers and deployers, commissioned teams, or independent third-party entities.

- **Compliance appraisal and attainment services:** Help organisations interpret regulatory and/or procurement requirements, improve processes to be compliant and prepare evidence of compliance.
- **Auditors:** Evaluate claims made by AI developers and deployers about AI product capabilities, safety and risk, and the adequacy of risk-management processes and/or compliance with regulation. They might perform assessments themselves (by providing testing services), or coordinate and draw on evidence provided by specialist evaluators. Auditors can be internal, commissioned second-party or third-party independent.
- **Certification bodies:** Attest that a product, process or organisation meets specified criteria.
- **Accreditation bodies:** Verify the competence, consistency and impartiality of testing, inspection and certification providers.

WHERE DOES GOVERNMENT COME IN?

Government institutions may participate directly in some parts of the AI-assurance process. For example, AI safety institutes already conduct some frontier-model evaluations using privileged access to frontier models. Regulators can also be established to participate in assurance activities and provide appropriate oversight mechanisms. However, governments are unlikely to have the capacity or breadth of expertise to deliver AI assurance at the scale and pace required.

Instead, the primary role of governments is to create the conditions under which robust AI-assurance processes can operate effectively. This includes setting clear expectations (for third-party model assessments or management-system audits, for example) through regulation and public procurement where appropriate, developing the legal and market

foundations that support assurance, such as liability and insurance frameworks, and investing in the institutions, standards, measurement science and professional capability that underpin assurance.

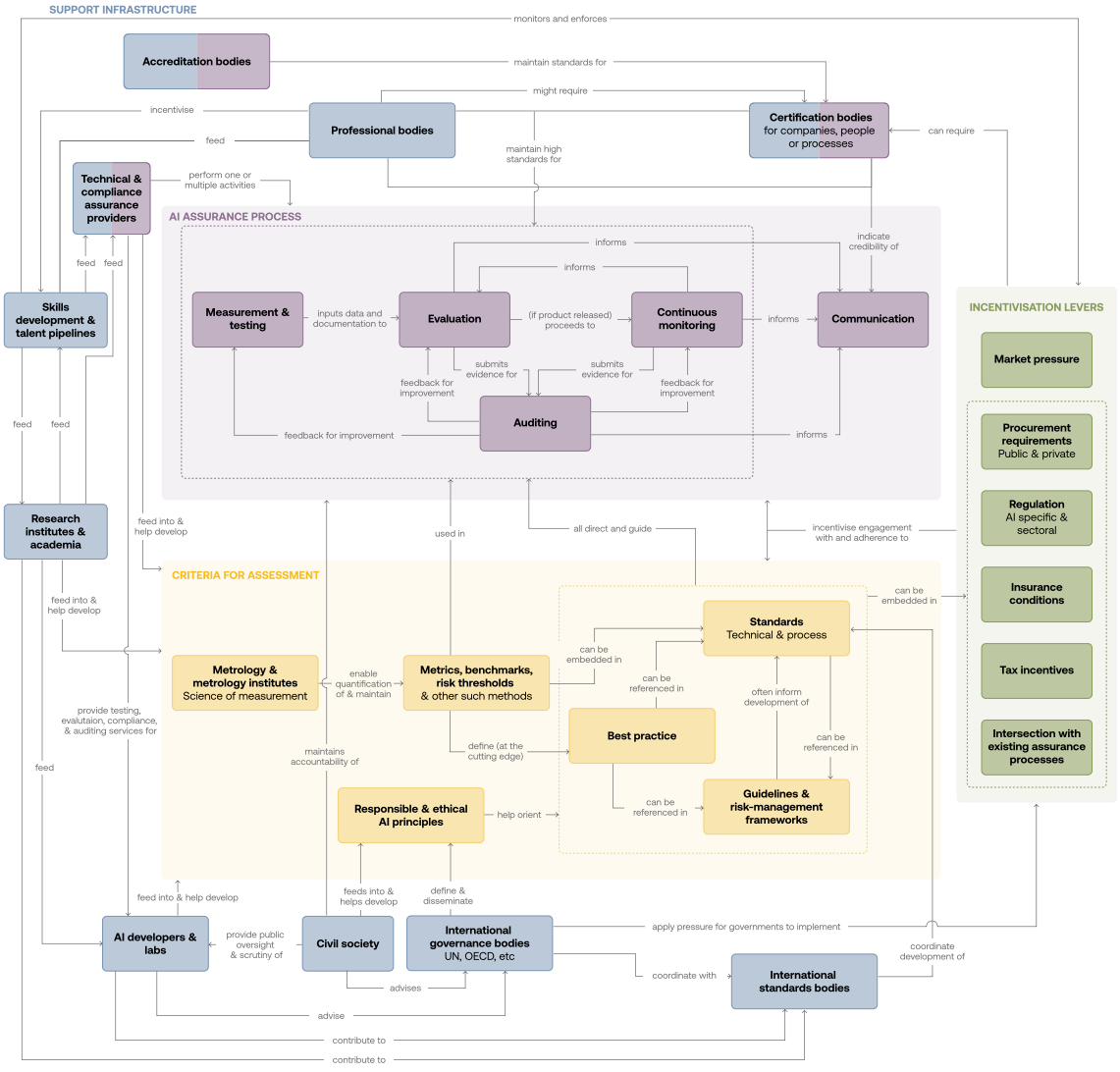
Why AI-Assurance Processes Must Be Embedded in an AI-Assurance Ecosystem

To be effectively deployed, adopted and maintained to a high standard, the AI-assurance process needs to be embedded in a broader AI-assurance ecosystem. This is the full network of institutions, standards, frameworks, policies and incentive mechanisms that supports, reinforces and, in some cases, catalyses the AI-assurance process. The broader AI-assurance ecosystem includes the following (with more details below Figure 3):

- **Criteria for assessment:** The standards, best practices and frameworks that set the bar for what good risk-management and assurance practice looks like.
- **Incentivisation levers:** Economic, policy and regulatory interventions that encourage or require engagement with AI-assurance processes.
- **Additional support infrastructure:** Surrounding network of organisations and institutions that facilitate assurance-ecosystem growth, and the maintenance of its quality, efficacy and credibility.

FIGURE 3

Expanded AI-assurance ecosystem map



Source: TBI[3]

1. CRITERIA FOR EVALUATIONS AND AUDITS

For assurance to be meaningful, there must be agreed criteria against which an AI system, organisational process or assurance provider can be assessed. Without common benchmarks or understanding of best practice, it is hard to differentiate between trustworthy systems and competent assurance providers and those that merely claim to be so. Shared criteria make assurance findings more consistent, comparable and credible, enabling organisations, regulators and users to interpret assurance results with confidence.

- **Standards** translate broad principles into specific, assessable requirements or technical benchmarks. They may focus on organisational processes, such as ISO/IEC 42001, which specifies requirements and provides guidance for establishing, implementing, maintaining and continuously improving an AI management system within the context of an organisation.[4] Other standards articulate technical methods, such as common approaches to testing, evaluating and measuring AI system performance, including stipulations for continuous monitoring of model performance and behaviour post-deployment. AIUC-1, for example, is an agent-focused standard developed to translate high-level requirements into auditable controls for agentic AI systems, backed by independent third-party audit and certification.[5] Meanwhile, international standards bodies, including ISO/IEC, IEEE and ITU, are increasingly developing standards covering AI governance, evaluation methodologies, robustness, transparency and other aspects of trustworthy AI. Importantly, different standards provide evidence about different objects of assurance. Certification against an organisational management-system standard such as ISO/IEC 42001 can provide evidence that appropriate governance and risk-management processes are in place, but does not in itself demonstrate that a particular AI system is safe, reliable or fit for purpose in its intended operating context. System-level claims require appropriate technical testing and evaluation against criteria relevant to the system, use case and risks involved.
- **Voluntary frameworks and guidelines** provide practical, non-binding approaches to AI risk management and assurance. Rather than prescribing specific requirements, they help organisations interpret and

implement good governance practices in different contexts. Examples include the NIST AI Risk Management Framework spearheaded by the US Center for AI Standards and Innovation, which provides a structured approach to managing AI risks across the lifecycle,[6] and Singapore's AI Verify governance and testing framework, which combines governance processes with technical testing guidance to support responsible AI deployment.[7] Other frameworks are more narrowly focused. The open-source AI Execution Framework, for example, provides a common structure for creating tamper-evident records of AI execution that can subsequently be independently verified, helping establish an evidence trail for audit and assurance.[8]

- **Best practice** here refers to the evolving state of the art in AI governance, measurement, testing, evaluation and assured development. It encompasses the methods, benchmarks and governance approaches that are widely regarded by experts as the most robust or effective, even where they have not yet been codified into formal standards. Because AI capabilities evolve and advance rapidly, best practice often develops more quickly than formal standards, allowing assurance methodologies to continuously improve as new evidence, techniques and risks emerge.

These criteria are typically developed through collaborative, multi-stakeholder processes involving governments, industry, academia and civil society. International standards organisations, national bodies (for example, national metrology institutes and AI safety institutes) and non-profits play a central role in sharing expertise, building consensus and ensuring assurance criteria remain technically robust, internationally interoperable and responsive to technological change.

2. INCENTIVISATION AND ENFORCEMENT LEVERS

Assurance improves AI systems only when organisations have reasons to invest in it; testing, evaluation, auditing and certification demand time, expertise and resources. Well-functioning assurance ecosystems combine incentives that reward trustworthy behaviour with mechanisms for

enforcement where necessary. Together, these levers create demand for assurance services, encourage continuous improvement and help make assurance a routine part of AI development and deployment.

