Discussion paper

Going pro?

Considerations for the emerging field of AI assurance

Emerging Technology & Industry Practice

Lara Groves
Ada Lovelace Institute

Amy Winecoff, Miranda Bogen
Center for Democracy and Technology

July 2025

Ada Lovelace Institute

CENTER FOR DEMOCRACY & TECHNOLOGY

Contents

How to read this paper

If you are a **policymaker or regulator**, read the 'Executive summary', which includes seven policy recommendations setting out options for professionalising AI assurance. You may want to read the 'Research findings' section exploring the regulatory and market drivers for professionalisation.

If you are an **AI assurance professional**, you might be interested in the findings that explore training and certification options.

If you are a **researcher** interested in the ongoing implementation of AI auditing, assurance and accountability practices, read the 'Research findings' section, and the 'Conclusion and further questions'.

Executive summary

AI systems may pose significant risks and impacts across the lifecycle of development and deployment, which will require a suite of methods to assure they are safe, fair and effective. Policy, civil society and industry have become increasingly interested in AI assurance – a set of practices that measure, evaluate and communicate the trustworthiness of AI systems[1] – as a promising way to support robust oversight and innovation in AI.[2][3][4] Practices that might fall under the banner of assurance include AI or algorithm audits, red teaming, conformity assessments and impact assessments (see 'Glossary').

Assurance practices have a long history of effective adoption in safety critical industries, like aviation and pharmaceuticals, and could likewise have an important role to play in supporting safe innovation in AI. Previous research by the Ada Lovelace Institute has found that people want AI systems in critical contexts like healthcare to be governed according to high standards of accountability and transparency,[5] and existing research on AI assurance suggests assurance may have a role in helping companies demonstrate trustworthiness to people and consumers.[6][7]

While assurance activities related to AI are already occurring, the landscape remains fragmented and efforts largely ad hoc. Proponents of AI assurance argue that professionalising the industry could help

1 UK Government, 'Introduction to AI Assurance' (GOV.UK) https://www.gov.uk/government/publications/introduction-to-ai-assurance/introduction-to-ai-assurance accessed 5 June 2025.

2 Digital Regulation Cooperation Forum, 'Ensuring Trustworthy AI: The Emerging AI Assurance Market' (www.drcf.org.uk, 16 July 2024) https://www.drcf.org.uk/publications/blogs/ensuring-trustworthy-ai-the-emerging-ai-assurance-market accessed 5 June 2025.

3 'BABL AI: Conducting Third-Party Audits for Automated Employment Decision Tools' (GOV.UK) https://www.gov.uk/ai-assurance-techniques/babl-ai-conducting-third-party-audits-for-automated-employment-decision-tools accessed 7 June 2025.

4 Abeba Birhane and others, 'AI Auditing: The Broken Bus on the Road to AI Accountability' (arXiv, 25 January 2024) http://arxiv.org/abs/2401.14462 accessed 31 January 2025. See also: Rosamund Powell and Marion Oswald, 'Assurance of Third-Party AI Systems for UK National Security' https://cetas.turing.ac.uk/publications/assurance-third-party-ai-systems-uk-national-security accessed 5 June 2025.

5 Ada Lovelace Institute and Alan Turing Institute, 'How Do People Feel about AI?' https://attitudestoai.uk/ accessed 20 June 2025.

6 Inioluwa Deborah Raji and others, 'Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance' (arXiv, 9 June 2022) http://arxiv.org/abs/2206.04737 accessed 4 August 2024.

7 Jakob Mokander and Luciano Floridi, 'Ethics-Based Auditing to Develop Trustworthy AI' (2021) 31 Minds and Machines 323 http://arxiv.org/abs/2105.00002 accessed 5 June 2025.

increase the effectiveness of the AI assurance industry in advancing sound practices for AI development and adoption.[8]

'Professionalising' an industry refers to the process of a field or occupation taking on 'professional qualities', which usually include training or certification, but can also include the creation of codes of conduct and membership bodies, and standardised practices. Many established professions including medicine[9] and law[10] have historically required regular assessments and certifications of professionals' competence and quality.

A professionalised industry may demonstrate trustworthiness and reliability in providing high-quality services or products. It may also increase the overall capability and productivity of the industry over time.[11]

At the time of writing, the global political economy has shifted towards a strategy of deregulation. In the UK, the Prime Minister has urged national regulators to take on a commitment to the 'growth agenda',[12] while a proposed national AI bill is on uncertain footing.[13] In the US, congressional leaders attempted, unsuccessfully, to introduce a decade-long moratorium on all state-level policymaking on AI.[14] The Trump administration has also directed the federal government to review and rescind policies and regulations that might constrain American competitiveness in global markets.

This has implications for AI assurance which, to date, has been incentivised to some extent through regulation. This includes US state-level bills like New York City's Local Law 144[15] and the European Union's Artificial

8 Inioluwa Deborah Raji and others, 'Outsider Oversight' (n 6).

9 Cathy Peck and others, 'Continuing Medical Education and Continuing Professional Development: International Comparisons' (2000) 320 BMJ 432 https://www.bmj.com/content/320/7232/432 accessed 5 June 2025.

10 'Family Law Accreditation: Re-Accreditation' https://www.lawsociety.org.uk/career-advice/individual-accreditations/family-law-accreditation/re-accreditation accessed 20 June 2025.

11 'Professional Bodies Add Huge Value to Society | CIOB' https://www.ciob.org/news/professional-bodies-add-huge-value-to-society accessed 5 June 2025.

12 'New Approach to Ensure Regulators and Regulation Support Growth' (GOV.UK) https://www.gov.uk/government/publications/a-new-approach-to-ensure-regulators-and-regulation-support-growth/new-approach-to-ensure-regulators-and-regulation-support-growth-html accessed 5 June 2025.

13 'Britain Goes Soft on AI after Trump's Bonfire of Rules' (*POLITICO*, 7 February 2025) https://www.politico.eu/article/britain-led-the-world-ai-safety-now-waiting-donald-trump/ accessed 5 June 2025.

14 Benj Edwards, 'GOP Sneaks Decade-Long AI Regulation Ban into Spending Bill' (*Ars Technica*, 13 May 2025) https://arstechnica.com/ai/2025/05/gop-sneaks-decade-long-ai-regulation-ban-into-spending-bill/ accessed 20 June 2025.

15 Ada Lovelace Institute, 'Code & Conduct' https://www.adalovelaceinstitute.org/report/code-conduct-ai/ accessed 5 June 2025.

Intelligence Act. However, previous research and policy development in AI assurance has identified that market forces may also be a strong incentive to shore up assurance practices and grow the industry.[16]

Previous research and policy has also proposed that professionalising the industry would enhance the positive impact of these market players by instilling confidence among actors like consumers and regulators that assurance practices are being conducted in accordance with accepted standards.

It is critical that policymakers, developers and deployers of AI make use of available mechanisms and practices to ensure the safety of tools and systems, and to minimise undue risks. Active professionalisation of AI assurance could support the development and adoption of safe, reliable and effective AI technologies that deliver benefits for people and society.

Collaboration between policymakers, regulators, industry and coalitions like standards development bodies will be required to take forward some options for professionalisation. This report outlines key considerations to inform efforts towards this goal.

Introducing our research

Existing research and policy has explored options for certifying AI assurance professionals.[17] [18] But there is limited qualitative evidence from professionals working on the ground about the state of play and how their experiences might inform efforts to professionalise AI assurance. It is crucial that efforts to professionalise the industry incorporate this evidence and take a global view to ensure robust and meaningful impact.

The Ada Lovelace Institute and the Center for Democracy & Technology (CDT) have collaborated to conduct qualitative research exploring the potential impacts of, and conditions needed for, professionalising the AI assurance industry.

16 'Unlocking the Growth Potential of the UK's AI Assurance Market' (*Frontier Economics*) https://www.frontier-economics.com/uk/en/news-and-insights/news/news-article-i21001-unlocking-the-growth-potential-of-the-uk-s-ai-assurance-market/ accessed 5 June 2025.

17 Inioluwa Deborah Raji and others, 'Outsider Oversight' (n 6).

18 UK Government, 'Introduction to AI Assurance' (n 1).

To build the evidence base for experiences of AI assurance professionals, we conducted interviews with 15 professionals and experts with knowledge and experience of:

- third- and second-party AI assurance or auditing
- third- and second-party model evaluations
- technical standards bodies
- AI assurance or AI governance training or certification.

We explored the following three research questions:

1. What is AI assurance and what is it setting out to achieve?
2. What role can a third-party professionalised industry play in ensuring AI assurance?
3. What is needed to ensure assurance works well?

For more information, see the 'Methodology' section.

Findings

Our research surfaced several important findings relevant to policymakers, AI assurance professionals and providers, and standards development bodies.

- **AI assurance must coalesce around a defined scope, required competencies, and core practices to become a professionalised field.** Our interviewees broadly agreed that the purpose of AI assurance is to evaluate AI systems and the organisations developing and deploying them, in order to validate claims about performance and risk. They agreed on three areas of expertise that assurance professionals should demonstrate: technical knowledge, legal fluency and risk management experience. However, they disagreed on how these competencies should be operationalised and applied to assurance practices, and to what degree.

- **Standards will have a direct impact on the scope of AI assurance, both enabling and constraining assurance activities.** On the one hand, there was support from interviewees for technical standards or risk management frameworks to provide consistency and clarity to assurance providers on the scope of their activities. On the other hand, some practitioners felt that in practice, standards often constrain or limit the overall process of assurance as well as its outcomes.

- **Accreditation or certification of AI assurance professionals is not a silver bullet that will solve all challenges in AI assurance adoption.** Our participants echoed existing research and policy that calls for accreditation and certification of AI assurance professionals and teams. However, we found disagreement about who is best placed to offer accreditation and certification. Interviewees also spoke of the risk that certification may not standardise and symmetrise practices to adequately raise the effectiveness of AI assurance.

- **Regulatory and market forces are likely to be the primary drivers of professionalisation.**
  Participants put forward regulation as a significant driver for the professionalisation of AI assurance. They also identified that market forces may create economic incentives for professionalisation and for wider AI assurance adoption, in light of global policymakers' deregulatory strategies.

Recommendations

Recommendation 1: Assurance practitioners and the organisations supporting them should clarify which competencies are relevant for assurance across AI systems generally, and which are relevant for specialised contexts.

Some competencies for AI assurance – like proficiency navigating risk management processes – are applicable to assurance practices in general, while other competencies will require specialised knowledge and experience. For example, a practitioner working in AI assurance in finance will need different competencies to one working in social media.

Recommendation 2: Certification of assurance professionals should consist of modular 'tracks'.

A flexible, modular approach to certification will allow professionals to demonstrate relevant expertise and create career development pathways, supporting various specialisms.

Recommendation 3: Policymakers should take action to promote skills development that could support AI assurance.
National skills agendas or educational programmes should drive skills development, which in turn could support professional development in AI assurance.

Recommendation 4: Standards setting bodies should create broadly applicable standards as well as tailored standards for specific system types and domain-specific use cases.

Some assurance methods are not widely applicable to all kinds of AI systems. Therefore, it is challenging to provide a holistic assurance assessment of an AI system based on a single standard.

Generative AI systems in particular will require unique assurance mechanisms since they have a variety of different downstream applications. Moreover, AI systems deployed within specific sectors and contexts may require forms of assurance that are not relevant in other sectors. Standards will need to reflect best practices that apply across system types and sectors as well as those applicable to more narrow contexts.

Recommendation 5: Professionalisation of AI assurance should be oriented towards supporting assurance adoption across the ecosystem, including downstream deployers of AI.

Since procurement can be a meaningful lever for the adoption of assurance practices,[19] policymakers should consider proposals related to professionalisation that encourage deployers to rely on assurance.

Recommendation 6: Standards, certification and training for AI assurance should demonstrate how assurance supports both clients' business priorities and accountability goals.

In the absence of regulatory pressure, tactics for adopting assurance across the ecosystem should focus on clear communication around how assurance services support business needs like market adoption and revenue. This should align with wider accountability and governance ambitions of AI assurance.

19 Ada Lovelace Institute, 'Buying AI' https://www.adalovelaceinstitute.org/report/buying-ai-procurement/ accessed 5 June 2025.

Recommendation 7: The AI assurance industry should cultivate strong professional norms and an ethical culture to complement formal standards and accountability structures.

The AI assurance industry should work collaboratively to define shared norms and practices and a strong professional culture. These normative commitments may support organisational buy-in and reinforce best practices, advancing professionalisation and increasing the degree of adoption of AI assurance.

A note on terminology

The authors of this report use several nascent terms. We have compiled a glossary, based on emerging best practice and literature for AI assurance. The definitions provided and used throughout the report do not always fully align with the definitions shared by interviewees, which we present in their original context.

Glossary

(AI) assurance

We follow the UK's Department for Science, Innovation and Technology's definition of AI assurance:

- **Assurance** is the process of measuring, evaluating and communicating something about a system or process, documentation, a product or an organisation. In the case of AI, assurance measures, evaluates and communicates the trustworthiness of AI systems.[20]

Understood this way, AI assurance is a process or a service exercised towards products or organisations.