- **Market demand:** Customers, users, investors and business partners can favour organisations that demonstrate trustworthy AI practices, making credible assurance a source of competitive advantage. As assurance becomes recognised and trusted, the more it will influence purchasing decisions, investment and market reputation. Market incentives are currently among the strongest drivers for organisations to adopt AI assurance practices. Many organisations already invest significantly in testing, evaluation and risk management beyond regulatory requirements, driven by customer expectations, reputational risk and commercial incentives.
- **Regulation:** Governments can require assurance activities for higher-risk AI systems, establish minimum requirements and impose consequences for non-compliance. Regulation often acts as a catalyst for assurance markets by creating demand for testing, auditing, certification and other services.
- **Procurement requirements:** Buyers from both the public and private sector can require evidence of testing, audits, certifications or compliance with recognised standards as a condition of procurement. As major purchasers of AI systems, governments can lead by example by using procurement to drive demand for assurance and to demonstrate what good AI governance looks like across the wider market.
- **Financial incentives:** Grants, tax incentives, vouchers and subsidies can reduce the cost of implementing assurance, particularly for startups and small and medium-sized enterprises (SMEs). These incentives encourage organisations to adopt assurance practices earlier and help build market capacity before assurance becomes widely expected or mandated.
- **Insurance markets:** Insurers can encourage stronger risk management by requiring evidence of effective governance and assurance practices, while using premiums, exclusions or coverage conditions to reward organisations that demonstrate lower AI-related risks.

3. ADDITIONAL SUPPORT INFRASTRUCTURE

Beyond assurance providers themselves, effective AI-assurance ecosystems rely on a wider network of institutions to develop the knowledge, capabilities and governance needed for assurance markets to function. These organisations help build technical and skills capacity, uphold the quality and credibility of assurance practices, maintain clear lines of accountability, and ensure the ecosystem adapts to evolving AI technologies and risks.

- **Standards bodies:** Organisations such as ISO/IEC, IEEE, ITU and CEN-CENELEC, as well as national standards bodies, convene governments, industry, researchers and civil society to develop internationally recognised standards and best practices. Beyond publishing standards, they foster expert communities, facilitate technical collaboration, provide training and professional certification, and help build international consensus around trustworthy AI.
- **Research and technical institutions and AI companies:** Universities, national laboratories and AI safety and security institutes (AISIs) together with leading AI labs develop the measurement science, evaluation methodologies, benchmarks and evidence base necessary for effective AI assurance. They advance the field by improving evaluation techniques, establishing responsible AI frameworks, identifying emerging risks and validating new assurance methods.
- **Professional bodies and networks:** Organisations such as the International Association of Privacy Professionals (IAPP), BCS, The Chartered Institute for IT and ISACA establish professional competencies, ethical codes and certification programmes while developing the skilled workforce needed to deliver assurance services. They also build communities of practice that facilitate knowledge sharing, professional development and the diffusion of emerging best practices.
- **Intergovernmental organisations:** Bodies such as the United Nations, the Organisation for Economic Co-operation and Development (OECD), G7 and G20 facilitate international cooperation by developing common principles, sharing evidence, building capacity and promoting greater alignment between national approaches to AI governance and assurance.

- **Civil society:** Independent researchers, advocacy organisations and public-interest groups provide external scrutiny, identify emerging harms and ensure assurance systems remain responsive to public values. By holding organisations accountable they help strengthen trust in AI systems and the assurance ecosystem.
- **Existing sectoral assurance systems:** Many sectors, including finance, health care, pharmaceuticals and aviation, already operate mature assurance ecosystems built around rigorous testing, auditing, certification and regulatory oversight. AI assurance can often be integrated into these existing compliance and assurance regimes, leveraging established institutions, technical expertise and cultures of risk management rather than building new systems from scratch.

Together, the elements of AI-assurance ecosystems form robust yet flexible systems of AI governance and risk management. They harness bureaucracy to embed institutional accountability, establish multiple layers of oversight, and create overlapping, mutually reinforcing incentives for responsible development and deployment. AI-assurance ecosystems retain the flexibility to evolve alongside technological change by continuously updating standards and best practices embedded within the assurance process. Due to the numerous interacting nodes within ecosystems, they can also be adapted to different regulatory and institutional environments.

02

The Benefits of AI Assurance

Building robust AI-assurance ecosystems is critical to navigating the AI transition – moving from an era focused almost exclusively on AI innovation and capability breakthroughs to one where AI is rolled out in real-world applications to drive public benefit and economic growth.

Done well, AI assurance will bolster consumer, business and investor confidence, provide middle and developing powers with new avenues to influence domestic and global AI governance, and establish AI assurance as an economic growth opportunity in its own right.

Deliver Better, Safer and More Trustworthy AI Products

The first purpose of any assurance ecosystem is straightforward: to improve the quality, safety and trustworthiness of the products it governs.

Assurance ecosystems already have a clear track record of improving safety and building trust in other sectors. Food-safety systems have reduced contamination and foodborne illness.[9] Pharmaceutical approval, inspection and monitoring systems have significantly improved drug safety and efficacy.[10] Aviation assurance systems were instrumental in making commercial air travel one of the safest forms of transportation.[11]

If assurance infrastructure is considered necessary for food, medicines, aircraft, vehicles and consumer products, it stands to reason that an assurance ecosystem should also be built around a new technology that will touch every aspect of daily life. Here is how:

Set a high bar for safety, reliability, and responsible development and deployment. Standards, benchmarks, testing protocols, evaluation methods, audit criteria, certification schemes and sector-specific guidance can translate broad AI governance principles into clear expectations for mitigating AI risk throughout the AI lifecycle. These expectations may relate to good practice, but also to outcomes, performance thresholds, documentation, risk management, continuous monitoring and accountability.

Provide mechanisms for incentivisation and enforcement. Assurance only improves safety if there are reasons to participate and consequences for falling short. In some cases, consumer pressure or reputational incentives may be enough. In others, regulation, procurement requirements, insurance conditions, liability rules, tax incentives, certification schemes, or market-access requirements may be needed. What is appropriate will vary by sector, use case, risk level and type of AI system.

Create reinforcing layers of accountability and oversight. As set out in the previous chapter, a mature assurance ecosystem cannot rely exclusively on self-assessment by developers/deployers or excessive reliance on a central government function for inspection. Assurance ecosystems can provide a mutually reinforcing network of actors and processes that increase robustness and minimise the risk that safety depends entirely on a single rule, institution, regulator, or assurance provider.

Keep up with rapid technological change. Traditional regulation can struggle to keep pace with fast-moving technologies because legal reform is often slow, formal and politically contested. Assurance ecosystems can be designed to be more agile, with government setting high-level objectives. Meanwhile, the underlying standards, best practices and evaluation methods can be flexibly adapted to incorporate cutting-edge methodologies as technology changes and new evidence emerges.

Expand available technical and compliance capacity. Effective AI governance requires specialised expertise in model testing, cyber-security, data quality, risk assessment and assured development. Many SMEs lack adequate in-house expertise and capacity to maintain adequate technical assurance practices and remain compliant with emerging requirements. Meanwhile, government and regulators might struggle to conduct all compliance checks they require internally. A distributed assurance ecosystem helps address both problems by creating a marketplace of specialist testing providers, evaluators and auditors.

Drive a competitive "race to the top". Together, the above features help set a standard baseline for the quality and safety of AI systems, but they also incentivise developers to go further through competition. For example, New Car Assessment Programmes (NCAPs), established in the United States in 1978, made comparative safety performance visible through independent testing and public ratings.[12] Manufacturers began competing for public recognition by designing vehicles to achieve higher safety scores, often exceeding regulatory requirements. It drove the development and adoption of Electronic Stability Control and Autonomous Emergency Braking before they were mandated. AI assurance can create a similar dynamic by informing downstream stakeholders about AI quality and safety.

Build Trust to Accelerate AI Adoption and Strengthen AI Markets

Markets depend on trust. Organisations invest, governments procure and citizens adopt new technologies only when they have confidence that products are safe, reliable and supported by credible systems of accountability. Where that confidence is absent, deployment slows, investment falls and promising technologies struggle to move beyond pilots. Where it is present, markets become larger, more resilient and better able to support continued innovation.

Indeed, TBI polling found that 38 per cent of UK adults cite a lack of trust in AI as the single biggest barrier to using the technology.[13] Meanwhile, 61 per cent of organisations report delaying or reducing planned AI investments because of concerns about AI-reliability risks.[14]

AI-assurance ecosystems work to address these concerns directly. Through robust systems of testing, evaluation, auditing, reporting and certification, they generate and communicate credible evidence that AI systems meet safety and reliability standards, implement effective governance processes, and ensure accountability when things go wrong. By reducing uncertainty for developers, deployers, investors and consumers, these ecosystems help build the confidence needed for AI markets to grow.

That confidence does more than encourage adoption: it also makes markets more resilient when failures do occur. Air travel remains a highly trusted mode of transport despite its occasional disasters because of how rigorously the industry investigates and responds. Following two consecutive Boeing 737 MAX crashes caused by a faulty automated flight-control system, the fleet was grounded for nearly two years. They were reinstated only after an independent, transparent review by the Federal Aviation Administration and international regulators forced a redesign and new pilot training.[15] In the months following the crashes, there was no market shock as consumer demand for flights continued broadly in line with previous trends.[16]

The same dynamic also helps entirely new markets emerge and scale. Markets characterised by novel technologies, complex risks and uncertainty require shared systems to demonstrate safety, reliability and compliance before participants will invest, procure and adopt at scale. The development of in vitro fertilisation (IVF) illustrates this: following the birth of the world's first IVF baby in 1978, IVF faced significant ethical, safety and governance concerns. The UK's Human Fertilisation and Embryology Act established a framework for licensing, inspection, oversight and transparent reporting through the Human Fertilisation and Embryology Authority, instilling confidence among patients, providers and policymakers.[17] This regulatory support transformed IVF from a controversial experimental procedure into a routine medical service and a multibillion-dollar industry accounting for one in 31 UK births today.[18]

AI-assurance ecosystems can play a similar role, providing the trusted governance infrastructure needed for innovative AI applications to move from experimentation to widespread adoption.