There are a number of different practices that fall under the banner of AI assurance. These include:

- Impact assessments, which explore how a particular AI system will affect people or society in positive or negative ways before the system is deployed.[21]
- Formal verification, a technique often used to assess software and hardware. Formal verification establishes whether a system satisfies specific requirements, often using formal mathematical methods and proofs.[22]
- Red teaming, which is an activity that involves the probing of a system in an adversarial way, to identify potential harmful outputs.[23]

The UK's AI Opportunities Action Plan refers to 'assurance tools', which reflects increasing interest in the role of automated, or sometimes AI-driven, tools or products that provide assurance.[24] One example of an

20 UK Government, 'Introduction to AI Assurance' (n 1).
21 Ada Lovelace Institute, 'AI Assurance?' https://www.adalovelaceinstitute.org/report/risks-ai-systems/ accessed 11 October 2024.
22 UK Government, 'Introduction to AI Assurance' (n 1).
23 Miranda Bogen, 'Assessing AI: Surveying the Spectrum of Approaches to Understanding and Auditing AI Systems' https://cdt.org/insights/assessing-ai-surveying-the-spectrum-of-approaches-to-understanding-and-auditing-ai-systems/.
24 UK Government, 'AI Opportunities Action Plan' (GOV.UK, 13 January 2025) https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan accessed 4 March 2025.

assurance tool is the UK government's AI Management Essentials (AIME) self-assessment tool. Another example is an AI-powered 'compliance dashboard', which gathers real-time data and metrics from companies to automatically produce a report on compliance with key regulations (such as the EU AI Act).

In this report, when we refer to 'assurance', we are referring to a process or service (or set of processes or services), unless otherwise stated.

(AI) audit

An AI or algorithmic audit is a process for scrutinising an AI system, or the policies and processes around it.[25] An audit can be considered a type of assurance practice, but the term is often used synonymously with assurance – to describe a process of measurement, evaluation and communication. Audits, depending on their specific design and implementation, can encompass different types of practices and can assess different types of risk.

For example, a technical audit is an audit of an AI system's inputs and outputs, measuring for accuracy or bias, while a compliance audit might be used to understand if a team or organisation has completed certain processes or regulatory requirements.

Professionalisation

'Professionalisation' refers to the process of giving professional qualities to a group or occupation, usually through training or certifications. Other components can include the creation of codes of conduct and membership bodies, standardised practices, and regular assessments of competence and quality. Some industries designate legally protected titles that demonstrate that a professional is trained or qualified to a particular standard, such as chartered surveyors or accountants.

25 Ada Lovelace Institute, 'AI Assurance?' (n 21)

First-, second- and third-party

First-, second- and third-party practices refer to the degree of independence between teams conducting the assurance process (such as audit):

- **First-party** refers to a company assessing its own products or practices (also referred to as internal assurance).
- **Second-party** describes a situation where assurance providers, such as an auditor, have a contractual relationship with the client (the organisation who is being subject to assurance), such as an AI company. This would be an example of external assurance.
- **Third-party** describes a situation where assurance is conducted by reviewers who have no contractual relationship with the auditee, which is also an example of external assurance, but one that provides a higher degree of independence.[26] [27] In other domains, third-party assurance providers might be appointed to audit a company by an institution like a government.[28]

In practice, the distinction between second- and third-party assurance may not be clear cut. Second- and third-party audits are sometimes both referred to together as an 'independent audit'.

Certification

Certification refers to a recognition of the competence of professionals that are employed by organisations and self-employed professionals within a particular industry. The process of becoming certified usually involves gaining special qualifications via training programmes and examinations.

26 Miranda Bogen, 'Assessing AI: Surveying the Spectrum of Approaches to Understanding and Auditing AI Systems' (n 23).

27 Inioluwa Deborah Raji and others, 'Outsider Oversight' (n 6).

28 Lesley K McAllister, 'Regulation by Third-Party Verification' (2012) 53 BCL Rev. 1 https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/bclr53&section=4&casa_token=ph5kGQx_4AUAAAAA:i8lgepByOzKDUkrxM6YaiDgGYiGY-w8LQunCexbLeyYIRFrCPPH9UZ0GutwfAxVwyr3lW9oS accessed 29 December 2024.

Accreditation

Historically, 'accreditation' refers to an assessment of the capability and competence of an organisation conducting assurance and whether they are in compliance with national and international standards.[29] Accreditation is generally conducted by an official body, such as a national accreditation body (in the UK, this is the United Kingdom Accreditation Service (UKAS)).

Standards

A standard is a document, developed through consensus and approved by an established body, that sets out rules, guidelines or specifications for activities and systems or their outcomes, with the goal of promoting consistency and achieving a desired degree of order within a particular context. Standards can be created by many kinds of organisations, which may vary widely in their level of formality, structure, focus and methods for standard development.[30] There are different types of standards:

- **Technical standards** are specifications that define precise requirements for a product, system or component, which can support quality and interoperability. Technical standards are distinct from safety standards.
- **Safety standards** establish requirements for products and systems that are intended to protect users, property or the environment from harms. They can involve setting thresholds for performance or risk, or detail necessary mitigations or designs that can reduce the likelihood or magnitude of negative impacts.[31]
- **Process-oriented standards** detail how work should be carried out rather than focusing on the system or product itself.

All three forms of standards play an important role in professionalisation.

29 UKAS, 'Accreditation vs Certification: What's the Difference?' (2022) https://www.ukas.com/accreditation/about/accreditation-vs-certification/ accessed 17 November 2024.

30 National Institute of Standards and Technology (US), 'A Plan for Global Engagement on AI Standards' (National Institute of Standards and Technology 2024) https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-5.pdf accessed 7 July 2025.

31 UK Government 'Technical Standards and Standard Development Organisations ' https://www.gov.uk/guidance/technical-standards-and-standard-development-organisations accessed 7 July 2025.

Introduction

Like the medicines on our shelves and the aircrafts we travel in, people expect AI technologies to be safe and effective.[32] Industries like pharmaceuticals and aviation have demonstrated and advanced trustworthiness through the use of robust independent safety testing and assessment regimes.[33]

These industries have consistently adopted mechanisms like auditing, assessment and external validation to evaluate systems and products, and to demonstrate that minimum industry or regulatory criteria for safety have been met. This provides strong assurances to people and society that products are safe and reliable.

AI systems are no different: ensuring AI systems are safe, effective and reliable will likely require regular assurance assessments of their technical components and the governance practices of companies that are developing and deploying them.

Like other safety critical industries such as pharmaceuticals, risks arising from AI systems can be dynamic, emerge over time and take on a domain-specificity when integrated in critical environments like healthcare settings. A whole lifecycle approach to assessment helps to provide continual assurance – from development to deployment.

Introducing AI assurance

The UK's Department for Science, Innovation and Technology (DSIT) defines assurance as the process of measuring, evaluating and communicating the trustworthiness of AI systems.[34] Assurance can involve a variety of methods that are already in active use in AI assessment, including auditing, external validation, and the red teaming of AI systems. (Indeed, the term 'assurance' is also sometimes used synonymously with 'auditing'.)

32 Ada Lovelace Institute and Alan Turing Institute 'How Do People Feel about AI?' (n 5).

33 Ada Lovelace Institute, 'Safe before Sale' https://www.adalovelaceinstitute.org/report/safe-before-sale/ accessed 17 May 2024.

34 UK Government, 'Introduction to AI Assurance' (n 1).

AI assurance sits alongside frameworks, policies and regulation

AI assurance represents a set of practices falling under the wider banner of AI governance, which might include assurance practices alongside frameworks, policies and regulation.

AI assurance can support multiple transparency and accountability goals for AI, including building consumer trust that AI products will function as intended, helping businesses feel confident in purchasing and adopting AI technologies, and providing regulators with the information they need to assist with monitoring compliance.[35]

Assurance in other domains, like accounting, has historically implied a degree of independence between the organisation conducting the assessment and the system or organisation being assessed.[36] In AI, however, assurance has so far been defined to encompass both internally and externally driven practices.

Internal teams often have deep knowledge of their organisation and the systems it builds. As such, they are well-positioned to thoroughly assess the functioning of systems and system components. They can identify risks that may have emerged from development choices, and determine whether these risks exceed the threshold of safety guarantees set by their organisation.

External teams do not have the same level of knowledge of systems as internal teams. However, they do have a higher degree of independence to support independent verification of an internal team's assessment of an AI system. They can also bring an alternative perspective about the necessary assessment activities or the interpretation of their results.[37]

What is AI assurance designed to achieve?

As AI systems are increasingly integrated into high-risk applications in contexts like healthcare, finance and critical infrastructure, robust AI assurance activities can help ensure AI systems function as intended and do not pose excess risk. Governments, businesses and consumers have shown that establishing the reliability and quality of AI systems,

35 Ada Lovelace Institute, 'Code & Conduct' (n 15)

36 Miranda Bogen, 'Assessing AI: Surveying the Spectrum of Approaches to Understanding and Auditing AI Systems' (n 23)

37 UK Government 'Introduction to AI Assurance' (n 1).

and the credibility of the organisations developing and deploying them, matter for trustworthiness and ultimately for adoption.[38] [39] [40]

> Evidence suggests that independently conducted assurance can be an effective accountability practice, due to the reduced risk of conflicts of interest.

This leads to higher quality assessments and stronger demonstrations of trustworthiness in products and services.[41]

In such cases, assurance can support the creation of accountability relationships between developers and deployers of technologies, and the people impacted by their technologies.[42]

Although it has shown some promise, the AI assurance industry remains highly emergent, with no standardised practices. There is also little consensus on the responsibilities of assurance professionals or the organisations employing them.[43]

Companies and deployers of AI have limited evidence to review the quality or efficacy of assurance providers, without professional norms and common frameworks for assessment. This leaves companies to make judgement calls with little guidance or accountability. Insufficient or poorly applied AI assurance practices leave people and society at risk of harm.

38 Ada Lovelace Institute and Alan Turing Institute 'How Do People Feel About AI?' (n 5).

39 Ada Lovelace Institute, 'AI Assurance? (n 21)

40 'How AI Assurance Can Support Trustworthy AI in Recruitment – Responsible Technology Adoption Unit Blog' (25 March 2024) https://rtau.blog.gov.uk/2024/03/25/how-ai-assurance-can-support-trustworthy-ai-in-recruitment/ accessed 5 June 2025.

41 Inioluwa Deborah Raji and others, 'Outsider Oversight' (n 6).

42 Trehu, Julia and Goodman, Ellen P., 'ALGORITHMIC AUDITING: CHASING AI ACCOUNTABILITY' (2023) 39 *Santa Clara High Technology Law Journal* 289 https://digitalcommons.law.scu.edu/chtlj/vol39/iss3/1 accessed 5 June 2025

43 IAPP 'AI Governance Profession Report 2025' https://iapp.org/resources/article/ai-governance-profession-report/ accessed 5 June 2025.

From a consumer perspective, professionalisation may help companies demonstrate trustworthiness

Professionalising the AI assurance industry will build trust, reduce inconsistency and uncertainty, and ensure that assurance services are both credible and effective.[44] [45]

What is a professional industry?

To assert authority within an industry, an occupation must ground its expertise in domain-specific knowledge, align its skills and activities with formalised standards, and earn public confidence by showing that its services are trustworthy.[46]

'Professionalising' an industry refers to the process of giving a group these 'professional' qualities, like developing a full-time workforce, creating professional associations, specialised education and training pathways, adopting formal codes of conduct, and establishing legal or institutional safeguards such as certification or licensure.[47]

A professionalised industry may provide higher quality services or products, via increased consistency and reliability, and increased knowledge sharing and emergent specialisms. This can result in career and industry-level opportunities for professionals.[48] Additionally, in a business context, from the perspective of a consumer, professionalisation may help companies demonstrate trustworthiness.

Professionalising an industry is rarely a simple process. Most occupations that are commonly recognised as professions, such as teaching, engineering or nursing, require advanced training or education. This requires curricula design and codification of credential requirements.

Practitioners may hold conflicting views about which competencies matter most, how their work should be carried out and what standards

44 UK Government, 'Introduction to AI Assurance' (n 1).

45 Public, 'Driving AI Assurance in the UK'https://view.publitas.com/public-1/driving-ai-assurance-in-the-uk/page/1 accessed 19 December 2024.

46 Harold L Wilensky, 'The Professionalization of Everyone?' (1964) 70 *American Journal of Sociology* 137 https://www.journals.uchicago.edu/doi/abs/10.1086/223790 accessed 5 June 2025.

47 ibid.

48 Jodi L Short, Michael W Toffel and Andrea R Hugill, 'Monitoring Global Supply Chains' (2016) 37 *Strategic Management Journal* 1878 https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.2417 accessed 5 June 2025.

should be used to assess it. In the case of AI assurance, there is still no shared agreement on what best practice looks like.

Given the rapid pace of technological change, the evidence base for many proposed practices is still in early stages of development. Formalising the evidence base will likely prove to be an ambitious endeavour.

Drivers of professionalisation of AI assurance

Many policymakers suggest AI assurance can contribute to AI oversight and accountability despite its challenges. Several laws and regulations mandating AI assurance assessments have been proposed, with recommendations for both internal and external assurance.