Streamline Compliance and Facilitate Access to Global Markets

Fragmented regulatory environments raise the cost of market entry, especially for smaller firms without the resources to navigate multiple rulebooks. Where firms must meet different requirements jurisdiction by jurisdiction, compliance becomes slower, more expensive and harder to scale. But when countries reference the same standards, recognise common best practices and accept one another's conformity assessments, a product checked once can be accepted across many markets.

For example, before the International Financial Reporting Standards (IFRS), a company listed in several countries had to restate its accounts under each national rulebook, but today it can file one set of IFRS accounts accepted across markets. Common rules make accounts comparable across borders, making them cheaper for companies to produce and clearer for investors.[19]

However, it is not a foregone conclusion that AI-assurance ecosystems will develop in this way. Binding rules such as the EU AI Act sit alongside a range of voluntary instruments such as the GPAI Code of Practice, Singapore's AI Verify framework and the US NIST AI Risk Management Framework. These all present overlapping but not identical expectations. Although different approaches allow for greater flexibility across jurisdictions as the technology develops, greater international collaboration will be essential if AI assurance is to deliver the compliance-streamlining and international market-access benefits seen in other assurance sectors.

Outcome-based regulation – which specifies the results a system must achieve such as defined levels of safety, reliability or performance, rather than prescribing the exact processes or documentation used to demonstrate them – can make compliance across jurisdictions substantially easier. By focusing on whether systems are able to be assured and deemed trustworthy in the first place, and on the evidence that demonstrates this, it allows the same underlying system to satisfy different regulatory regimes even where their detailed requirements differ. As AI capabilities advance, much of the routine "formatting" of compliance is likely to become easier to produce and even automate, but only where the underlying system has been designed so it is capable of being assured from the outset.

The UK's Military Aviation Authority offers an example.[20] Its approach to assuring AI in safety-critical air systems is outcome-focused. It treats the use of AI as conceptually the same as any other system-development approach, requiring that the resulting output be assured to a level of confidence commensurate with the risk of failure, with performance targets defined for the operating context. Where AI replaces an existing capability, it can be assessed against whether it delivers at least equivalent performance, rather than against a prescribed method, and applicants are directed to a range of recognised assurance methodologies rather than a single mandated format. This emphasis on demonstrated outcomes over prescribed process is what makes assurance evidence more portable – the same result can be evidenced in more than one way and still be recognised. This is the foundation on which cross-border acceptance is built.

Underpin AI Sovereignty

For most countries, AI sovereignty is not about building and controlling every layer of the stack domestically, as we explored in a previous paper: *Sovereignty in the Age of AI: Strategic Choices, Structural Dependencies and the Long Game Ahead*. It is about retaining strategic agency, being able to make informed choices about which technologies to procure, how they are deployed, what risks are acceptable, and what standards of safety, trustworthiness and accountability they must meet.

Countries that build credible assurance capacity can help determine what responsible AI use looks like at home and abroad.

A strong domestic AI-assurance ecosystem underpins sovereignty in two ways.

It helps countries defend their own interests in AI procurement and adoption. Even if they are not major producers of AI, countries still need the capacity to make informed decisions about the systems they procure and deploy. They need to know what good looks like, what evidence to require and what assurance mechanisms are credible for their own market and public sector. That means building domestic testing, evaluation and measurement capacity and, when appropriate, pooling it with trusted partners. This is part of the intended function of building national AI safety institutes, which aim to equip governments with a scientific understanding of the risks posed by advanced AI, and the expertise needed to test, evaluate and collaborate.[21] A broader assurance ecosystem, including private providers and certification systems, can help extend technical capacity beyond the public sector.

Building domestic assurance capability provides a route to influencing global AI governance and shaping the future of AI development. However, assurance capability alone is not enough. Countries need to be active participants in the development and deployment of AI. A strong assurance ecosystem promotes safer and more confident AI adoption at

home while demonstrating governance models that others can observe and emulate. Early movers will shape their own domestic rules, as well as the precedents, methods and standards that other countries later adopt.

It is a common phenomenon across assurance markets. Sweden, for example, pioneered its Vision Zero approach to road safety in 1997, shifting some responsibility for safety from individual road users to transport-system designers.[22] This approach has since influenced road-safety frameworks from Australia and Europe to Latin America and Africa. In another examples, Singapore became the first country to approve the commercial sale of cultivated ("lab grown") meat in 2020, establishing a regulatory pathway that has become a common reference point for other countries developing frameworks for accelerating the introduction and regulation of novel foods such as cultivated meat and synthetic-protein products.[23]

Similar dynamics are beginning to emerge around AI assurance. The US has helped establish an early benchmark through NIST's AI Risk Management Framework (NIST AI RMF) – a voluntary framework that has become one of the world's most widely referenced approaches to AI risk management.[24] The UK is seeking to play a complementary role through its new Centre for AI Measurement, housed within the National Physical Laboratory (NPL), which aims to advance the science of AI measurement and evaluation. The idea is to also position the UK as a global hub for developing and maintaining the benchmarks, measurement techniques and evaluation methodologies that underpin AI assurance.[25]

Beyond building domestic institutions, countries can also shape global AI governance by influencing the international standards and technical guidance that support assurance markets. Technical standards developed through bodies such as ISO/IEC, IEEE and the International Telecommunication Union (ITU) increasingly define how AI systems are evaluated, documented and governed across jurisdictions.[26,27,28] Governments that invest in these processes, contribute technical expertise, lead working groups or help develop reference methodologies can shape the global benchmarks for trustworthy AI.

Capture the Economic Opportunity of the AI-Assurance Industry

Assurance is not just standards and policy. It is the service layer that grows up around them – testing, evaluation, auditing, certification, accreditation and advisory work – that turns a standard into something tangible. Supplying these services is a large and fast-growing market. Modelling by the Institute for AI Policy and Strategy (IAPS) suggests the global market for AI-assurance technology could reach $276 billion by 2030.[29]

AI systems are not one-off products: they are updated, retrained, monitored and redeployed, making assurance a recurring need rather than a single compliance event. That creates scope (indeed, a necessity) for continuous testing, monitoring, audit, certification and compliance tooling, rather than a one-time legal check.

For countries with the right building blocks – professional services depth, standards institutions, regulatory expertise, accreditation capacity and skilled technical talent – this is a significant economic opportunity. Countries that build assurance capability early can capture a disproportionate share of a fast-growing global market, sell assurance services abroad, and help shape the standards and methods others later adopt.

Even where capturing a global assurance market is unrealistic, there is still a notable opportunity for developing a domestic market. This is particularly true for services that ensure AI models and applications reflect local linguistic, cultural, regulatory and risk contexts, rather than relying on imported assessments designed for high-income, English-speaking settings.

Is There a Third-Party Assurance Market for Frontier AI?

Most commercial AI-assurance activity today focuses on enterprise AI rather than frontier foundation models. This reflects where AI is currently being deployed at scale and where organisations require external support to assess risks, demonstrate compliance and build trust. As a result, a growing ecosystem of testing, auditing and governance providers has emerged to support enterprise AI adoption.

The market for independent assurance of frontier AI models is much less developed. Today, much of this work is undertaken either within government institutions, such as AISIs, and by a small number of specialist non-profit organisations, such as METR and AVERI. Commercial providers of independent frontier-model testing and evaluation remain relatively rare, with firms such as Faculty AI among a small number operating in this space.

Several factors have slowed the emergence of a commercial market for frontier-AI assurance:

- **Limited commercial demand.** There are only a handful of frontier AI developers, creating a very small customer base for specialist assurance providers.
- **Weak incentives to procure independent assurance.** Frontier labs can often obtain external evaluations through non-profit collaborations that are partly motivated by public-interest objectives. Whether equivalent demand exists for commercial assurance services is uncertain. Regulatory requirements, procurement expectations or insurance markets may be needed to create sustained demand.

- **Expertise remains highly concentrated.** Much of the world's frontier evaluation expertise is concentrated within leading AI laboratories and a small number of specialist organisations. While frontier evaluations are technically demanding, the narrow career pipeline has also likely limited the growth of an independent assurance workforce.
- **Trust and access create barriers to entry.** Privileged access to frontier-model internals (such as model weights, training information or unrestricted querying) is required for thorough assessment of their capabilities. Frontier labs typically grant model access to organisations they already know and trust. This is appropriate given the sensitivity of frontier models but makes it difficult for new assurance providers to gain experience, build credibility and compete.

These barriers may not last indefinitely. As frontier AI systems are deployed in government, critical infrastructure and other high-risk settings, demand for independent assurance is likely to grow. Specialised risks like biosecurity, cyber-capability or loss-of-control evaluations may naturally support independent specialist providers whose expertise is uneconomical for individual frontier labs to maintain in-house.

However, a competitive market for frontier-AI assurance is unlikely to emerge on its own. Like assurance markets in many other sectors, governments must create the conditions under which independent assurance providers can compete and scale. Risk-proportionate regulation, public-procurement requirements, liability frameworks and insurance markets can all help generate sustained demand, while investment in workforce development, evaluation science and independent assurance institutions can expand the supply of capable providers.

It is highly unlikely that small independent not-for-profits remain the solution for frontier assurance in future, as the legal, technological and scale demands will necessitate large private enterprise involvement.

03

How Leaders Can Drive AI-Assurance Ecosystem Growth

Well-functioning AI-assurance ecosystems can improve the safety and quality of AI systems, accelerate adoption, strengthen public trust, reduce compliance burdens and create new economic opportunities. But identifying the components of an assurance ecosystem is only the first step. The greater challenge is to coalesce them into a coherent system that creates the right incentives, provides sufficient clarity to market participants and remains flexible enough to evolve alongside a rapidly changing technology.

Political leaders need to create the conditions that allow AI-assurance ecosystems to emerge and mature. Governments do not need to deliver every assurance service themselves. Rather, their role is to establish clear expectations, create demand for assurance, invest in enabling infrastructure where markets alone will underprovide it, and to encourage international interoperability of assurance standards and processes from the outset.

If assurance ecosystems are to develop successfully, leaders will need to address some key challenges:

- **Limited mechanisms for validating and communicating assurance**. Many organisations already invest substantially in AI testing, evaluation and risk management, driven by commercial, reputational and regulatory considerations. However, inconsistent standards, benchmarks and independent verification make it difficult for customers, regulators and insurers to distinguish between assurance approaches and compare their quality, while demand for independent third-party assurance remains uneven.
- **Limited clarity about what assurance should deliver**. Uncertainty around government expectations, acceptable assurance methods and evidentiary requirements makes it difficult for providers and customers to invest with confidence.

- **Difficulty identifying trusted assurance providers**. Organisations often lack reliable ways to assess the competence, independence and quality of assurance providers, making procurement difficult and limiting confidence in assurance markets.
- **Immature measurement science**. Many aspects of AI performance, safety and trustworthiness remain difficult to measure consistently, limiting comparability across systems and reducing confidence in assurance outcomes.
- **Limited access to AI systems and information**. Independent assurance requires privileged access to technical information or model internals that developers often have legitimate reasons to restrict. Reliance on developer-granted access can limit the depth of assessments, create barriers to entry and generate conflicts of interest that undermine assurer independence.
- **Poorly understood risks**. Unlike established assurance domains where hazards are well mapped, AI assurance must address risks that remain uncertain, difficult to observe directly and only emerge in deployment.
- **Standards that struggle to keep pace with technological change and provide sufficiently specific guidance**. Consensus-based standards promote interoperability and stability, but often materialise more slowly than frontier AI capabilities and remain too high-level to answer many of the practical questions developers encounter.
- **Workforce and capability constraints**. There is a shortage of professionals with the multidisciplinary expertise needed to evaluate AI systems effectively, and limited organisational capability to procure and use assurance services well.
- **Fragmented international standards and governance approaches**. Multiple frameworks and standards (for example, NIST AI RMF, ISO/IEC 42001, CEN-CENELEC standards, and emerging sector-specific approaches) provide valuable guidance, but their differing scopes and requirements increase complexity, raise compliance costs and make international interoperability more difficult.

The chapters that follow are organised around three complementary roles for government in building AI-assurance ecosystems.

- **Establish AI assurance as a strategic national priority** and use the policy levers they already possess to create demand and set clear expectations.
- **Develop the new foundations and market mechanisms** needed to support a mature assurance ecosystem.
- **Invest in the capabilities, infrastructure and international partnerships** that allow those ecosystems to grow and deliver their full benefits.

Together, these interventions enable countries to reap the opportunities of AI assurance.

04

Make AI Assurance a Strategic National Priority

The previous chapters argued that AI assurance is more than a governance tool; it is a strategic capability. Countries that succeed in building robust AI-assurance ecosystems will be better placed to accelerate AI adoption, strengthen public trust, reduce compliance burdens, capture new economic opportunities, and shape the international norms and standards that govern AI. This must now be recognised as a strategic national priority.

Establishing AI assurance as a national priority involves aligning government policy levers to translate consumer and investor demand for more trustworthy AI into consistent demand for high-quality AI assurance. This means shifting from broad expectations of "responsible AI" to providing clear incentives for organisations to invest in testing, evaluation, auditing and certification.

Regulation and public procurement are among the most powerful tools immediately at governments' disposal for driving this shift while establishing clear expectations about what credible AI assurance should demonstrate.

Recommendation: *Establish a flexible regulatory framework for AI assurance that creates enforceable obligations to enable an independent assurance ecosystem.*

As AI systems become more capable and are deployed in increasingly consequential contexts, self-assessment alone is not sufficient. Stakeholders interviewed for this paper consistently highlighted two closely related challenges: the absence of clear, enforceable expectations about what high-quality AI assurance should demonstrate, and the need for independent scrutiny of the most critical AI systems. Regulation provides the foundation for addressing both. Governments should establish legal obligations that require appropriate testing, evaluation and ongoing monitoring for highly capable frontier models and AI systems deployed in high-risk contexts, while creating the institutional framework that allows for

independent assessment and enforcement of those obligations. This includes recognised standards, accreditation, and independent third-party verification and audit. Together, these measures will create sustained demand for high-quality assurance services while maintaining accountability.

A key question is how those regulatory obligations should be implemented. At one end of the spectrum, legislation can establish high-level obligations by requiring appropriate testing, evaluation and monitoring while leaving recognised standards, assurance providers and expert bodies to determine the specific methodologies. For example, the Independent Verification Organisation (IVO) model proposed by non-profit organisation Fathom reflects this logic: government sets outcome-based safety goals and authorises a market of independent, expert-led verification bodies to innovate and compete to develop the best technical criteria and methods for assessing whether AI products meet those goals.[30] This model has recently gained legislative traction in California, where the state legislature passed SB183 in August: it directs the Government Operations Agency to establish a system for designating qualified IVOs to independently assess AI systems and models.[31]

At the other end of the spectrum, legislation could specify detailed technical requirements directly, such as particular evaluation protocols, documentation requirements, information-sharing obligations or minimum standards for independent assessments.

Both approaches have strengths and weaknesses. High-level regulation provides greater flexibility, distributes technical expertise across the assurance ecosystem and allows evaluation methodologies to evolve alongside the technology. However, it also creates a wider range of acceptable assurance practices, increasing the importance of complementary mechanisms such as accreditation, professional certification and independent verification organisations. More prescriptive regulation provides greater consistency and clarity, but also risks becoming outdated as AI capabilities and assurance methodologies rapidly evolve. Stakeholders pointed to the implementation of the EU AI Act through the CEN-CENELEC standards process as an illustration of this challenge.

In practice, mature assurance systems are likely to converge on a hybrid approach. This follows the same legislative principle outlined in a previous TBI paper, *Getting the UK's Legislative Strategy for AI Right*. Primary legislation should establish stable legal duties while delegating rapidly changing technical questions to regulators and expert institutions. For example, for high-stakes AI systems, "assurance cases" can provide a structured way to demonstrate that these obligations have been met. Legislation establishes the high-level claims that organisations must demonstrate, while regulators and technical bodies determine how those claims should be substantiated and what constitutes sufficient evidence.[32] Independent third-party assurance providers then play a central role in assessing the claims and evidence. Such an approach preserves the stability of regulation while allowing assurance practice to move with the technology.

Finally, AI-specific assurance requirements should build on existing sectoral regulation wherever appropriate. Sectors such as health care, financial services and aviation already possess mature assurance and compliance infrastructure. Rather than creating parallel systems, governments should clarify whether existing regulatory obligations explicitly apply to AI-enabled products and services. For example, the UK's Medicines and Healthcare products Regulatory Agency (MHRA) regulates AI as a subset of software under the existing Medical Devices Regulations 2002 (MDR 2002).[33] The MDR 2002 provides an established system of pre-market assessment and post-market monitoring for medical devices. Rather than creating a separate regime for AI, the MHRA uses initiatives such as its AI Airlock programme to test how the existing regulatory framework applies to AI-enabled medical devices and where targeted changes are needed.[34] Such an approach avoids gaps in coverage while leveraging well-established assurance mechanisms.

Recommendation: *Establish public-procurement requirements for AI that clearly specify the assurance evidence government expects suppliers to provide, and the governance, monitoring and maintenance practices expected throughout an AI system's lifecycle.*

Public procurement is one of the most effective ways governments can shape emerging markets. While generally less coercive than regulation, procurement requirements send a strong signal about what good AI assurance looks like. By clearly setting out the testing, evaluation, documentation, monitoring and governance processes expected, governments create demand for high-quality assurance while giving developers, deployers and assurance providers much-needed clarity about what is expected of them.

During stakeholder interviews, AI developers and deployers described struggling to understand what evidence customers wanted to see, while procurement teams reported uncertainty about what questions they should ask when purchasing AI systems. Well-designed procurement requirements help solve both problems by establishing a common set of assurance expectations for government suppliers.

Government purchasing decisions frequently become reference points for private-sector procurement, particularly in emerging technology markets where best practice is still evolving. Procurement requirements can therefore accelerate the development of broader industry standards and encourage organisations to adopt robust assurance practices, even where regulation has not yet made them mandatory.

Towards this end, in 2025, IEEE published IEEE 3119, the first international standard dedicated specifically to the procurement of AI and automated decision systems.[35] The standard provides practical guidance for embedding AI risk management throughout the procurement lifecycle, illustrating how procurement itself is becoming more important for advancing trustworthy AI. Meanwhile, in the US, the Office of Management and Budget's 2025 memorandum M-25-22 similarly established government-wide requirements for federal AI acquisition, including additional safeguards for high-impact AI.[36] These requirements respond directly to challenges documented by the US Government Accountability Office, pertaining to difficulties defining AI requirements, evaluating systems before purchase, and monitoring their performance over time.[37] While it is

too early to assess their overall effectiveness, the US approach illustrates how procurement policy can translate assurance needs into concrete expectations for government suppliers.

Governments should not delay in establishing procurement requirements simply because it is unclear which assurance methodologies, standards or measurement approaches will prove most effective in the long term. One of procurement's greatest strengths is its ability to adapt alongside technology. Unlike legislation, procurement requirements can be updated relatively quickly as assurance methodologies mature, measurement science advances and new best practices emerge. Governments should use this flexibility to get started, set clear expectations for the market and refine them over time.