The EU's AI Act is the most notable example. The AI Act requires companies to conduct a variety of assurance activities. These include generating and providing system and governance documentation, evaluating system robustness and validating high-risk systems externally. Other regulations include New York City's Local Law 144, which requires providers of hiring algorithms to conduct bias audits.[49]

Proposed legislation like the United States' Validation and Evaluation for Trustworthy (VET) AI Act would require both internal and external assurance of AI systems.[50] Similarly, a proposed California bill would establish a third-party assessment model by multi-stakeholder regulatory organisations (MRO) which would be certified by the Attorney General.[51]

> Assurance efforts can help an AI company demonstrate compliance with legal requirements

49 Ada Lovelace Institute, 'Code & Conduct' (n 15)

50 'Congress.Gov | Library of Congress' https://www.congress.gov/bill/118th-congress/senate-bill/4769/text accessed 5 June 2025.

51 'SB 813- AMENDED' 813 https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB813#99INT accessed 20 June 2025.

These existing and proposed policy proposals and regulations have helped to shape an emerging second- and third-party AI assurance ecosystem and create demand for AI assurance providers.[52] Assurance efforts can help an AI company demonstrate compliance with legal requirements in jurisdictions where regulation applies.

Outside of policy, a variety of industry and civil society actors have proposed voluntary assurance standards and frameworks to provide a complementary approach to risk management.

These include: the International Organisation for Standardisation (ISO)'s standard 42001, which provides requirements for internal organisational management practice to implement and maintain AI-based products or services; the European Telecommunications Standards Institute (ETSI) Technical Specification 104 223, which synthesises cybersecurity best practices toward standards for secure and safe AI systems,[53] and the National Institute of Standards & Technology (NIST)'s AI Risk Management Framework, which provides detailed guidance for internal risk management of AI systems.

There are also a growing number of assurance companies operating globally that offer second- and third-party assurance services to willing organisations. Although there is no horizontal AI regulation in the UK, it is estimated that there are more than 500 companies offering AI assurance goods and services, with around 80 designated, specialised AI assurance companies.[54]

These policies, standards and frameworks provide a focal point for developers and deployers of AI to adopt practices that manage the risks across the development and deployment lifecycle. This also provides a floor for professionalisation efforts to work from.

At the time of writing, national governments and the global political economy are increasingly moving toward a deregulatory position for AI

52 Lara Groves and others, 'Auditing Work: Exploring the New York City Algorithmic Bias Audit Regime', *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2024) https://dl.acm.org/doi/10.1145/3630106.3658959 accessed 8 July 2024.

53 Phil Muncaster, 'ETSI Unveils New Baseline Requirements for Securing AI' (Infosecurity Magazine, 24 April 2025) https://www.infosecurity-magazine.com/news/etsi-baseline-requirements/ accessed 5 June 2025.

54 'Where Are We Now with AI Assurance?' https://www.icaew.com/insights/viewpoints-on-the-news/2025/feb-2025/where-are-we-now-with-ai-assurance accessed 20 June 2025.

'Competitive advantage' may be an incentive for companies to voluntarily adopt assuranc

and other technologies. Research and development, economic policy, and industrial strategy are focusing more on investment and innovation.[55]

This shift in norms and commitments has implications for AI assurance as a governance and safety-focused set of practices. This is because compliance will likely be removed as a motivator for companies to adopt assurance.

Instead, market-driven factors, like preventing reputational damage stemming from unassessed and underperforming systems, or increasing customer trust, may provide a 'competitive advantage' incentive for companies to voluntarily adopt assurance.

Similarly, adopting assurance can signal to individual and institutional investors that a company has meaningfully reduced the risk of high-profile or high-cost failures. These strategies already exist as incentives for businesses, and the professionalisation of AI assurance could better support these goals.

The uncertain political economic climate and absence of definitive evidence about what assurance activities are effective, and under what conditions, should not delay efforts to leverage assurance practices to mitigate risk. Instead, it underscores the need for adaptive frameworks that can evolve alongside both the technology and the growing body of evidence on effective risk management.

Building a professionalised AI assurance industry will require institutional mechanisms that support this kind of iterative learning, standard-setting and course correction over time. Growing AI assurance into a professional industry can support the development and adoption of safe, reliable and effective AI technologies that deliver benefits for people and society.

55 For example, see US Vice President JD Vance's speech at the 2025 AI Action Summit in Paris, France. https://www.youtube.com/watch?v=64E9O1Gv99o

Research findings

In our interviews, AI assurance practitioners and experts were confident about the role professionalisation could play in supporting their field but highlighted several areas of ambiguity. These must be resolved in order to move towards a more professionalised industry. We found:

- To become a professionalised field, AI assurance must coalesce around a well-defined scope, required competencies and core practices.
- Standards will have a direct impact on the scope of AI assurance, both enabling and constraining assurance activities.
- Accreditation or certification of AI assurance professionals is not a silver bullet that will solve all challenges in AI assurance adoption.
- Regulatory and market forces are likely to be the primary drivers of professionalisation.

Our findings help inform ongoing professionalisation efforts in the AI assurance industry and signal the importance of strengthening the impact of AI assurance practices. While assurance activities should not be considered a panacea for AI risk, our findings suggest that further coordination and standardisation will better enable positive outcomes for AI for people and society.

To aid interpretation of the findings, we note that all quotes are presented in their original context – some definitions of key terms, such as 'audit' or 'assurance', may not have parity with those used by the researchers and provided in the glossary.

AI assurance must coalesce around a defined scope, required competencies and core practices to become a professionalised field

There is no consensus about the activities that make up AI assurance

In our interviews, practitioners broadly agreed that the purpose of AI assurance is to help establish the trustworthiness of AI systems, the organisations that develop them and claims made about AI systems' behaviour or performance.

However, when asked to elaborate on the necessary core competencies and central practices for these purposes, interviewees provided a wide array of methods, disciplinary approaches and knowledge bases. This suggests that the surface consensus may mask deeper disagreements.

For example, although many practitioners discussed system evaluation as a general assurance activity, they differed in their definition of evaluation. Several interviewees with experience in fields like financial services emphasised that the term 'audit' carries specific and regulatory significance and should not be used informally.

These practitioners found it important to define AI audits with precision – especially in terms of the relationships between stakeholders, the standards applied and the consequences of audit findings.

Others, however, viewed the term 'audit' more pragmatically and saw it as a useful shorthand to communicate assurance services to clients and stakeholders. As a result, our interviews revealed no coherent set of activities that can be described as 'AI assurance'. The boundaries of assurance work remain blurry.

A range of competencies are required for AI assurance

At a high level, interviewees identified three core knowledge areas as essential for AI assurance professionals: technical fluency, legal and policy acumen, and an understanding of risk management.

Many interviewees emphasised the importance of technical fluency to evaluate systems rigorously and interpret results accurately. Many felt a solid foundation in data science and statistics was critical. For instance, one interviewee with a background in auditing described collaborating with expert data scientists from the financial sector, who applied concepts from model risk management. Their expertise from the financial domain, an area with well-established model evaluation practices, proved valuable across other AI domains as well.[56]

In addition to technical knowledge, participants discussed the importance of facility in legal and policy concepts, as well as the ability to translate them into technical approaches. As one participant noted:

'We definitely need people that can turn legal requirements into technical specifications.'[57]

This translation requires both an understanding of relevant laws, standards and regulatory frameworks in a given jurisdiction, and the skill to convert them into actionable recommendations relating to AI systems or organisational processes.

In sectors like healthcare or finance, professionals must also be able to interpret how industry-specific regulations intersect with broader AI assurance frameworks.

Some interviewees emphasised that a solid understanding of corporate risk management is also essential to effective AI assurance. Practitioners with risk management expertise are adept at systematically identifying, assessing and mitigating potential threats to a company's goals, helping to reduce losses and inform strategic decisions.

Such expertise can, for example, give practitioners a sound understanding of the distinction 'between providing pre-audit services[58] and being an auditor'[59] and recognising the professional boundaries between these roles to ensure auditor independence is not compromised.

56 P12, AI assurance

57 P3, AI auditing

58 A pre-audit service could include data cleaning in preparation of the audit, or providing an assurance tool to support internal risk management https://dl.acm.org/doi/fullHtml/10.1145/3630106.3658959

59 P1, AI auditing

Some practitioners further added that those with knowledge of how to conduct assurance from other industries – such as finance or cybersecurity – were best-suited to translate that knowledge into risk management within the AI domain.

Practitioners generally agreed that all three core competencies are essential to effective assurance, but acknowledged that it's unlikely any one professional would have deep expertise in all of them.

Instead, they emphasised that assurance is best conducted by a team whose members collectively bring strong, specialised knowledge across the areas, rather than by generalists with shallow familiarity across each.

Having a team of interdisciplinary experts has a number of advantages. Different forms of assurance often require different expertise. As one interviewee said: 'The competencies and the skill are going to depend a great deal on what kind of claims you're trying to verify, and what kind of evidence you're trying to draw from.'[60]

The necessary expertise to assure a system or organisation may often be unclear upfront. If teams only reflect a narrow range of competencies, they may not have sufficient knowledge to address all of the relevant considerations once the requirements become clearer.[61]

Interdisciplinary teams are also best positioned to approach AI appropriately as a sociotechnical system, a perspective that is especially important to adopt when assurance focuses on organisational claims or the impacts of AI systems on people and society.

And while technical expertise is crucial for verifying system components and conducting quantitative analyses, these efforts are often strengthened and their gaps filled by qualitative methods and interpretive insights from the social sciences.

60 P14, AI assurance

61 This may have implications for accreditation of an assurance firm, who in order to gain the credentials, may need to demonstrate they employ people with the collective array of skills.

One participant with a background in auditing framed it as:

> 'You cannot code for a world you don't understand.'[62]

For example, practitioners with deep sociological expertise in how bias unfolds in housing markets may be well equipped to define fairness tests for AI systems that risk allocating housing resources inequitably.

Core practices in AI assurance will differ between narrow and general-purpose AI systems

While interviewees broadly agreed on the key knowledge areas and skills needed for AI assurance, they highlighted challenges in translating these abstract categories into concrete practices suited to specific systems and deployment contexts.

In particular, participants noted that the skills and competencies required for AI assurance may differ between traditional machine learning systems and foundation models, especially the most capable foundation models.[63]

Foundation models, including generative AI systems, generate a wide range of outputs and can be applied across diverse use cases. This makes it difficult to determine what, exactly, should be assured – and what collection of assurance activities can support robust claims about trustworthiness.

Practitioners working with traditional predictive machine learning systems often focus on identifying and mitigating well-known risks such as bias and misinformation, or other established risks to trust and safety. In these cases, many evaluation and auditing methods are reasonably well-established, providing a basis for consensus on best practices.[64]

62 P3, AI auditing

63 Sometimes such models are referred to as 'frontier models.' See: 'Explainer: What Is a Foundation Model?' https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/ accessed 24 July 2023.

64 For example, 'ISO/IEC TR 24027:2021(En), Information Technology — Artificial Intelligence (AI) — Bias in AI Systems and AI Aided Decision Making' https://www.iso.org/obp/ui/en/#iso:std:iso-iec:tr:24027:ed-1:v1:en accessed 20 June 2025.

For generative AI systems, on the other hand, the rapid pace of development has outstripped the growth of corresponding methods for evaluation, assurance and auditing, even though they pose many of the same types of risks as traditional systems.

Part of the challenge of evaluating generative AI systems is that their outputs are not deterministic. In other words, when a system is given the same prompt more than once, it is likely to produce slightly different answers. In some cases, seemingly minor variations to input prompts (for example, spelling or synonyms) can result in notably different outputs.[65]

Therefore, a crucial part of the skillset for assurance of generative AI systems is prompt engineering and robustness testing. This involves designing and testing prompts to evaluate system behaviour, identify risks and ensure outputs align with desired standards or requirements.

> As generative systems continue to evolve, the specific skills necessary to sufficiently elicit the underlying capabilities of the system or its risks will likewise need to adapt.

Measurement approaches from the empirical and social sciences offer a way forward for defining and operationalising measurements of complex concepts that resist easy quantification.[66]

Core practices for frontier AI systems – the most capable generative systems on the market – will likely differ both from traditional predictive machine learning AI and other forms of less capable generative systems.

Several practitioners highlighted that frontier systems may raise new categories of risk, including autonomous research and development,[67] deception, and threats involving chemical, biological, radiological or

65 Norah Alzahrani and others, 'When Benchmarks Are Targets: Revealing the Sensitivity of Large Language Model Leaderboards' (*arXiv*, 1 February 2024) http://arxiv.org/abs/2402.01781 accessed 15 March 2024.

66 Hanna Wallach and others, 'Evaluating Generative AI Systems Is a Social Science Measurement Challenge' (*arXiv*, 17 November 2024) http://arxiv.org/abs/2411.10939 accessed 20 March 2025.

67 Yutaro Yamada and others, 'The AI Scientist-v2: Workshop-Level Automated Scientific Discovery via Agentic Tree Search' (*arXiv*, 10 April 2025) http://arxiv.org/abs/2504.08066 accessed 5 June 2025.

nuclear (CBRN) capabilities, which may require different practices and domain expertise to evaluate.