05

Establish the Foundations for AI Assurance to Function Effectively

Governments will need to establish the next generation of foundations for AI-assurance ecosystems to function effectively in the long term. These foundations move beyond creating demand for assurance towards embedding it within the legal and economic structures that shape how AI systems are developed, deployed and brought to market.

Liability frameworks and AI insurance markets in particular are critical for supporting meaningful AI assurance, allocating responsibility and reinforcing good risk management.

Recommendation: *Establish clear liability frameworks as a critical foundation for effective AI assurance.*

Liability is the scaffold on which effective AI-assurance incentives are built. Regulation, procurement and insurance all depend on organisations understanding who ultimately bears responsibility if AI systems cause harm. Without clear liability, incentives weaken. Organisations can shift responsibility to others, which means assurance becomes a compliance exercise rather than a genuine risk-management tool, while victims face uncertainty about who should be held accountable.

Effective liability frameworks should allocate responsibility to those best placed to understand and mitigate the relevant risks.[38] They should encourage organisations to exercise reasonable care – not by imposing liability for every adverse outcome, but by creating strong incentives for organisations to follow recognised AI-assurance practices and demonstrate that they have met the expected standard of care.

For most enterprise AI applications, this means deployers should remain responsible for harms arising from their use of AI systems, just as they remain responsible for the employees, software and equipment they choose to use. Deployers decide whether, where and how AI is used, what

safeguards are implemented and whether a system is appropriate for a given context. End users should bear responsibility only where they have clearly disregarded documented instructions or explicit conditions of use.

Frontier AI systems present a more difficult challenge. Deployers often lack sufficient visibility into the capabilities, limitations and risks of highly capable foundation models, particularly where models are proprietary and continually updated. Developers are often the only actors capable of evaluating certain model-level risks before release. Liability frameworks should acknowledge that frontier-model developers are responsible for testing, evaluating and documenting models, and for providing sufficient information about capabilities, limitations and known risks so downstream deployers can make informed decisions. Responsibility should reflect the information and control available to each actor, rather than being placed solely on developers or deployers.

The picture becomes even more complex as AI begins to automate parts of the assurance process itself. Automated testing, evaluation and compliance tools will likely become essential if assurance is to scale alongside AI deployment and operate continuously as systems and deployment contexts change. These systems introduce their own liability chains, raising questions about the responsibilities of assurance providers, developers of assurance tools and organisations relying on their outputs.[39]

Overall, liability frameworks need to clearly distribute responsibility across the AI lifecycle, reflecting each actor's visibility into relevant risks, degree of control and ability to prevent or mitigate harm. This clarity is essential both to avoid gaps in accountability and to ensure that each actor has incentives to manage the risks within its control. Governments should begin clarifying these relationships now, before automated assurance becomes commonplace.

Recommendation: *Build towards mandatory insurance for AI systems whose failure could cause material harm to individuals, organisations or society.*

AI-insurance markets have the potential to become one of the most powerful market mechanisms for driving meaningful AI assurance. Because insurers have a direct financial interest in reducing risk, they are motivated to differentiate between genuine assurance and superficial compliance. This incentivises them to reward organisations that display strong risk-management practices.

Over time, governments should aim to require insurance for AI systems capable of causing material harm, as they do for other activities such as driving. Under such a system, governments would determine where insurance is required, while insurers would determine what testing, evaluation and risk management they need to see before underwriting a system. This market-based approach creates continuous incentives for organisations to maintain meaningful AI assurance rather than treating it as a tick-box compliance exercise. Meanwhile insurers can flexibly update underwriting expectations as assurance methodologies improve and new evidence emerges.

But mandatory insurance will only work if insurers can assess and price the risks they are being asked to cover. Those conditions do not yet exist across much of the AI market. The market for standalone AI insurance – policies designed specifically to cover AI-related risks – remains nascent, and insurers have limited data both on how frequently different AI failures occur and on the financial losses that result. If governments require insurance before insurers can confidently price these risks, the policy could backfire: insurers may decline to provide cover, charge prohibitively high premiums or offer restrictive policies that provide little meaningful protection. Where insurance is a condition of deployment, this could prevent otherwise viable AI systems from reaching the market and disproportionately disadvantage startups and smaller providers.

Governments should therefore focus first on building the evidence base and underwriting capability needed to support a viable market. Incident-reporting and information-sharing mechanisms can help by generating evidence about what types of AI failures occur, how frequently they occur and the harms and financial costs that result. Stakeholders interviewed for this paper repeatedly called for such mechanisms to strengthen the

evidence base available for underwriting. Insurers also need clear standards for how to evaluate how effectively individual organisations manage AI risks. AIUC-1, for example, was developed with insurers to provide a standardised assessment of governance and technical controls for AI agents that can inform underwriting.

There is also an immediate challenge in how AI risks are treated within existing insurance. AI-related harms may fall within established insurance lines, such as professional liability, technology errors and omissions and cyber insurance, creating uncertainty where policies do not explicitly address AI-related risks. Governments should therefore ensure that relevant policies clearly state whether AI-related risks are covered or excluded, avoiding "silent AI". For example, in the response to "silent cyber" concerns, the UK Prudential Regulation Authority required insurers to clarify how existing policies treated cyber risks, rather than leaving it uncertain whether cyber-related losses were covered.

As the evidence base and underwriting capability mature, governments can progressively introduce insurance requirements, beginning in contexts where risks and potential losses are sufficiently well understood and expanding them as the market develops. For jurisdictions that favour market-based approaches to governance, insurance could prove a particularly valuable complement to regulation, aligning commercial incentives with public-safety objectives while reducing the need for highly prescriptive regulatory requirements.

Can AI Insurance Scale to Include Frontier AI?

Today's standalone AI-insurance market is largely focused on discrete, measurable harms. Frontier AI presents a much harder challenge.

Potential harms are significantly greater, may be highly correlated across thousands of downstream users and are difficult to quantify because there is little historical data on which to estimate probabilities or price premiums. Frontier-AI developers are typically well capitalised to absorb ordinary litigation costs themselves, while the largest catastrophic risks may exceed the underwriting capacity of conventional insurance markets. These factors help explain why today's standalone AI-insurance market remains tightly circumscribed.

Rather than viewing these challenges as reasons to abandon insurance, researchers are exploring new insurance structures designed specifically for frontier-AI risks.[40,41,42,43] Potential approaches include:

- **Catastrophe bonds and other capital-market instruments:** Portions of catastrophic AI risk are transferred to investors willing to bear low-probability, high-consequence events. This expands the available capital to address extreme losses that traditional insurance markets may not be able to cover, making it more feasible to insure risks that exceed the capacity of conventional insurers.
- **Insurance pools and mutuals:** Frontier-AI developers and insurers collectively pool capital and share catastrophic risks that would be difficult for any single insurer to absorb. Because participants collectively

bear the costs of failures, they also have strong incentives to maintain high safety standards, share information about emerging risks and jointly invest in independent testing, evaluation and measurement infrastructure.

- **Parametric insurance:** Policies pay out automatically once predefined technical or operational triggers are met, obviating lengthy claims assessments. This may prove particularly valuable for distributed AI failures, where establishing causation and calculating thousands of individual losses could otherwise delay compensation for months or years.
- **Dynamic underwriting:** Rather than pricing AI risk once at the start of a policy, insurers could continuously update premiums, coverage conditions and assurance requirements as new evaluation results, incidents or risk indicators emerge. This would allow insurance to evolve alongside rapidly changing AI capabilities while creating ongoing incentives for developers to maintain high safety standards rather than treating assurance as a one-off compliance exercise.
- **Government backstops:** Similar to arrangements used for terrorism, flood and nuclear risks, governments could provide last-resort protection for losses that exceed the capacity of private insurance markets. This would allow private insurers to participate in markets that might otherwise be considered uninsurable, while limiting their exposure to truly catastrophic tail risks. Socialising losses in this way is not ideal and therefore should be considered carefully before implementing: it can suppress risk-reflective insurance pricing if government protection is not priced at the full cost of the risk and can create a moral hazard if firms expect the state to absorb extreme losses.

None of these approaches is yet mature enough in application for underwriting frontier AI, but they are promising leads on how to develop the extension of insurance models to frontier AI. Rather than serving solely as a mechanism for compensating losses, insurance could become an increasingly important driver of better assurance, stronger risk management and higher safety standards across the frontier-AI ecosystem.

06

Growing the Ecosystem to Realise AI Assurance and Its Benefits at Scale

The final challenge is bringing the ecosystem together. Even with strong incentives and well-designed legal and market frameworks, AI assurance cannot succeed in delivering safer and more trustworthy AI, accelerating adoption, or creating new economic opportunities unless organisations can readily access the expertise, institutions and technical infrastructure needed to deliver it in practice.

The recommendations below outline how governments can bring these components together by developing domestic capability, leveraging international partnerships, and strengthening the markets and institutions that enable organisations to access and engage confidently in high-quality AI assurance. By investing strategically, avoiding unnecessary duplication and building on shared international infrastructure wherever appropriate, governments can develop assurance ecosystems that are more effective and more efficient while strengthening their role within the emerging global AI-assurance economy.

Recommendation: *Decide when to build, partner or depend in order to access AI-assurance capability and maintain international cohesion.*

Bringing an AI-assurance ecosystem together requires governments to think strategically about which capabilities should be developed domestically, which are best built collaboratively and which can be accessed through mature international markets. AI-assurance ecosystems are inherently both national and international. While every government will require sufficient domestic capability to procure, oversee and use AI assurance effectively, many of the ecosystem's underlying components (notably for measurement science, testing methodologies, standards, accreditation systems and technical infrastructure) are shared public goods that are more effective when developed collaboratively across borders.

No country will build a world-leading capability across every part of the AI-assurance ecosystem, nor should they seek to. Governments should instead make deliberate strategic decisions about which capabilities to develop domestically, which to build collaboratively with international partners, and which to access through mature international assurance markets and institutions. This allows countries to concentrate investment where it delivers the greatest strategic value while avoiding unnecessary duplication of effort.

Fragmentation presents a particular challenge for AI assurance. If every jurisdiction develops its own testing methodologies, assurance frameworks, accreditation systems and technical standards in isolation, developers, deployers and assurance providers will face duplicative requirements, higher compliance costs and reduced interoperability. Wherever possible, governments should align with, contribute to and build upon existing international initiatives rather than creating competing national approaches. A globally connected assurance ecosystem will ultimately be more efficient, more credible and better able to support the international development and deployment of AI.

Governments should therefore strategically decide where to build, partner or depend when developing AI-assurance capability and infrastructure (mirroring the choices countries face in securing access to technical AI resources, explored further in TBI's paper *Sovereignty in the Age of AI: Strategic Choices, Structural Dependencies and the Long Game Ahead*).

Build when there is strategic advantage. Invest in domestic assurance capabilities that align with national strengths, economic priorities or sovereign requirements, or where a country can make a distinctive contribution to the global assurance ecosystem. The UK, for example, is leveraging longstanding strengths in AI research, safety and measurement science, investing through institutions such as the AI Security Institute and the NPL's Centre for AI Measurement to build assurance capabilities with international as well as domestic value. Singapore is similarly building on its position as a trusted, internationally connected technology and business

hub, investing in the AI Verify Foundation, the Global AI Assurance Sandbox and internationally oriented governance frameworks to position itself as a trusted testbed and convening hub for AI governance assurance.

Partner when capability is best developed collectively. Many of the foundational components of AI assurance (for example, measurement science, benchmark development, testing methodologies and standards) function as shared infrastructure for the global assurance ecosystem. This does not mean no country should build them; rather, where one country develops such capability from a position of national strength, as with the UK's investment in measurement science, the benefits are best realised when the results are shared and adopted internationally rather than duplicated jurisdiction by jurisdiction. Stakeholders consistently identified that the lack of common approaches to measuring AI performance, safety and trustworthiness has been consistently cited as one of the greatest barriers to scaling assurance markets internationally. Without internationally recognised methods and benchmarks, assurance results become difficult to compare across providers and jurisdictions, increasing costs, reducing trust and contributing to regulatory fragmentation.

Developing this kind of infrastructure is rarely feasible for countries to do independently. Just as international metrology systems provide the common measurement foundations for global trade, AI assurance will require internationally recognised methods for measuring the safety, reliability and performance of AI systems. National metrology institutes can play an important role in advancing this science and upholding "best practice" concepts that can be referenced in global standards, guidelines and regulation, but its value hinges on international coordination and recognition through networks of peer institutions. The same principle applies across the wider assurance ecosystem. Countries can pool expertise, share costs and shape international best practice through collaborative initiatives on standards, accreditation and assurance methodologies, while avoiding unnecessary duplication of complex technical infrastructure.

A growing range of international initiatives illustrate how different elements of this shared infrastructure can be provided collectively. For example, the International Network for Advanced AI Measurement, Evaluation and

Science (NAAIMES) brings together government institutes representing nine countries and the EU, in order to strengthen shared scientific understanding and internationally recognised approaches to measuring and evaluating advanced AI capabilities.[44] The newly established Appia Foundation addresses a related challenge further along the assurance process, convening organisations from across the AI value chain and jurisdictions to develop specifications to help translate foundational international standards into assessable criteria. These can support conformity assessment across different legal, regulatory and contractual contexts, for example.[45]

The AI Quality Infrastructure (AIQI) Consortium, meanwhile, connects accreditation bodies, conformity-assessment bodies, standards organisations and research institutes to advance common standards, harmonised approaches to AI testing and mutual recognition of AI certifications.[46] Building further, the Global Accreditation Cooperation Incorporated (Global ACI) is working to establish a multilateral recognition arrangement to support international acceptance of accredited conformity-assessment results, thereby reducing duplicative assessment across markets.[47][48]

However, international accreditation does not automatically translate into regulatory recognition. For example, under the EU AI Act certain AI systems must be assessed by an independent body formally authorised within the EU (a “notified body”). This means an internationally accredited assurance provider may still be unable to provide the assessment required for access to the EU market. International interoperability therefore depends not only on common standards and accreditation, but also on regulatory recognition of equivalent providers and assurance results across borders.

Governments should examine these interfaces when designing AI-assurance regimes and pursue mutual recognition or equivalent mechanisms where appropriate, allowing technically credible assurance undertaken in other jurisdictions to be recognised without unnecessary duplication while preserving appropriate regulatory oversight.

This approach will be particularly important for smaller economies and developing countries. Rather than attempting to build frontier evaluation capability themselves, countries might establish a national AI-assurance or procurement office responsible for setting assurance requirements, recognising trusted providers and maintaining partnerships with international testing institutes. This would allow countries to retain the domestic capability needed to determine what assurance they require while accessing specialist technical capabilities through trusted international partners.

Depend strategically when mature international capability already exists. Governments can often achieve better outcomes by relying on established, trusted international assurance markets rather than recreating equivalent domestic capability. This mirrors the approach of sectors such as pharmaceuticals and aviation, where regulators routinely recognise internationally accredited testing laboratories, certification bodies and inspection organisations. Strategic dependence is not a sign of weakness. It enables governments to benefit from globally recognised expertise while focusing scarce domestic resources on capabilities that are crucial to their own national priorities. It also prevents increasing fragmentation of the AI-assurance space.

Recommendation: *Build the technical capability and human capacity needed to engage with AI assurance.*

AI assurance depends on both technical capability and human capacity. On the technical side, assurance providers need access to robust measurement methods, evaluation tools, benchmarks, testing environments, monitoring infrastructure and, where appropriate, sufficient access to AI systems and data to assess them rigorously. On the human side, they need people with the multidisciplinary expertise to design evaluations, interpret evidence and exercise judgement about whether systems are trustworthy in their intended context.

Stakeholders interviewed for this paper consistently identified constraints across both areas as barriers to AI-assurance ecosystem development. Measurement science is still maturing, technical tooling is uneven across

different types of AI systems and risks, and much of the leading technical expertise remains concentrated within frontier-AI companies and a relatively small number of specialist organisations. At the same time, shortages of skilled practitioners risk becoming a bottleneck as AI deployment accelerates.

Governments should therefore treat technical infrastructure and workforce development as complementary investments. Building a larger assurance profession will achieve little if providers lack the tools, methods and access required to generate credible evidence; equally, sophisticated evaluation infrastructure will have limited value without skilled practitioners to deploy it effectively.

Build industry technical capability to deliver AI assurance. Governments should focus on addressing barriers that prevent specialist providers from developing, validating and scaling technical-assurance services. They can fund shared or open technical resources where these would otherwise be underprovided; provide assurance providers with access to testing facilities, specialist compute and relevant data sets; use challenge programmes and testbeds to enable providers to develop and demonstrate their methodologies against real AI systems; and use public procurement to create early demand for technically rigorous assurance services. Where useful, governments can also support independent validation and comparison of assurance tools, helping buyers understand what different tools measure, how robust their methods are and where their limitations lie.

Emerging initiatives provide models to build on. The UK's AI Assurance Innovation Fund is providing £11 million to support the development of novel assurance tools and services,[49] while the US NIST ARIA programme provides common evaluation infrastructure for model testing, red-teaming and field testing, with the aim of producing reusable methodologies, metrics and tools for the wider ecosystem.[50] These approaches need not be replicated by every country. Governments without the resources or technical base to develop equivalent infrastructure domestically should seek access through international partnerships, shared facilities and established assurance markets, while concentrating domestic investment on the capabilities needed to procure and use these services effectively.

Address information-access barriers to independent assurance. Access to AI systems presents a particular structural challenge for independent assurance. For some forms of assessment, access through a public interface may be sufficient. But rigorous evaluation of highly capable or high-risk systems may require privileged access to technical documentation, system logs, evaluation and training information, higher or unrestricted querying limits or, in some cases, model internals. Without sufficient access, third-party providers risk being limited to relatively shallow assessments of externally observable behaviour, constraining the confidence that can reasonably be placed in their findings.[51]

Yet granting deeper access creates its own challenges. AI developers have legitimate reasons to restrict sensitive information relating to proprietary systems, cyber-security vulnerabilities or potentially dangerous capabilities. Providers receiving privileged access may also become bound by confidentiality requirements that constrain what findings they can communicate. More fundamentally, if developers themselves determine which organisations are sufficiently trusted to receive access, incumbent providers with established relationships are advantaged over new entrants. This risks creating a market in which the organisations being assured exercise significant control over who is capable of assuring them, weakening competition and perceptions of independence.

Governments should therefore treat secure-access arrangements as part of the institutional infrastructure for independent AI assurance, rather than leaving them entirely to bilateral relationships between developers and assurance providers. There are already emerging models to build on. Under the EU AI Act, notified bodies assessing certain high-risk AI systems can receive access to technical documentation and relevant training, validation and testing data sets and, where other means of assessing conformity are insufficient, can request access to trained models and relevant parameters.[52] Similar principles are long established in financial audit, where independent statutory auditors have legal rights to access company records and obtain information necessary to perform their assessments.[53]

Governments should explore how comparable arrangements can be adapted to AI assurance. Depending on the risk and deployment context, this could include standardised access requirements attached to regulatory or procurement obligations; common confidentiality and disclosure protocols specifying what providers can access and what findings they can communicate; secure evaluation environments in which sensitive systems can be examined without information being transferred more widely; and accreditation or competency requirements that provide a trusted gateway to privileged access. For the most sensitive systems, accredited independent providers could receive controlled access under regulatory supervision rather than solely at the discretion of the developer.