> Looking ahead, the emergence of increasingly autonomous or 'agentic' systems introduces even greater complexity.[68] [69]

As one practitioner noted, the evaluation practices for these systems differ significantly from those for existing generative models. In their words, these approaches are '80% non-overlapping'.[70]

Some assurance providers for frontier systems expressed concerns about AI systems that could either improve their own functioning or autonomously spread themselves to new domains. If these risks were to materialise, systems with these capabilities would likely also demand different assurance approaches.

Implications for professionalisation of AI assurance

Efforts to professionalise AI assurance face a range of tensions: between standardisation and adaptability, specialisation and integration, and individual breadth and team diversity.

The importance of shared and flexible definitions

For AI assurance to professionalise, the industry needs to define the core knowledge and practices that its practitioners should share. But they may not capture the full array of risks nor keep pace with the evolution of AI technologies if these are defined too narrowly or rigidly.

Competencies that are effective today may prove inadequate for emerging systems, especially as new capabilities introduce new risks.

---

68 Ruchika Joshi, 'Before AI Agents Act, We Need Answers | TechPolicy.Press' (Tech Policy Press, 17 April 2025) https://techpolicy.press/before-ai-agents-act-we-need-answers accessed 5 June 2025.

69 Miranda Bogen, 'It's (Getting) Personal: How Advanced AI Systems Are Personalized' (Center for Democracy and Technology, 2 May 2025) https://cdt.org/insights/its-getting-personal-how-advanced-ai-systems-are-personalized/ accessed 5 June 2025.

70 P9

At the same time, without some agreement on what practitioners should know and do, the assurance industry will struggle to build a cohesive professional identity.

Breadth of knowledge and skills can come at the expense of depth but too much specialisation may limit adaptability

As our findings suggest, assembling a diverse team with deep specialisms would be more effective than reliance on a single professional with a more generalist skillset (for example, varying degrees of technical, legal and risk management proficiency).

However, there is a risk that assurance providers will find this difficult to operationalise, with companies facing pressure to make more hires than they can otherwise support.

The importance of effective translation and communication across teams

Our research shows that teams of assurance professionals with a diversity of competencies may be beneficial to deliver a holistic approach to assessing AI systems and organisational claims.

However, it is not a given that such individuals can work effectively together: a necessary competency is the ability to translate and communicate across disciplines, weaving multiple methods and approaches into a coherent effort.

For example, a technical professional may be challenged by the task of interpreting a legal doctrine for the development of a relevant quantitative model evaluation. A legal-professional colleague might be able to help with interpretation, but would also need to be able to effectively translate and communicate relevant details to their technically minded colleague.

Similarly, a technical professional who can interpret and communicate the results of technical artefacts to policy-oriented colleagues may help an assurance team identify and prioritise relevant risks more effectively.

Even if assurance organisations assemble teams that represent relevant disciplines, cross-discipline project teams will only be effective in completing assurance efforts if their members have the skills to collaborate at the intersection of their disciplines.

Standards will have a direct impact on the scope of AI assurance, both enabling and constraining assurance activities

In AI assurance, standards – such as frameworks, specifications, or criteria – serve as valuable bedrocks for defining how systems ought to be developed, evaluated and maintained to ensure consistency and safety. They can also specify the governance processes that organisations ought to follow when building AI systems.

Interviewees stressed the need for an authoritative source of standards to ensure consistency – a normative framework that defines 'what [companies] ought to do',[71] while noting that no such universally accepted sets of standards currently exists.

Some practitioners rely on evolving ISO standards, certain IEEE (Institute of Electrical and Electronics Engineers) standards or specific regulatory requirements.[72] Though not a formally recognised standard like ISO or IEEE standards, many interviewees reported turning to NIST's AI Risk Management Framework for guidance.[73]

Participants in our research expressed concern that existing standards proposals are neither precise enough nor sufficiently flexible to support effective assurance. They observed that some standards function as binary checklists, instructing assurance professionals to assess safety or efficacy based on whether a particular organisational practice is in place.

However, these standards often fail to capture how well the organisation actually implements that practice – for example, whether an impact assessment is thorough and meaningful. Such constraints could undermine the utility that professionalisation of AI assurance is envisioned to provide.

71 P15, AI assurance

72 CEN and CENELEC are the European standardisation bodies responsible for, respectively, general industry and electrotechnology. The European Commission has charged them with turning the AI Act's broad legal requirements into concrete harmonised standards. Once adopted, these standards will give AI providers a presumption of compliance when placing high-risk systems on the EU market. Companies operating across jurisdictions may face strategic decisions about whether to maintain separate versions of their products. And as alternative standards emerge outside the EU, multinational firms will increasingly need to navigate—and reconcile—overlapping or conflicting requirements for AI systems and their governance.

73 National Institute for Science & Technology (US), 'AI Risk Management Framework' [2021] NIST https://www.nist.gov/itl/ai-risk-management-framework accessed 20 June 2025.

Our interviewees reflected on several notable roles that standards might play in the assurance process. Standards can provide clear metrics and promote consistency in how assurance is conducted across providers and organisations, ensuring that all assessments follow comparable methods and meet common expectations, even across different providers.

Consistency would improve the reproducibility of assurance processes and provide clear guidance for both evaluators and those undergoing assessment. As one interviewee with an auditing background explained:

'Standards allow auditors to know for a given requirement exactly what they're looking for to determine compliance, and those seeking compliance know exactly what they must provide to meet that requirement.'[74]

Beyond consistency, several interviewees noted that clearly defined and operationalised standards can also help reinforce the independence of assurance providers. Without externally defined, credible standards, assurance providers may struggle to demonstrate that their findings are 'objective' or methodologically sound.

In this view, standards not only guide what assurance looks like but also help uphold the integrity and trustworthiness of the practitioners themselves. Standards can also help enable comparisons – it is far easier for a customer or a regulator to see which of two systems assessed against the same standard has fared better, than for systems assessed against disparate criteria.

While there was broad agreement on the value of unified standards, there was little consensus on who should set them. Far from being an apolitical process, standards present an opportunity for stakeholders seeking to influence what assurance entails and who gets to define it.

Our evidence reflected this dynamic, as participants from several different organisations suggested that their own organisations were best positioned to drive standard development.

74 P1, AI auditing

Apart from debates about who should define standards, interviewees also disagreed on their ideal substance and the appropriate level of abstraction. As one participant involved in AI assurance certification noted: 'It's really difficult to assure something [against] a generic standard.'[75]

For instance, ISO's industry-agnostic standard for AI management was seen as too vague. Participants noted that it lacks the specificity needed for consistent application – both within industries (for example, two assurance providers in finance might interpret it differently) and across industries (for example, finance versus healthcare).

Without concrete guidance around operationalisation of such generic guidance, 'the discretion is left up to [the assurance provider] to determine what's acceptable or not based on the system that they're evaluating at the time.'[76] Effective standards must strike a balance between being specific enough to ensure consistency and flexible enough to fit the varied contexts where AI is deployed.

As the field of AI assurance evolves, more specialised standards tailored to specific domains may develop; this is likely to evolve in a similar manner to how building codes eventually established precise definitions, such as what counts as 'triple-pane glass'.[77]

One participant with experience in AI auditing suggested that, in the meantime, standardised 'templates' could be created for assurance in common application areas.[78]

In domains like hiring or content moderation, stakeholder concerns about AI are already well understood and could serve as a foundation for creating assurance templates specific to those contexts.

These templates could then be adapted to related use cases; for example, a template designed to evaluate AI systems that flag toxic content on social media might also be useful for assessing potentially harmful outputs from large language models (LLMs).

75 P4, AI assurance certification
76 P4, AI assurance certification
77 P4, AI assurance certification
78 P6, AI auditing

While mapping stakeholder concerns and operationalising them into templates in a new domain requires significant effort, that work could inform similar efforts across a broader set of related applications.

Implications for professionalisation of AI assurance

In grounding AI assurance efforts in a set of common expectations and definitions, standards are essential to the professionalisation of the AI assurance field. But developing standards that both support this goal and remain adaptable to the diverse needs of clients presents a significant challenge.

There are different levels of tolerance for 'imperfect' standards

Among all the themes raised in our research, interviewees were most aligned on the importance of developing consensus-based standards. Yet, interviewees raised concern about how poorly designed standards could undermine AI assurance, with existing standards falling short of meaningfully improving the safety or functionality of AI systems.

Some viewed the adoption of even imperfect standards as a step toward progress, creating momentum and shared direction. However, others worried that even short-term adoption of inadequate standards risks reducing assurance to a symbolic compliance exercise, ultimately undermining both the impact of assurance efforts and the credibility of the profession.

Several interviewees pointed to New York City's Local Law 144, which mandates independent bias audits for automated employment decision tools, as an example of the limitations of current approaches. The law specifies the use of a particular approach to bias measurement: impact ratio. This is a concept from US federal civil rights law grounded in the legal theory of disparate impact that considers disproportionate exclusion or harm to protected groups such as those defined by race, gender or age.[79]

79 Lara Groves and others, 'Auditing Work' (n 52).

While New York City's law prescribes a method (bias audits) and metric (impact ratio), many participants viewed the law's approach as an inadequate proxy for AI system risks. As one interviewee explained, it 'focuses on measuring bias using a specific metric but does not account for a system's overall accuracy'.[80]

This leaves a significant gap, given that even unbiased systems can be harmful if they are inaccurate. Nevertheless, some saw the law as a meaningful step toward normalising assurance practices, even if its initial implementation is incomplete.

Process-based and binary standards may not raise the quality of assurance, nor improve the functionality or safety of AI systems

Participants expressed scepticism about reliance on process-based standards, especially those centred on organisational practices. Many pointed to ISO's Artificial Intelligence Management System (AIMS) standard (ISO 42001) as a key example of why such frameworks may provide limited utility for assurance goals.

A common criticism was that these standards often rely on binary criteria – assessing only whether a process exists, without evaluating the quality or rigour of its implementation. As one participant explained:

'There are certain criteria in ISO 42001 where I could comply by doing the littlest amount or by doing an enormous amount. Both count as compliance, and if I could build a Walmart between the two, then it's not really useful.'[81]

In other words, binary process checks lack the nuance needed to drive high-quality system development.

Notably, these standards don't necessarily translate into better AI outcomes. While strong governance and risk management processes should, in theory, reduce the likelihood of harm, participants noted that this connection is tenuous – especially when compliance with process-based standards can be achieved through minimal or superficial efforts.

80 P8, AI assurance
81 P1, AI auditing

Accreditation or certification of AI assurance professionals is not a silver bullet that will solve all challenges in AI assurance adoption

Nearly all our participants felt that some form of accreditation and certification of AI assurance was beneficial to professionals and would-be professionals. Many argued that by undergoing certification, AI assurance practitioners would be better equipped to market the quality and value of their services to clients, which in turn would help demonstrate trustworthiness.

What is accreditation and certification?

Both accreditation and certification are considered core components of a professional industry. While these two terms are sometimes used interchangeably, they serve distinct roles within the context of assurance.

At a high level, **certification** entails verifying that a person (for example, an assurance professional) or thing (for example, an AI product) meets specific criteria, while **accreditation** entails recognising that the organisation doing the certifying (for example, an assurance service provider) is qualified to conduct that assessment.[82]

In the context of professionalisation, 'certification' is more commonly used to refer to a recognition of competence for the professionals within a particular industry (either employed by organisations or self-employed), which typically involves gaining special qualifications via training programmes and examinations.[83]

However, we note that in AI assurance in theory and practice, 'certification' has been used to refer to *both* professionals and their organisations[84] and AI products or systems.[85] For example, companies may claim their AI products are 'certified' if they are placed on the market with a certificate detailing that certain safety standards have been met. A UK survey report finds that 53 per

82 Within our study, interviewees typically discussed certification with respect to professionals rather than systems. For a discussion of whether systems comply with pre-defined criteria, see Section X on standards.

83 Inioluwa Deborah Raji and others, 'Outsider Oversight' (n 6).

84 Philip Moreira Tomei, Rupal Jain and Matija Franklin, 'AI Governance through Markets' (*arXiv*, 29 January 2025) http://arxiv.org/abs/2501.17755 accessed 11 February 2025.

85 Peter Cihon and others, 'AI CERTIFICATION: Advancing Ethical Practice by Reducing Information Asymmetries' [2021] *IEEE Transactions on Technology and Society* 1 https://ieeexplore.ieee.org/document/9427056/ accessed 12 October 2021.

cent of respondents wrongly believe AI tools are required to be certified before reaching the market.[86]

Historically, accreditation refers to an assessment of the capability and competence of an *organisation* conducting assurance (as opposed to individual employees of that organisation) and whether they are in compliance with national and international standards.[87]

Accreditation is generally conducted by an official body. For example, the Public Company Accounting Oversight Board (PCAOB) oversees and accredits auditors for financial auditing in the US. Similarly, the UK Accreditation Service (UKAS) can accredit institutions like medical laboratories, which conveys to regulators, patients and others that the institution meets relevant ISO standards for quality and competence.