Professionalise AI assurance. Professionalisation provides one of the most effective mechanisms for building AI-assurance capability. It is not only about upskilling practitioners and maintaining consistent standards of competence, but also about establishing AI assurance as a recognised and respected profession that attracts and retains top talent.[54] Professional bodies can define competency frameworks, establish career pathways, provide continuing professional development and uphold ethical standards, helping the workforce keep pace with rapidly evolving AI technologies.[55] Organisations such as IAPP and BCS, The Chartered Institute for IT are already playing an important role in developing the professional standards and communities needed to support a mature AI-assurance ecosystem.[56]

Develop talent pipelines into AI assurance. Governments should invest in long-term talent pipelines to attract new entrants into AI assurance and equip like-minded professionals to transition into the field. This could include university programmes, postgraduate qualifications, apprenticeships, executive education and mid-career conversion programmes spanning both the technical and governance dimensions of AI assurance. Alongside developing technical evaluators, governments should also support pathways for professionals in fields such as auditing, cyber-security, risk management, law and compliance to specialise in AI assurance. Such initiatives could be modelled on the EU's AI Skills Academy and Singapore's AI Apprenticeship Programme, adapting broader AI-workforce-development efforts to the specific competencies required for AI assurance.[57,58]

Harness frontier expertise to strengthen the wider assurance ecosystem. Much of the world's leading expertise in evaluating frontier AI systems resides within a handful of frontier-AI developers. Stakeholders consistently noted that the AI-assurance ecosystem will only keep pace with frontier AI if this expertise helps define the evolving best practices that underpin independent assurance. Governments should therefore create incentives for frontier-AI developers to contribute to the development of shared public goods such as evaluation methodologies, measurement science, benchmarks and technical standards, through participation in standards bodies, collaborative research programmes, and contributions to national metrology institutes and AI standards hubs. Some industry-led initiatives already perform elements of this function. For example, the Frontier Model Forum facilitates coordination and the development of emerging best practices among frontier-AI developers, while MLCommons brings together industry and other stakeholders to develop shared evaluation methods and benchmarks. These institutions can then translate frontier technical knowledge into independent best practice that can be referenced by regulators, procurers, insurers and assurance providers across the wider ecosystem.

At the same time, it is important that standards-setting and methodology development are not dominated by the largest AI companies. The challenge is to ensure frontier expertise is channelled through governance arrangements that remain balanced, transparent and operationally independent. Governments could, for example, explore institutional models such as the US Financial Industry Regulatory Authority (FINRA), where industry collectively supports independent institutions that serve the wider public interest while operating independently of any individual firm.[59] The US government is currently exploring this approach to establishing an AI regulatory body.[60]

Facilitate the responsible use of AI for AI assurance. Governments should actively support the development and adoption of AI-enabled assurance tools that expand the scale, speed and affordability of AI assurance. This includes funding assurance sandboxes and pilot programmes, publishing guidance on where automation is appropriate, and ensuring regulatory and

procurement frameworks recognise the responsible use of automated assurance. Findings from Singapore's Global AI Assurance Sandbox suggest that AI-assisted approaches will be essential for scaling assurance and should be used to augment rather than replace human experts – automating repetitive, scalable tasks such as evidence collection, documentation, continuous monitoring and aspects of technical evaluation, while leaving humans to design evaluation methodologies, interpret results and exercise judgement. Governments can accelerate responsible adoption by using sandboxes to further contribute findings into recognised best practice.

Recommendation: *Help organisations identify and access trustworthy AI-assurance providers.*

Building capability is only valuable if organisations can readily access it. Having invested in AI-assurance capability, governments must also ensure that organisations can access it. Interviewed developers and deployers have noted the difficulty in distinguishing between the quality and claims of different assurance providers in what remains a young and rapidly evolving market. Likewise, assurance providers similarly reported worries about new customers finding them and choosing their services in an increasingly saturated field.

Governments can help make AI-assurance markets more transparent and easier to navigate. This could be done, for example, by maintaining public directories or approved provider lists, either directly or through delegated institutions such as accreditation bodies, professional associations or independent assurance organisations, making it easier for organisations to identify trusted providers. Another option is to directly facilitate collaboration between AI developers, deployers and assurance providers to test and refine assurance methodologies, as demonstrated by Singapore's Global AI Assurance Sandbox pilot programme. Such initiatives not only improve assurance practices but also help organisations identify capable providers with demonstrated expertise.

Governments should support strong systems of professionalisation, accreditation and certification. Professionalisation can establish recognised competencies and career standards, while accreditation can provide a trusted signal that assurance providers are competent to perform specified assessment activities. Certification can provide similarly useful evidence that a defined product, process or organisation meets specified criteria. In each case, the signal is only as meaningful as the scope and quality of the underlying assessment.

As outlined at the outset of this paper, AI assurance serves two complementary purposes: improving the safety and trustworthiness of AI systems and communicating that trustworthiness so others can make informed decisions. Helping organisations identify trustworthy assurance providers is therefore an essential part of a functioning AI-assurance ecosystem. Ultimately, it is the combination of trustworthy AI and trustworthy assurance that will enable AI adoption to accelerate with confidence.

07

Conclusion

AI is moving from an era defined by capability breakthroughs into one increasingly characterised by widespread deployment and adoption. The countries that benefit most from this transition will not simply be those that build the most powerful AI models. They will be those that build the processes and institutions needed to deploy AI safely, confidently and at scale.

AI-assurance ecosystems provide a practical path forward. They enable governments to establish governance that is robust enough to maintain accountability, yet flexible enough to evolve alongside the technology. Done well, they improve the quality and safety of AI systems, build public and market confidence, reduce compliance burdens, create new economic opportunities and, above all, accelerate the adoption and scaling of AI across economies.

The key is for political leaders to get started. The temptation will be to wait until all the building blocks are set, but AI-assurance ecosystems cannot be designed entirely in the abstract. Assurance providers need demand before they invest. Buyers need trusted providers before they procure. Insurers need evidence before they can price risk. Standards improve through practical application, while regulation depends on capable assurance markets to enforce it. Every part of the ecosystem depends on the others, making it tempting to wait until every piece is in place before acting. That would be a mistake.

One of the defining strengths of AI assurance is that it is designed to evolve. Governments do not need to build the perfect assurance ecosystem before taking the first step. They need to create the conditions for one to emerge. Early procurement requirements, targeted regulation, pilot programmes, sandboxes and initial standards will all be imperfect. They should be. They create the feedback loops through which markets mature, methodologies improve and institutions strengthen. The greatest risk is not getting the first step wrong. It is waiting until every uncertainty has been resolved before taking one at all.

The countries that lead the age of AI will not necessarily be those that build the most capable models. They will be those that build the trust, institutions and governance needed to deploy AI confidently at scale. The time to start building those foundations is now.

08

Annex

Summary: A Roadmap for Building an AI-Assurance Ecosystem

Governments do not need to implement every intervention at once. The appropriate pathway will depend on existing institutions, domestic capabilities and patterns of AI adoption. But governments can start with the policy levers already available to them, progressively build the foundations and capabilities of a mature assurance market, and refine the system as evidence accumulates.

Step 1: Start with the levers government already controls. Make AI assurance a strategic national priority and establish clear expectations for trustworthy AI.

Recommendations

- **Set risk-proportionate assurance requirements.** Require appropriate testing, evaluation and ongoing monitoring for highly capable models and AI deployed in high-risk contexts; use outcome-based requirements and independent third-party assessment where risks justify it.
- **Use outcome-based regulation.** Set stable legal duties and the outcomes that must be demonstrated, while allowing regulators, standards bodies and technical experts to determine appropriate methodologies and evidence as technology evolves. For high-stakes applications, consider structured assurance cases through which organisations demonstrate that required claims have been satisfied.
- **Build on existing sectoral regimes.** Map existing assurance requirements before introducing new AI-specific obligations and embed AI assurance within established regulatory, conformity-assessment and supervisory processes where possible.

- **Use public procurement.** Specify the assurance evidence, governance, monitoring and maintenance practices expected from government AI suppliers; introduce requirements early and update them as methodologies mature.

2. Put the legal and market foundations in place. Make meaningful assurance part of the normal operation of AI markets, rather than something driven solely by government requirements.

Recommendations

- **Clarify liability across the AI lifecycle.** Allocate responsibility according to actors' information, control and ability to mitigate risk; clarify the respective responsibilities of developers and deployers and begin addressing liability for automated assurance tools.
- **Build towards viable AI-insurance markets.** Build the evidence and underwriting capability needed to price AI risk through incident reporting and information sharing; ensure existing policies explicitly address AI-related harms; and progressively introduce insurance requirements where risks become sufficiently measurable and insurable.

3. Secure the capabilities needed to deliver assurance. Ensure organisations can access the technical infrastructure, expertise and skilled people required for high-quality assurance.

Recommendations

- **Choose what to build, partner on or depend on.** Map national strengths and gaps; invest domestically where capability has strategic value; pool resources internationally for shared infrastructure; and use mature international markets where domestic duplication adds little value.
- **Build industry technical capability.** Support the translation of measurement science into usable tools and services; provide shared testing resources, compute and data sets where needed; use testbeds and challenge programmes; and support independent validation of assurance tools.

- **Enable meaningful independent access to AI models, systems and documentation.** Develop secure mechanisms through which qualified providers can access the information and systems required for rigorous assessment, including common confidentiality protocols, secure evaluation environments and appropriate accreditation or competency gateways.

4. Build a credible and scalable assurance profession and market. Create the institutions and workforce that allow assurance to be delivered consistently, independently and at scale.