Accredited or certified providers can differentiate themselves as offering high quality services and products within an industry compared to others who lack those credentials. This can benefit competitive, consumer and regulatory outcomes. Accreditation and certification also confer legitimacy on professionals, and help to acknowledge and institutionalise knowledge within the profession.[88]

In the context of AI assurance, certification and accreditation may serve multiple goals. Certifying AI assurance professionals may enhance the perceived legitimacy and trustworthiness of the services they offer, particularly where those services are somewhat novel,[89] [90] as well as help filter out actors who lack necessary competencies. This may also help to legitimise the project of 'AI assurance' as a whole.

Several international certification efforts for AI assurance are already underway. Organisations such as BABL AI and ForHumanity train and certify AI auditors according to their own training programmes. Relatedly, the International Association of Privacy Professionals (IAPP) has drawn on their experience providing training to privacy professionals and data protection officers to develop an AI governance training course and related certification. The

86 Lydia Preston, 'Polling Data: Consumer-Facing Certificates as an Incentive to Improve Frontier AI Safety and Security' https://www.longtermresilience.org/wp-content/uploads/2025/02/Polling-Briefing-CLTR-.pdf accessed 7 July 2025.

87 UKAS, 'Accreditation vs Certification: What's the Difference?' (2022) https://www.ukas.com/accreditation/about/accreditation-vs-certification/ accessed 15 June 2025.

88 Jennifer L Bartlett and Josef Pallas, 'Accreditation and Certification' in Craig Carroll (ed), The SAGE Encyclopedia of Corporate Reputation (Sage 2016) https://us.sagepub.com/en-us/nam/the-sage-encyclopedia-of-corporate-reputation/book244532 accessed 20 June 2025.

89 Apollo Research, 'A Causal Framework for AI Regulation and Auditing' (Apollo Research) https://www.apolloresearch.ai/research/a-causal-framework-for-ai-regulation-and-auditing accessed 20 January 2025.

90 Khoa Lam and others, 'A Framework for Assurance Audits of Algorithmic Systems', The 2024 ACM Conference on Fairness, Accountability, and Transparency (2024) http://arxiv.org/abs/2401.14908 accessed 5 June 2025.

International Association for Algorithm Auditors (IAAA) also offers training programmes and codes of conduct for AI auditors.

One of our interviewees defined certification in AI assurance as a combination of elements that could comprise training, education and legibility of qualification, while others highlighted that certification might help clarify responsibilities and roles of assurance providers.

Certification could have the effect of increasing the adoption rate of AI in different contexts. One participant reflected that: 'Companies may feel there's less risk if there's some certification from which they can draw talent to conduct evaluations, which are then assured.'[91]

Some felt that streamlining emerging training programmes and curricula for AI governance and assurance[92] into accreditation and certification programmes would provide economic benefits to the AI assurance industry as a whole. Accredited assurance firms may be seen by potential clients as providing more consistent and higher quality services compared to those without accreditation.

As a result, accreditation could act as a market differentiator. Over time, this may drive broader adoption of accreditation and encourage the professionalisation of assurance services. This would also support a competitive industry, with SMEs and startups in AI assurance able to access the market.

Despite these benefits, efforts toward certification face substantial challenges. No clear consensus emerged from our interviews on who should conduct certification – many interviewees suggested that their own organisation was best placed.

91 P11, AI model evaluations

92 Where AI governance training and certification might pertain to a broader knowledge and skills base - including fluency of AI policy and regulation - than AI assurance, which is more focused on measuring, evaluating and communicating claims about an AI system and the organisational policies around it

It was generally agreed that certification should be supported or institutionalised in some form by governments or national accreditation schemes. This should take the form of a trusted body that could make arbitrations of quality and would therefore be well positioned to adjudicate accreditations. In the absence of such a body, the AI assurance industry risks capture by actors whose primary motivations might not be the interests of the sector.

The development of certification and accreditation processes also depends on agreement about what professionals and organisations should be accredited to do. The lack of consensus on core competencies and practices in AI assurance remains a major obstacle. As one interviewee with a stake in AI assurance certification said:

> 'We need to agree on the capabilities that somebody would need to professionally or responsibly act in that field. So what kind of skills do they need? What kind of capabilities? [...] There can be multiple ways of thinking about it, but there has to be some at least common ways of thinking about those capabilities.'[93]

The scope of certifications will meaningfully influence their impact in the space. On one hand, simpler and broader requirements would lower barriers to entry but give less indication of competency or specialisation. On the other hand, narrow requirements or requirements that demand real world experience of relevant assurance activities could have a limiting effect on participation.

In order to enable broad access to AI assurance certification, one participant suggested that a certification scheme could have 'multiple layers of certification'.[94] This would include one 'layer' for professionals that already have experience in AI assurance and another for those who were newly joining the industry.

93 P15, AI assurance
94 P8, AI assurance

> ‘Layered’ certification might clarify professional pathways and enable longer term career development by setting different expectations for entry-level practitioners compared to advanced levels.

This would also enable specialisation to demonstrate that a practitioner has achieved more method-specific or domain-relevant expertise. This would follow practice in other professions that make use of certification, such as pharmacy.[95]

Interviewees indicated that actors across the ecosystem may need to collaborate on enacting certification. For example, one interviewee felt the assurance industry would be equipped to sufficiently organise and upskill assurance professionals, but thought the role of policymakers would be to support the certification of those professionals into third-party auditors or assurance professionals specifically.[96]

> An enforcement mechanism for malpractice will be required, regardless of the certification scheme’s shape and who is enacting it.

Two interviewees spoke about the importance of ensuring that certifications are enacted with due diligence and that professionals are held to account. They both suggested regulators could provide an enforcement function for individuals or organisations who are found to be violating professional norms or failing to preserve the qualities that they relied on to obtain their credentials where necessary.

One interviewee proposed that the United Kingdom Accreditation Service (UKAS), the UK’s national accreditation body for both products and services, should fulfil an accreditation function for UK AI assurance industry, providing oversight and laying out the rules for a certified assurance provider.

95 Board of Pharmacy Specialties, ‘Why BPS Certification Matters’ (Board of Pharmacy Specialties) https://bpsweb.org/why-bps-certification-matters/ accessed 6 June 2025.

96 P7

Implications for professionalisation of AI assurance

Certification and accreditation in AI assurance can help enforce quality and consistency in services while distinguishing individual professionals and firms. However, designing and implementing these schemes in ways that advance the field's professionalisation meaningfully could be challenging.

Certification may be perceived as having questionable value

Implementing certification schemes for AI assurance professionals may carry some risks. For one, certification schemes do not inherently reduce the information asymmetry between assurance providers and clients, if clients do not have an ability to judge the legitimacy of the credential. Or as one participant invested in certification framed it:

> 'Some [certifications] just look nice, and some of them actually imply some capability. But it's hard to know which.'[97]

Additionally, professionals already working in the field successfully may have little incentive to pursue certification. As one interviewee explained:

> 'There's a lot of people in the trenches that already have a lot of expertise that don't need another certification.'[98]

As with other now-established industries, certification processes could be slow to gain traction among those with established experience, as they may offer little added value. They may also struggle to gain traction among seasoned practitioners whose pre-existing credibility might otherwise transfer to the certification as a whole. In such a scenario, certification could ironically become a signal that someone lacks practical expertise, marking those newer to the field in stark contrast to those with deep, hands-on experience.

97 P8, AI assurance
98 P13, AI assurance

Both internal and external assessments will be needed for evaluating AI systems' efficacy and safety

The difference between internal and external AI assurance certifications

Existing certification efforts in AI and the long tail of certification in other industries, such as aviation[99] and financial services[100] have focused on the role of certification in a third-party or independent assessment ecosystem, as opposed to internal practices.

The current AI assurance context differs in that there is also work on auditing, assessment and evaluation being conducted by internal safety teams at AI labs and some companies. Internal assurance has a different function and different aims to second- and third-party auditing. Particularly as there may only be a small degree of independence from internal assurance practitioners and developers and research scientists.

Some existing AI assurance and AI governance certification schemes are tailored towards internal practitioners,[101] where certification may need to equip professionals with organisational risk management expertise. Schemes tailored towards third-party assurance providers – where external assurance has more of a verification and evaluation function – therefore may not be suitable for internal teams.

It is likely both internal and external assessments will need to work in tandem to make overall evaluations of efficacy and safety for AI systems.[102] For example, one interviewee suggested an internal (first-party) audit could be conducted first, as a technical audit to assess components of the system. This would be followed by a second- or third-party audit to ensure compliance with regulatory requirements or standards. This is similar to the 'four lines of defence' model in financial services assurance.[103] This should be facilitated by certification schemes that address both internal and external practices.

99 Sophie Williams, Jonas Schuett and Markus Anderljung, 'On Regulating Downstream AI Developers' (*arXiv*, 14 March 2025) http://arxiv.org/abs/2503.11922 accessed 28 March 2025.

100 Ada Lovelace Institute, 'New Rules?' https://www.adalovelaceinstitute.org/report/new-rules-ai-regulation/ accessed 16 December 2024.

101 For example, IAPP's Artificial Intelligence Governance Professional training

102 Miranda Bogen, 'Assessing AI: Surveying the Spectrum of Approaches to Understanding and Auditing AI Systems' (n 23).

103 ICAEW, 'The Four Lines of Defence | Assurance Practical Guidance | ICAEW' https://www.icaew.com/technical/audit-and-assurance/assurance/what-is-assurance/four-lines-of-defence accessed 20 June 2025.

Certification schemes may become quickly outdated or obsolete

Certification schemes recognise and reflect a moment in time, which poses two distinct challenges: rapid technological developments could mean professionals may be forced to revisit certifications or work outside them to a large degree. A shifting external policy or legal context may implicate the utility of certifications in some contexts.

As above, some types of 'certification' in AI assurance may be directed towards certifying or validating a product, such as the EU AI Act's 'conformity assessments'. Certification schemes for professionals are likely to be developed and validated rigorously, which is often a time-consuming process.

There is a risk that reliance on skillsets for assessing specific systems leads to certifications losing relevance if those systems become outdated or evolve significantly. This issue is especially acute for frontier AI systems,[104] which reflect rapid advances in AI technology in recent years and necessitate different evaluation skillsets than those used for traditional predictive machine learning models. Advancements in AI have considerable implications for certification schemes for professionals, which need to be designed and implemented to stay up-to-date with AI research and development.

Therefore, any certification framework intended for AI in general must be flexible enough to accommodate assurance of systems that vary widely in function and evolve at different rates. However, if such a framework is too broad, it risks being of little practical use. One of our interviewees characterised the problem:

> 'You need to find something that is general enough, that it actually implies having competency, even if the field moves fast, but also precise enough that it actually has any meaning at all.'[105]

The second pitfall is around the external context changing: certifications that equip professionals to conduct practice under certain laws may

104 Ada Lovelace Institute, 'Emerging Processes for Frontier AI Safety' https://www.adalovelaceinstitute.org/project/emerging-processes-for-frontier-ai-safety/ accessed 15 June 2025.

105 P9

become obsolete if those laws cease to apply or if new ones that require significantly different assurance activities come into force.

A third challenge raised by interviewees is that many current AI assurance certification schemes treat assurance as a horizontal effort, spanning multiple contexts where AI is deployed, such as healthcare, energy, or finance. However, industries often have their own domain-specific standards, regulations, and legal frameworks.

As a result, general certification may intersect with existing requirements in complex ways, which could affect the perceived legitimacy and practical value of a professionalised AI assurance industry. For instance, an AI assurance professional certified through a general programme might find that their training does not sufficiently prepare them to assess compliance or evaluate AI products in the context of financial services.

This gap could erode confidence – both of professionals themselves and of others – in the adequacy of certification as a foundation for and signal of competent practice. A potentially more harmful risk is that a practitioner might *assume* their general certification equips them with the necessary expertise for a specific sector. This could lead to AI systems being incorrectly deemed safe for deployment in environments where they may pose significant domain-specific risks.

Despite the promise of training and certification programmes, the lack of international agreement on the standards or norms for certifying AI assurance professionals may hinder progress toward professionalisation of the field.

Requiring assurance professionals to obtain multiple certifications to evaluate the suitability of products for different international markets would be impractical. Instead, 'reciprocal recognition',[106] where professionals' certifications are recognised by other jurisdictions could address this concern.

Further research and analysis is needed to ascertain whether certification to assess compliance under the Colorado AI Act, for example, could enable reciprocal certification to assess compliance with

106 IFOMPT, 'Reciprocal Recognition' https://www.ifompt.org/About+IFOMPT/Reciprocal+Recognition.html accessed 5 June 2025.

UK evidence shows that a lack of standards and regulation is causing businesses to not adopt AI

the EU AI Act. Additionally, important questions remain on who should be given certification powers, according to what criteria, and how that criteria aligns with emerging best practice for AI assessment.

Regulatory and market forces are likely to be the primary drivers of professionalisation

Our interviewees overwhelmingly put forward two primary drivers of professionalisation of AI assurance: regulation and market drivers. Many referred to AI assurance practices proposed as part of regulations like the EU's Digital Services Act (DSA) and New York City Local Law 144 as helping to routinise and standardise the field, which is likely to be necessary for professionalisation.