Recommendations

- **Professionalise AI assurance.** Support competency frameworks, ethical standards, continuing professional development and recognised career pathways.
- **Build talent pipelines.** Develop university, postgraduate, apprenticeship and mid-career routes into assurance, including pathways for auditors, cyber-security specialists, risk professionals, lawyers and compliance practitioners.
- **Make provider competence legible.** Strengthen accreditation and appropriate certification, and establish directories or approved-provider lists to help organisations identify providers competent to perform particular assurance activities.
- **Harness frontier expertise while protecting independence.** Create mechanisms for frontier developers to contribute to shared methodologies, benchmarks, measurement science and standards while ensuring these processes are not dominated by individual firms.

5. Scale, connect and continuously improve the ecosystem. Use emerging technology, international cooperation and practical experience to make assurance cheaper, more interoperable and more effective over time.

Recommendations

- **Use AI to scale assurance.** Fund pilots and sandboxes for AI-enabled assurance; develop guidance on appropriate automation; and feed findings into recognised best practice.
- **Build international interoperability.** Participate actively in international standards development, support shared measurement and evaluation infrastructure, and pursue mutual-recognition mechanisms that reduce unnecessary duplication across markets.
- **Learn through implementation.** Use procurement, regulation, pilots, sandboxes, incident data and monitoring to test assurance requirements in real-world deployment, and refine them as evidence and methodologies improve.

09

Contributors and Acknowledgements

This paper has been developed through broad stakeholder consultation across industry, academia and civil society. Note: contribution does not equal endorsement of points made in the paper.

Contributors

The following have provided substantive feedback and put forward inclusions to this report.

Maria Axente, Responsible Intelligence

David Bholat, Faculty AI

Tom Clarke, Faculty AI

Callum Cockburn, Synoptix

Phil Dawson, Armilla AI

Tom Fowler, Kainos

Kasia Jakimowicz, GovAI

Tim McGarr, BSI Group

Mishka Nemes, Trilateral Research

Adam Leon Smith, AIQI Consortium

Naomi Solomon, GovAI

David Sully, Advai

Alexandru Voica, Synthesia

Peter Wedge, Testudo

Pingshan Zhang, GovAI

Acknowledgements

The authors would further like to thank the following experts for their input and expertise.

George Balston, AVERI

Peter Cihon, AI standards and policy expert

Gwendoline Grollier, T3

Lara Groves, Ada Lovelace Institute

Annika Hallensleben, Apollo Research

Emil Bender Lassen, AIUC

Emma McGuigan, AI Assurance Stakeholder Consortium

Lee Wan Sie, AI Verify Foundation

Adam Pettman, 2i

Endnotes

1 https://partnershiponai.org/resource/strengthening-the-ai-assurance-ecosystem/

2 https://www.aisi.gov.uk/blog/early-lessons-from-evaluating-frontier-ai-systems

3 See recent work by the Partnership on AI for additional useful ecosystem-mapping references. https://partnershiponai.org/resource/strengthening-the-ai-assurance-ecosystem/

4 The standard specifies requirements and provides guidance for establishing, implementing, maintaining and continually improving an AI management system within the context of an organisation. https://www.iso.org/standard/42001; https://knowledge.bsigroup.com/products/information-technology-artificial-intelligence-management-system-1

5 https://www.aiuc-1.com/

6 https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10

7 https://aisecurityandsafety.org/en/frameworks/singapore-ai-verify/

8 https://aief.dev/

9 https://www.fsis.usda.gov/news-events/news-press-releases/reflecting-25-years-haccp

10 https://www.fda.gov/about-fda/fda-history/milestones-us-food-and-drug-law

11 https://www.iata.org/en/programs/safety/

12 https://www.aptiv.com/en/insights/article/what-is-ncap

13 https://institute.global/insights/tech-and-digitalisation/what-the-uk-thinks-about-ai-building-public-trust-to-accelerate-adoption

14 https://www.qlik.com/us/news/company/press-room/press-releases/61-percent-of-global-businesses-are-scaling-back-ai-investment-as-a-result-of-trust-issues

15 https://www.faa.gov/newsroom/faa-updates-boeing-737-max-0

16 https://www.iata.org/en/iata-repository/publications/economic-reports/air-passenger-monthly---dec-2019/

17 https://www.hfea.gov.uk/celebrating-40-years-of-ivf/

18 https://www.hfea.gov.uk/about-us/news-and-press-releases/2026/number-of-ivf-patients-triples-in-30-years-says-fertility-regulator/

19 https://www.ifrs.org/use-around-the-world/why-global-accounting-standards/

20 https://assets.publishing.service.gov.uk/media/68f0b97c1c9076042263ef2a/MAA%5FRN%5F2025%5F04.pdf

21 These include, for example, the UK AI Security Institute (UK AISI), the US Center for AI Standards and Innovation (CAISI), the Singapore AI Safety Institute (Singapore AISI), the Canadian Artificial Intelligence Safety Institute (CAISI), the Japan AI Safety Institute (J-AISI), and the French National Institute for AI Evaluation and Security (INESIA), and others, many of which are now organised into an International Network for Advanced AI Measurement, Evaluation and Science (NAAIMES); https://www.aisi.gov.uk/blog/international-evaluation-best-practice-and-open-questions-in-ai-measurement

22 https://www.itf-oecd.org/sites/default/files/docs/zero-road-deaths.pdf

23 https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/09/meat-protein-alternatives%5F54e42940/387d30cf-en.pdf

24 https://www.nist.gov/itl/ai-risk-management-framework

25 https://www.npl.co.uk/news/npl-to-establish-new-centre-for-ai-measurement

26 https://www.iec.ch/ai

27 https://standards.ieee.org/beyond-standards/artificial-intelligence-ai-and-cybersecurity-emerging-risks-big-opportunities-and-the-path-to-trust/

28 https://aiforgood.itu.int/ai-standards-exchange/

29 https://www.iaps.ai/research/assuring-growth

30 https://ivo.fathom.org/

31 https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB813#90ENR

32 https://www.york.ac.uk/assuring-autonomy/guidance/big-argument-ai-safety-cases/

33 https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device/software-and-artificial-intelligence-ai-as-a-medical-device

34 https://www.gov.uk/government/collections/ai-airlock-the-regulatory-sandbox-for-aiamd

35 https://standards.ieee.org/ieee/3119/10729/

36 Office of Management and Budget, ***M-25-22: Driving Efficient Acquisition of Artificial Intelligence in Government***, 3 April 2025. Requirements apply to contracts under solicitations issued on or after 30 September 2025.

37 https://files.gao.gov/reports/GAO-26-107859/index.html

38 https://www.hyperdimensional.co/p/how-should-ai-liability-work-part-3df

39 For example, in 2026, a data breach at the AI recruiting firm Mercor became the subject of class-action litigation that named not only Mercor but also LiteLLM, the software it relied on, and Delve Technologies, the automated-compliance provider that had certified that software's security. The complaint alleges that Delve issued fabricated compliance certifications that Mercor relied upon in its vendor due diligence, extending potential liability along the chain from the organisation that suffered the breach to the tool it deployed and the provider that assured it. This case is still in process. ***White and Beltran versus Mercor.io Corporation, Delve Technologies, Inc. and Berrie AI Incorporated***, US District Court for the Northern District of Texas, Case No. 6:26-CV-00143-H (2026). https://news.bloomberglaw.com/litigation/ai-talent-recruiter-mercor-hit-with-suit-over-march-data-breach

40 https://autopoesis.substack.com/p/machina-economica-part-ii-the-commodification

41 https://arxiv.org/abs/2607.11999

42 https://papers.ssrn.com/sol3/papers.cfm?abstract%5Fid=5505759

43 https://arxiv.org/abs/2512.06597

44 NAAIMES, formerly referred to as the International Network of AI Safety Institutes, brings together government AI institutes and counterpart bodies from ten jurisdictions to develop internationally recognised approaches and best practices for measuring and evaluating advanced AI capabilities. The network is currently comprised of organisations from ten participating jurisdictions: Australia, Canada, the EU, France, Japan, Kenya, South Korea, Singapore, the UK and the US; https://www.aisi.gov.uk/blog/international-evaluation-best-practice-and-open-questions-in-ai-measurement

45 https://appiafoundation.org/wp-content/uploads/sites/6/2026/06/appia_foundation_execsum061626a.pdf

46 https://aiqi.squarespace.com/about-us

47 https://global-aci.org/en/home/

48 https://iaf.nu/en/news/global-accreditation-cooperation-incorporated-launch-unifies-international-accreditation-organisations-and-strengthens-worldwide-trust/

49 https://www.gov.uk/government/publications/trusted-third-party-ai-assurance-roadmap/

trusted-third-party-ai-assurance-roadmap

50 https://ai-challenges.nist.gov/aria

51 https://www.aisi.gov.uk/blog/early-lessons-from-evaluating-frontier-ai-systems

52 https://eur-lex.europa.eu/eli/reg/2024/1689/2026-07-27/eng

53 For example, see Section 499 of the UK Companies Act 2006: https://www.legislation.gov.uk/ukpga/2006/46/pdfs/ukpga%5F20060046%5Fen%5F002.pdf

54 https://www.adalovelaceinstitute.org/report/going-pro/

55 https://www.gov.uk/government/publications/trusted-third-party-ai-assurance-roadmap

56 https://www.techuk.org/resource/techuk-paper-mapping-the-responsible-ai-profession-a-field-in-formation.html

57 https://digital-strategy.ec.europa.eu/en/policies/ai-talent-skills-and-literacy

58 https://aiap.sg/apprenticeship/

59 https://www.finra.org/index.php/about

60 https://www.cfr.org/articles/the-u-s-is-about-to-design-an-ai-regulator-heres-how-to-get-it-right

TONY BLAIR
INSTITUTE FOR
GLOBAL CHANGE

Follow us

facebook.com/instituteglobal

x.com/instituteGC

instagram.com/institutegc

General enquiries

info@institute.global