It should be noted that many AI assurance professionalisation efforts to date have been led by self-organised consortia, including industry actors. As we have shown, evidence suggests that products and services that have been verified to be high quality and trustworthy are more likely to bring about consumer confidence, allowing consumers to make informed choices and rank different products on metrics like safety.[107]

Recent UK evidence also shows that a current lack of standards and regulation around AI is causing confusion for businesses, resulting in reticence to adopt AI.[108] Consumer confidence is likely to translate to boosted sales, and therefore contribute positively to a company's bottom line. A professionalised industry will perform an important market shaping function in this regard.

The following case study is an illustrative example of the interplay between assurance, innovation, adoption and trust, which may inform efforts to professionalise AI assurance in accordance with those goals.

107 Rafe Uddin and Cristina Criddle, 'Microsoft to Rank "Safety" of AI Models Sold to Cloud Customers' *Financial Times* (7 June 2025) https://www.ft.com/content/02f39b33-fa6e-4bb7-b1f4-8171b50738af accessed 11 June 2025.

108 UK Government, 'Barriers and Enablers to Advanced Technology Adoption for UK Businesses' (GOV.UK) https://www.gov.uk/government/publications/barriers-and-enablers-to-advanced-technology-adoption-for-uk-businesses/barriers-and-enablers-to-advanced-technology-adoption-for-uk-businesses accessed 20 June 2025.

How can market forces support both assurance and innovation? An example from car safety

The car safety regime offers a potentially informative model for AI systems to be assured across the lifecycle, particularly for internal assurance testing by teams developing systems.

Establishing the safety of a car involves evaluating both individual components as well as the vehicle as a whole, ensuring that each part – and its integration into a more complex system – meets minimum safety standards.

The wider quality management mechanisms and organisational structures for governance also require assessment.[109] Likewise, AI assurance should involve the evaluation of components of AI systems – datasets, models, APIs etc. – as well as AI systems overall. Comprehensive AI assurance should also address any organisational process issues that could contribute to safety failures.[110]

The positive impact of car safety assurance also offers a potential model for how AI companies could pursue safe innovation that generates industry and regulatory benefits. One of our interviewees with a background in AI auditing pointed to how Volvo's safety innovation in seatbelts ultimately drove industry adoption of safer designs:

> 'I often use the example of Volvo – [Volvo's] seatbelts were not developed by policymakers. The industry developed around Volvo's seatbelts and then once Volvo showed that they were a good solution to safety in cars, policymakers started to demand that seatbelts are in all cars.'[111]

The 'three-point' seatbelts that are now universal to all automotive vehicles today were designed by an engineer at Volvo, the Swedish manufacturing company, in 1959.[112] Volvo had spearheaded a company culture of safety since its inception, and waived its patent rights to the seatbelt's design.[113]

Volvo conducted a systematic programme of crash testing with their seatbelt innovation to generate evidence demonstrating their increased protection

109 Federica Pizzuti and others, 'An Infrastructure for Safety and Trust in European AI' https://www.adalovelaceinstitute.org/blog/an-infrastructure-for-safety-and-trust-in-european-ai/ accessed 5 June 2025.

110 Jakob Mökander and others, 'Auditing Large Language Models: A Three-Layered Approach' (*arXiv*, 16 February 2023) http://arxiv.org/abs/2302.08500 accessed 13 April 2023.

111 P3, AI auditing

112 Classics World, 'A Brief History of Car Safety Innovations - Classics World' https://classicsworld.co.uk/history/a-brief-history-of-car-safety-innovations/ accessed 5 June 2025.

113 Volvo, 'The Three-Point Safety Belt - over 1 Million Lives Saved' https://www.volvogroup.com/en/about-us/heritage/three-point-safety-belt.html accessed 20 June 2025.

Assurance must be supported by accountability and enforcement mechanisms to protect businesses, people and society from harm

to car occupants.[114] This evidence base helped to make the case for wider dissemination of the three-point seatbelts across the industry.

As industry coalesced around three-point seatbelt adoption, regulation responded: in the UK, seatbelt production in cars became law for car manufacturers in 1965, with the requirement for drivers and passengers to wear them becoming law in 1983 and 1991 respectively.[115] Millions of lives have been saved as a result of this market pull, and the resulting introduction of national seat belt laws.[116] Volvo have also successfully continued to advance their market share on a remit of safety and safe innovation.

This example shows how internally-driven assurance practices can create utility and impact beyond just the level of the company – toward an entire industry, and ultimately, to people and society.

However, another example from car safety – the Volkswagen emissions scandal[117] – highlights the risk that internal assurance standards are used to whitewash unethical practices.

Accordingly, assurance practices must be buttressed by accountability and enforcement mechanisms to protect businesses, people and society from harm, in instances where assurance fails or is thwarted. This will involve coordination between multiple actors across the ecosystem. For example, investors should leverage due diligence processes to create scrutiny and help shape market norms.[118]

Lessons for AI assurance and its professionalisation:

- **Providing trustworthy evidence of safety can offer a competitive advantage for companies developing and deploying AI:** Companies can differentiate themselves by innovating on quality AI safety and assurance practices, which may have positive outcomes towards their bottom line.
- **AI assurance should be underpinned by empirical evidence on emerging**

114 Volvo, 'A MILLION LIVES SAVED SINCE VOLVO INVENTED THE THREE-POINT SAFETY BELT' https://www.media.volvocars.com/uk/en-gb/media/pressreleases/20505/ accessed 5 June 2025.

115 UK Government, 'Thirty Years of Seatbelt Safety' (GOV.UK) https://www.gov.uk/government/news/thirty-years-of-seatbelt-safety accessed 5 June 2025.

116 United Nations, 'Buckling up to Save Lives: UN Celebrates Five Decades of Seat Belt Laws | UN News' https://news.un.org/en/story/2023/06/1137412 accessed 20 June 2025.

117 The German car company, Volkswagen, had marketed thousands of cars on the premise of low diesel emissions for some years. In 2015, the US Environmental Protection Agency (EPA) found that many cars sold in the US had been fitted with 'defeat device' software, so that when under emissions test conditions, diesel engines could artificially inflate the performance of the car. See:'Learn About Volkswagen Violations | US EPA' https://www.epa.gov/vw/learn-about-volkswagen-violations accessed 20 June 2025. And Guilbert Gates and others, 'How Volkswagen's "Defeat Devices" Worked' *The New York Times* (8 October 2015) https://www.nytimes.com/interactive/2015/business/international/vw-diesel-emissions-scandal-explained.html, https://www.nytimes.com/interactive/2015/business/international/vw-diesel-emissions-scandal-explained.html accessed 20 June 2025.

118 Tomei, Jain and Franklin (n 84).

best practice for assessment: Volvo's seatbelt design was constructed alongside a robust programme of safety testing. AI assurance should follow this example: standards and norms for professionalising the industry should emerge from empirical research.

- **Open standards and knowledge can build the case for an AI assurance ecosystem:** Making safety innovations public helped to raise the bar for safety across an entire industry, while maintaining a competitive advantage. The wide dissemination of standards and methods for AI assurance may confer similar benefits.

Regulatory drivers of professionalisation may support wider uptake of assurance

Regulation and market forces were overwhelmingly the two drivers put forward by interviewees as being the most impactful for professionalisation of the AI assurance industry. Interviewees who expressed positive sentiment about the need for regulation to shape and professionalise assurance were often driven by the goal of creating accountability outcomes.

Regulation was identified as impacting professionalisation in two ways: first, through defining what is required of assurance activities, and second, explicitly defining the goalposts for professionalisation. Two interviewees felt that the market for AI assurance services was better formed in well-regulated industries with a lower appetite for risk:

> 'Banks and healthcare companies, manufacturing, all of these are where the auditors or assurers that I speak to are finding there to be a lot of work, because there's a lower risk profile in companies that are already in well-regulated industries.'[119]

119 P15, AI assurance

> Adoption of AI assurance practices in healthcare could build on processes already widely adopted and understood within healthcare, like governance structures and ethical standards, rather than starting from scratch.

Two interviewees suggested that there was not sufficient incentive to adopt assurance practices without regulation and liability regimes in AI in particular – which would place strict guardrails around malpractice from AI developers and also support people in pursuing legal routes for redress in the event of harm.

The enactment of regulatory regimes provides a backdrop for organisations to bake in assurance and audit within their business processes. One interviewee working on AI audit certification felt that the most straightforward path to professionalising the industry would come from widespread mandating of auditing:

> 'For example, [we should have] mandatory annual audits in 2027. That would give us a couple of years to build up the ecosystem [...] The great value of mandatory annual audits is that if you know someone's going to come and check your work every year, you know what to do. You are proactively compliant.'[120]

Mandatory independent audits can create the incentives for creation of a clearly defined and scoped audit process with accompanying metrics. This may in turn address the challenge of scoping required for professionalisation. Regulators themselves were deemed to have an important role in setting goalposts for professionalisation of AI assurance:

> 'What does it mean to be a professional in AI assurance? I would love regulators to weigh in on what is AI assurance and what are the features it needs to have to be AI assurance.'[121]

120 P1, AI auditing
121 P15, AI assurance

Overall, however, most of our interviewees were circumspect about the degree in which regulation could be relied upon to drive professionalisation of AI assurance, given the current advancement towards deregulation in AI.

The market could be a driver of professionalisation

Nearly all our interviewees agreed that an AI assurance industry with professional qualities would help to increase confidence in both the ambition of assurance in AI, and AI systems overall, which could translate to business and market confidence.

We found that, on one hand, the implicit business value of assurance could be a driver of professionalisation. On the other hand, professionalisation itself could contribute to market confidence.

AI systems that have been routinely tested and assured, situated in a market where third-party assurance and evaluation becomes well-rehearsed, would improve the quality and efficacy of systems. They would also serve as a driver for AI adoption by increasing customer trust:

> 'One of the things we see is that audited AI is better at performance [...] over time, there's going to be a strong driver [for assurance industry] in the actual performance.' [122]

This interviewee explained that this incentive is particularly strong as it would benefit both actors using AI, and firms developing or deploying AI. Another interviewee working on delivering audit services also concurred that the evolution of assurance practices under professionalisation may be generated directly 'from a business need'.[123]

If audited or assured systems provide better overall functionality, with the secondary effect of improving safety, this could be a compelling argument for businesses adopting AI assurance.

Assurance was also identified as having a market function between downstream organisations buying AI systems from upstream

122 P3, AI assurance
123 P14, AI assurance

developers and adapting them for their business context. To conduct this transaction, deploying organisations are likely to want a degree of assurance about the system or product they are procuring.

For example, a company could bake in assurance requirements at the point of procurement where it decides to build or iterate on a large language model (LLM) to create a customer service chatbot. One interviewee with experience in AI evaluations explains the potential for assurance and the 'de-risking' effect:

> 'The key mission piece is to reassure companies [...] their liability will be fine and they can make a decision to adopt with more confidence. There's a lot of uncertainty especially if you're a large company as to your legal risk.' [124]

> 'These enterprise companies see high stakes. They're gonna get potentially blocked out of markets if it's the EU AI Act or get significant fines if it's some financial regulator. So those enterprise companies will require a push for AI assurance, even if it's not required, because they might have an internal system that they're developing that they want to make sure that it's being governed or tested in the right way.'[125]

Another interviewee provided insight on how certification could drive business between firms:

> 'We [multinational consulting firm] asked a company we were working with to go through a certification process and for them, it was very intensive, but in the end that allowed them to engage not just with us but with many of our peers. So I suppose from an SME perspective, it's a clear incentive [to undertake certification] to show the world how well they've performed.'[126]

According to our interviewees, organisations may be incentivised to seek out AI assurance for consumer customers in addition to business or enterprise customers.

124 P11, AI model evaluations
125 P15, AI assurance
126 P12, AI auditing

One interviewee pointed out that people and society are often distrustful of AI and sceptical about its proposed benefits, which may drive down demand for certain kinds of AI. Assurances to the public about the safety and fairness of AI systems that comprise products may help to demonstrate trustworthiness and increase consumer confidence in AI. As one interviewee explained:

> 'We should make sure that we build mechanisms that combine the ability of these companies to make money, but in conditions of safety for the users.' [127]

Consumer reluctance to use AI may drive down demand for certain kinds of AI-powered products. In response, companies may see voluntary AI assurance not only as a demonstration of corporate social responsibility, but also as a strategy to boost trust and, in turn, increase adoption and economic returns.

Several interviewees emphasised that professionalising the AI assurance industry could strengthen confidence in assurance providers – and, by extension, in the AI systems they evaluate. One interviewee felt C-suite executives would have a strong role to play in helping to shape a culture of good assurance practice internally:

> 'We can't really talk about a robust third-party assurance ecosystem until there is a little bit of momentum built internally, including how well you're doing on validating your own models.'[128]

Other proposed audiences for assurance include tech industry investors or venture capital companies, or boards of large companies who make consequential decisions about the direction of travel for AI development

127 P4, AI auditing
128 P12, AI auditing

and adoption. They all require credible information to guide their strategy, as well as insurers who might insure against liability risk.

Professionalisation efforts that focus on increasing the legibility of both AI assurance outputs – particularly as the effective or optimal output for an assurance process is still contested – and what the resulting certification for an assurance professional should contain, was felt to be impactful.

Many organisations already bring implicit commitments to certain dimensions of safety as a core company value, such that they might voluntarily adopt assurance. For example, firms might conduct internal forms of assurance – including processes like technical audit and risk assessment – on their AI systems before deploying them on the market, as part of organisational due diligence commitments.

One interviewee working on AI evaluations reflected on the internal and business incentives for developers of foundation models or frontier systems. This was considered as particularly important for foundation models or frontier systems that might introduce systemic categories of risk.

> 'I think that, directionally, labs have a lot of incentive to work and frame safety work as something that happens before you deploy a product. [...] For our capabilities of concern that are national security and catastrophic risk type stuff, we're worried about them for internal development also. I think this is something that's kind of unique to AI, it's maybe more similar to nuclear.'[129]

Finally, one interviewee shared their thoughts on the potential relationship between market and regulatory forces, working together to support AI assurance professionalisation:

> 'I think that regulators also have a responsibility towards the market and regulation exists to facilitate a market and conditions of safety and protection, but also conditions of these companies to make money.'[130]

129 P5, AI model evaluations
130 P3, AI auditing

Personal motivations may contribute to professionalising AI assurance

Interviewees also identified that assurance practitioners' personal motivations for ethical practice could help motivate professionalisation. Many practitioners currently conducting assurance practices felt strongly motivated to conduct work in the space as part of a normative or moral incentive to create positive social outcomes:

> 'I think as an organisation you want to do a couple of things. Number one, you wanna have a clean conscience that whatever you're doing is actually going down the right trajectory and not negatively impacting people, but that hopefully it is positively impacting people. So that is on the individual level or you know, from a social responsibility perspective?'[131]

This interviewee also suggested that high motivation from professionals can create a groundswell of support at the organisational level of AI companies. Other interviewees spoke of a desire to see the industry professionalise to increase their own capacity and skills as practitioners and professionals of assurance. We suggest that leveraging these incentives could support startups and SMEs to demonstrate good assurance behaviours, to gain competitive advantage.

Implications for professionalisation of AI assurance

Overall, professionals' personal motivations and other internal organisational incentives were felt to be weaker drivers of change for professionalisation than external forces. Participants suggested relying on the role of regulation in the current landscape may be a risky strategy, as well as putting forward some limitations of market forces to professionalisation. It will be important for policymakers to consider some of these as potential pitfalls when advancing professionalisation of AI assurance.

131 P13, AI assurance

Regulation creates risk of capture in the AI assurance market

One interviewee suggested that assurance practices driven by regulation may hinder a flourishing assurance market, by creating requirements that are too onerous or technically specific. This would result in an ecosystem where certain practices and services are dominated by a small number of actors, usually those with a high degree of resources or specialisms, which has implications for professionalisation efforts. They put forward an example from a current AI regulation regime which has created capture:

> 'My first reaction when I saw the DSA [EU Digital Services Act] was "Oh my God! Do they intend to create auditing requirements that you know only extremely large auditing organisations are going to be able to provide because of how logistically challenging the requirements they've created [are]. Are they trying to hand off a market to the Big 4 [consultancy firms]?" That's pretty much what's happened so far with that market.'[132]

Demand-side drivers for assurance market growth may also be too diffuse or weak to fully advance and professionalise AI assurance

Over-indexing on market forces for professionalising AI assurance brings risks in a turbulent economic climate, where existing norms, behaviours and priorities by market actors may shift. Evidence has shown that demand for, in particular, third-party AI assurance practices has lagged behind supply.[133] Market-driven levers appear likely to be a key driver of assurance adoption as regulation continues to remain patchy.

132 P14, AI auditing

133 Digital Regulation Cooperation Forum, 'Ensuring Trustworthy AI: The Emerging AI Assurance Market' (n 2).

Recommendations

Our conversations with AI assurance practitioners and experts underscored the important role assurance can play in fostering a healthy AI market and ensuring that AI products are both safe and effective. Safe and effective products, in turn, ensure that people and society can access the benefits of AI technologies while reducing the potential for harm.

At the same time, interviewees pointed to several gaps the field must address to ensure that assurance efforts are robust and adaptable as AI technologies and their applications continue to evolve. Addressing these gaps will be essential not only for improving assurance practices but also for advancing the professionalisation of the field as a whole.

We recognise that our findings do not necessarily reflect the realities of all assurance practitioners and may not have captured all the issues facing the sector today. In this section, we offer recommendations on the basis of our findings, while recognising that our evidence is partial. It reflects the complex and potentially misaligned incentives evident among our interviewees, each from different organisations within the AI assurance ecosystem, and with different aspirations for themselves and their organisation.

Nevertheless, we conclude that there is considerable opportunity for a multi-stakeholder coalition of actors to support professionalisation of AI assurance. This includes civil society, industry bodies, international standards development organisations and national policymakers. Such efforts, as we have argued, will require support from policymakers and regulators – for example, policymakers enacting funding initiatives or subsidies to support uptake of certification schemes.

Recommendation 1: Assurance practitioners and the organisations that support them should clarify which competencies are relevant for assurance across AI systems generally, and which are relevant for specialised contexts

While some core competencies – such as a solid understanding of foundational statistical concepts or knowledge of AI risk management approaches – are broadly applicable across AI assurance, many skills must be tailored to specific contexts and systems. For example, an assurance practitioner working in finance may need different technical, legal and risk management expertise than one working in the social media domain.

Similarly, those evaluating traditional predictive or classification systems may require different skillsets than those aiming to assure generative AI systems. In short, AI assurance demands both field-wide fundamentals and domain-specific specialisations.

To address these challenges, we recommend that AI assurance curriculum developers and standards-setting bodies define core competencies that reflect both the type and capabilities of AI systems and the specific demands of different application domains.

Often, assurance professionals must develop domain-specific expertise – such as navigating HIPAA compliance in healthcare, assessing financial risk under regulatory constraints in banking, or evaluating legal accountability in automated decision-making systems – to ensure AI systems meet the unique standards and risks of each sector.

Policymakers designing AI certification schemes should ensure that their frameworks accommodate this range of expertise. Our recommendation for layered certification tracks (see Recommendation 2) is intended to support this need, enabling professionals to develop and demonstrate competencies aligned with both system type and application context.

Academic institutions and curriculum developers can further this goal by offering specialised modules within relevant degree programmes – for example, focused training in evaluating AI systems related to chemical, biological, radiological and nuclear (CBRN) risks or environmental impacts.

Recommendation 2: Certification of assurance professionals should consist of modular certification 'tracks'

Given the wide variation in required competencies across domains and system types, a one-size-fits-all approach to certification is unlikely to be effective. Instead, certification schemes should mirror this diversity by offering flexible, modular pathways that allow professionals to build and demonstrate relevant expertise over time.

This would follow the approach taken in other professions that require continuing education (such as the legal profession) and recertification on a periodic basis (such as toxicology, which requires professionals to recertify every five years).[134]

One promising approach to modular certification is to create certification programmes tailored to specific skill sets. In support of this idea, one interviewee proposed offering multiple 'tracks', for instance focusing on technical testing, executing risk management processes, or evaluating legal aspects, to ensure a diverse set of capabilities is represented in the assurance process. This also helps to ensure professionals can advance in their professional development by gaining new or auxiliary certifications that reflect certain specialisms.

For example, while certification is not necessary to practice in privacy roles, professionals interested in doing so often start by obtaining a Certified Informational Privacy Professional (CIPP) certification, delivered by the International Association of Privacy Professionals (IAPP), but can then add specialised certifications based on their choice of career. A privacy professional wanting to work in healthcare privacy compliance, can take the HealthCare Information Security and Privacy Practitioner (HCISPP) certification.

In the context of AI assurance, governments can support modular certification by either creating their own AI assurance certifications or accrediting organisations offering these auxiliary certifications. As in the case of privacy, there may be multiple accrediting organisations that can support this.

134 Royal Society of Biology, 'UK Register of Toxicologists' (RSB) https://www.rsb.org.uk/careers-and-cpd/registers/uk-register-of-toxicologists accessed 5 June 2025.

Recommendation 3: Policymakers should take action to promote skills development that could support AI assurance

Policymakers could consider adopting a number of initiatives to drive skills development for AI assurance. For example, policymakers should include AI assurance into wider AI literacy objectives or national skills agendas.

National skills agendas are government-led approaches for identifying, developing, and aligning the workforce skills needed to support economic development, technological innovation, and social responsibility. By including AI assurance certification in national skills agendas, governments can prioritise an inclusive and accessible approach to AI assurance skills development and help to increase participation in AI assurance.

This could include funding for training programmes for professionals exploring options to reskill into AI assurance, and for those already working in the profession. In the UK, policymakers should consider broadening the proposal outlined in the UK AI Opportunities Action Plan to work with Skills England, the planned executive agency at the Department for Education (DfE), and DSIT's dedicated skills department to deliver a programme focused on AI assurance skills, to support the growth of a certified professional industry.

The UK government should also consider including AI assurance and governance skills into the TechFirst digital skills curriculum for schools.[135] In the US, policymakers might consider whether AI assurance-related competencies could be incorporated into efforts like the National AI Literacy Campaign, put forward by the National Artificial Intelligence Advisory Committee, which proposes supporting educational institutions and professional associations in ensuring professionals are equipped with relevant skills.[136]

135 UK Government, 'PM Launches National Skills Drive to Unlock Opportunities for Young People in Tech' (GOV.UK) https://www.gov.uk/government/news/pm-launches-national-skills-drive-to-unlock-opportunities-for-young-people-in-tech accessed 9 June 2025.

136 The White House, 'Executive Order: Advancing Artificial Intelligence Education for American Youth' (The White House, 23 April 2025) https://www.whitehouse.gov/presidential-actions/2025/04/advancing-artificial-intelligence-education-for-american-youth/ accessed 20 June 2025.

Investing in AI assurance will support policymakers' goals of creating beneficial economic impact

Investing in pathways for people to become AI assurance professionals will also support national policymakers' desire to harness AI to deliver thousands of jobs,[137] and ensure AI assurance continues to create beneficial economic impact.[138]

Recommendation 4: Standards setting bodies should create broadly applicable standards as well as tailored standards for specific system types and domain-specific use case

Our interviewees consistently emphasised that if assurance methods aren't well-suited to the type of AI system under review, assurance may fail to provide the protections or guarantees it is intended to deliver. For example, generative AI systems present particularly difficult challenges for assurance because their outputs are non-deterministic and can vary widely in response to the same input.

As these systems become more widespread and more agentic, with the ability to act autonomously, it is increasingly important to establish reliable and valid methods for evaluating their behaviour. Because of the variability in generative AI systems' outputs, evaluation methods for such systems must account for both average behaviour and the range of potential outputs.

These challenges are further compounded in multimodal generative systems that integrate text, audio, images, and video, where interactions across modalities introduce additional complexity and novel forms of risk. Technical standards, including metrics and evaluation procedures, must reflect these realities. Because it is not feasible to account for every possible output, standards must also support the interpretation of performance measurements under uncertainty.

Testing of more specific templates of standards could be supported under controlled environments like sandboxes, in collaboration with national standardisation or national AI Safety/Security

137 UK Government, 'Prime Minister Sets out Blueprint to Turbocharge AI' (GOV.UK, 13 January 2025) https://www.gov.uk/government/news/prime-minister-sets-out-blueprint-to-turbocharge-ai accessed 5 June 2025.

138 Frontier Economics, 'Unlocking the Growth Potential of the UK's AI Assurance Market' (n 16)

Institutes (AISIs), to increase specificity. For example, the recent ISO/IEC TS 42119-7[139] standard for AI red teaming could be tested in a sandbox with an assurance provider against a specific AI application context, like a mental health chatbot.

More broadly, many participants raised concerns about whether existing or proposed evaluation metrics are meaningful or valid proxies for real-world system behaviour. Simplified metrics commonly used to assess AI performance or risk are often only weakly connected to how systems behave in practice.[140 141 142 143 144] Yet many emerging standards already specify which metrics should be used.

Without a stronger empirical basis, such standards risk creating the illusion of oversight without ensuring meaningful protections. Policymakers should therefore invest in research that evaluates the validity and reliability of these metrics, particularly for generative models, and that develops new measurement approaches better suited to their unique characteristics. National AISIs, the private sector, and academic and civil society research groups must all be involved in developing consensus on how to evaluate both traditional predictive machine learning AI and generative AI systems.

While some standards may be broadly applicable across AI systems, many assurance practices will require more narrowly defined specifications. AI assurance can benefit from the experience of well-regulated sectors like healthcare or financial services, which have well-established practices for assurance and risk management. Standards-setting bodies should draw on evidence from how AI

139 'ISO/IEC AWI TS 42119-7' (ISO) https://www.iso.org/standard/91240.html accessed 5 June 2025.

140 Sean M McNee, John Riedl and Joseph A Konstan, 'Being Accurate Is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems', *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Association for Computing Machinery 2006) https://doi.org/10.1145/1125451.1125659 accessed 5 June 2025.

141 Seraphina Goldfarb-Tarrant and others, 'Intrinsic Bias Metrics Do Not Correlate with Application Bias' (*arXiv*, 8 June 2021) http://arxiv.org/abs/2012.15859 accessed 5 June 2025.

142 Ryan Steed and others, 'Upstream Mitigation Is Not All You Need: Testing the Bias Transfer Hypothesis in Pre-Trained Language Models' in Smaranda Muresan, Preslav Nakov and Aline Villavicencio (eds), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Association for Computational Linguistics 2022) https://aclanthology.org/2022.acl-long.247/ accessed 5 June 2025.

143 Laura Weidinger and others, 'Toward an Evaluation Science for Generative AI Systems' (*arXiv*, 13 March 2025) http://arxiv.org/abs/2503.05336 accessed 17 July 2025.

144 Su Lin Blodgett and others, 'Stereotyping Norwegian Salmon: An Inventory of Pitfalls in Fairness Benchmark Datasets' in Chengqing Zong and others (eds), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (Association for Computational Linguistics 2021) https://aclanthology.org/2021.acl-long.81/ accessed 5 June 2025.

assurance is already being implemented in these domains to develop recommendations that are tailored to domain-specific use cases.

One potential model for a collaborative, government-supported investment into research for AI assessment and assurance best practices could be the approach taken by the national government of Singapore. The Infocomm Media Development Authority (IMDA) has developed 'AI Verify', an 'AI governance testing framework and software toolkit'[145] developed in tandem with companies from multiple scales and sectors. AI Verify is designed to allow different organisations to contribute to methods and standards for effective AI assurance and governance processes.

Initiatives like AI Verify contribute to empirical understanding of measurement and assessment of AI systems of varying classifications, which in turn can inform standards development at the international level.

Finally, to ensure these standards are both practical and responsive to real world concerns for specific use cases, standards setting bodies should collaborate with civil society groups and trade unions, whose research, advocacy, and community engagement offer valuable insight into communities and groups impacted by AI, and the related risks that assurance providers may be called on to prevent.

Recommendation 5: Professionalisation of AI assurance should be oriented towards supporting assurance adoption across the ecosystem, including downstream deployers of AI

Efforts to professionalise AI assurance should take an ecosystem-wide approach, extending beyond developers to include both the organisations that deploy AI and the systems they implement. This is especially important when deployers fine-tune or modify upstream models, as these changes can lead to impacts that differ from those assessed by the original developers.

145 Infocomm Media Development Authority, 'Singapore Launches AI Verify Foundation 2023' (*Infocomm Media Development Authority*) https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/singapore-launches-ai-verify-foundation accessed 5 June 2025.

Safety testing, auditing and assurance on upstream foundation models remains pertinent, so AISIs should continue to develop and run evaluations of foundation models, and developers should continue to contract third parties for additional verification (particularly as these actors are likely to be well-resourced in comparison to downstream actors).[146]

By embedding assurance practices such as third-party audits into real-world deployment contexts, the field can better respond to context- and industry-specific risks throughout the AI supply chain. Organisations developing certification schemes (per Recommendation 2) should also adopt an ecosystem perspective to ensure that certifications are designed with the context of AI applications in mind and are responsive to sector-specific concerns and regulations.

As covered in the previous sections of this paper, evidence suggests proactive risk management by companies has the potential to translate to higher profits.[147] To enable flourishing ecosystems of AI assurance towards that goal, policymakers will need to ensure that SMEs and startups are supported to implement assurance. Governments could, for example, offer subsidies or grants for implementation of assurance from second- or third-party providers. In demonstrating trustworthiness via publicised adoption of AI assurance, firms adopting AI could see increased consumer demand.

Policymakers should also consider that firms purchasing AI systems can influence vendor behaviour by requiring certain forms of assurance, such as audit reports or transparency about system capabilities. Policymakers can help by promoting tools like DSIT's AI Management Essentials that support firms in governing the AI systems they procure.

146 AI Now Institute, 'Artificial Power: 2025 Landscape Report' (*AI Now Institute*, 3 June 2025) https://ainowinstitute.org/publications/research/ai-now-2025-landscape-report accessed 5 June 2025.

147 Frontier Economics, 'Unlocking the Growth Potential of the UK's AI Assurance Market' (n 16).

Recommendation 6: Standards, certifications and training for AI assurance should demonstrate how assurance can support both clients' business priorities and accountability goals

Sometimes advanced capabilities of AI systems enable harm, such as facilitating malware development or enabling automated 'spear-phishing' campaigns.[148] More frequently, however, systems create risks when they fail at their intended function. For example, a computer vision model designed to identify cancer in radiographic images that cannot detect areas of concern in images of older female patients demonstrates both a safety failure and a core functionality problem. Such functional shortcomings could directly impact businesses' outcomes such as market adoption or revenue.

As we show in our findings, market adoption and revenue may be powerful incentives for assurance, so if assurance providers frame their evaluations around whether the product consistently delivers the value it promises, they meet companies where their priorities already lie. Over time, second- and third-party assurance providers can build relationships with companies that may allow them to also address product risks beyond basic functionality.

In global markets where regulation plays a smaller role in motivating adoption of AI assurance, for AI assurance to become the norm, providers will need to demonstrate the clear value of their services to clients. If the development of standards, training, and certifications fails to align with business priorities, market adoption of AI assurance could stall.

Conversely, if assurance efforts are clearly tied to the outcomes that matter most to businesses, the field may grow rapidly, even without strong regulatory pressure. However, business priorities must align with some of the accountability goals in AI assurance.

To ensure that AI systems are both effective and do not pose undue risk to people and society, assurance providers will need to balance attention to business-relevant concerns with risks that may not appear directly tied to a firm's short-term success.

148 Fred Heiding and others, 'Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects' (arXiv, 30 November 2024) http://arxiv.org/abs/2412.00586 accessed 5 June 2025.

Recommendation 7: The AI assurance field should cultivate strong professional norms and an ethical culture to complement formal standards and accountability structures

The field of AI assurance should work to define shared norms of practice (for example, through codes of conduct) and actively cultivate a professional culture grounded in ethical responsibility and a commitment to the public good (for example, through regular professional gatherings).

Many practitioners are already motivated by a strong sense of moral purpose and a desire to ensure AI systems contribute positively to society. These normative commitments can help secure organisational buy-in and reinforce best practices.

However, while such motivations are valuable, they are not sufficient on their own. Although practitioners' normative motivations and cultural norms can be influential, they are difficult to scale and may erode under market pressures to deliver quickly or satisfy client demands.

A strong ethical culture must therefore be complemented by external standards, formal mechanisms of accountability, and structural supports that reinforce responsible conduct–especially as the field grows and professionalises.

Conclusion and further questions

Our research provides policymakers with valuable insights from practitioners and experts about the current state of play for the field of AI assurance, along with guidance for advancing professionalisation in the field.

Our research also highlights that professionalisation of AI assurance could bring important benefits, such as helping companies more clearly demonstrate trustworthiness to consumers – and as a result, help reduce the harmful impacts of AI systems by creating safer products.

Among other advantages, a more professionalised industry could clarify expectations between assurance providers and clients, establish credentials that ensure a baseline level of competence, and offer clearer pathways for individuals entering the field.

At the same time, our conversations with practitioners highlighted that the road to realising these benefits is not straightforward. AI technologies are evolving rapidly, and so too are their potential impacts and the risks to be managed. Tensions among stakeholder groups and other systemic challenges – like a disconnect between assurance focused on organisational process and assurance focused on AI system components or outcomes – may further complicate progress. Without deliberate attention to these obstacles, efforts to professionalise the field and to promote the broader adoption of AI assurance may falter.

Because our work is grounded in the experiences of current practitioners, it sheds light on the day-to-day challenges that assurance professionals face now and may continue to face as the industry moves towards professionalisation. This practice-oriented perspective complements the more theoretical work of other research and advocacy organisations that explore the potential benefits of a professionalised AI assurance field.

Efforts to develop recommendations for professionalising AI assurance without incorporating the perspectives of practitioners risk producing guidance that is misaligned with the realities of the field. Firsthand insight into the challenges, constraints, and motivations shaping day-to-day

assurance work, can help pinpoint ways in which proposed standards or structures may prove impractical, ineffective, or difficult to adopt.

Our research also points towards remaining questions about the professionalisation of AI assurance that future research could answer.

How can the AI assurance field reconcile horizontal standards with domain-specific needs?

Our interviewees acknowledged the value in aligning the field around shared language and broad principles, as part of a project to professionalise, but they expressed reservations about the current trajectory of standards-setting in AI assurance. A key question is whether horizontal guidance (guidance that applies across sectors), such as ISO organisational risk management guidelines or NIST's AI Risk Management Framework address all the relevant considerations when applied to specific domains. Several participants suggested that domain-specific standards could help fill the gaps left by generalised frameworks or even serve as more effective substitutes in certain contexts.

However, developing domain-specific AI standards may require more expertise and effort to align with existing sector-specific regulations, standards, and laws. This added complexity could slow both the adoption of AI assurance practices and the broader push towards professionalisation.

Future research could explore how AI assurance maps onto other well-regulated industries, such as healthcare or financial services, and how existing frameworks in those sectors might inform domain-relevant standards for assurance professionals.

How does the absence of well-established evaluation methods for frontier AI systems pose challenges for AI assurance efforts?

The capabilities of frontier foundation models and generative AI systems are rapidly evolving. Existing evidence suggests these systems may not

always perform consistently or reliably,[149] making it essential to establish best practices for evaluation amidst uncertainty.

However, the evolving nature of frontier model risks and capabilities complicates efforts to develop systematic evaluation approaches, whether focused on technical performance, ethical concerns or societal impacts, since the results of a given test may not hold if the system is altered or updated.[150]

Some researchers have proposed that frontier system evaluation, like traditional system evaluation, could adapt measurement approaches for generative AI systems from established methods in the social sciences, which are similarly tasked with assessing abstract concepts (for example, 'fairness') in complex and evolving subjects such as individuals, communities or economies.[151]

While such proposals offer a crucial theoretical foundation for measuring frontier systems, they stop short of outlining concrete implementation details for conducting reliable evaluations of generative AI. Because those methods are still underdeveloped, determining how to codify them into standards and certifications for AI assurance is an acutely complex and unresolved challenge where future research could be beneficial.

An additional consideration is that, for general-purpose systems like foundation models, other AI systems are often used to conduct scaled or automated evaluations.[152] Recent research indicates that Meta may soon use AI to automate 90% of all risk assessments on products, algorithms or features.[153] This raises critical questions for AI assurance and auditing, and further work may be needed to explore which types of assessment must be conducted by humans and which could be automated.

149 Wallach and others 'Evaluating Generative AI Systems Is a Social Science Measurement Challenge '(n 66).

150 Ada Lovelace Institute, 'Under the Radar?' https://www.adalovelaceinstitute.org/report/under-the-radar/ accessed 5 June 2025.

151 Wallach and others 'Evaluating Generative AI Systems Is a Social Science Measurement Challenge' (n 66).

152 These "LLM-as-a-judge" approaches offer greater scalability than human validation studies, enabling the evaluation of hundreds or even thousands of responses for each measurement target (e.g., legality, promotion of self-harm, or the presence of biased or unfair stereotypes). For example, Anthropic uses Clio, an LLM, to categorise user interactions from their Claude model to produce analysis that informs safety and governance. See: Anthropic, 'Clio: Privacy-Preserving Insights into Real-World AI Use' https://www.anthropic.com/research/clio accessed 5 June 2025.

153 Bobby Allyn, 'Meta Plans to Replace Humans with AI to Assess Privacy and Societal Risks' *NPR* (31 May 2025) https://www.npr.org/2025/05/31/nx-s1-5407870/meta-ai-facebook-instagram-risks accessed 11 June 2025.

In addition to the challenges for accountability and liability challenges in the event of safety incidents,[154] this may also present a potential professionalisation challenge – should the AI systems used to validate foundation models be certified, and by whom?

Who is best placed to certify AI assurance providers?

In addition to limited existing consensus around who should conduct assurance and audits of AI,[155] who should certify AI assurance providers is an open question. Certification of professionals may raise the quality of services and provide a trusted institutional structure. Evidence from other industries has continually shown that trusted third-party actors should develop and enact certification.[156] [157]

However, the emergence of certification schemes may not be driven by policy or regulation: as with car safety, enterprise-led innovation can promote industry-level safety improvements, which national and international standards can then buttress through certifications.

Further research could shed light on how certification in different national or industry-specific contexts could offer a path forward. However, we also call upon governments to consider their critical role in scoping, funding and convening the right bodies in order to enact certification schemes for AI assurance professionals.

How should we understand and confront differences in binary and more nuanced standards for AI assurance?

Interviewees report that AI assurance can take one of two forms. Assurance providers could conduct a binary verification that a process has been completed, or they could conduct a more nuanced and granular assessment of the quality, robustness and efficacy of the process undergoing assurance.

154 Ada Lovelace Institute, 'Safe before Sale' (n 33).

155 Merlin Stein and others, 'Public vs Private Bodies: Who Should Run Advanced AI Evaluations and Audits? A Three-Step Logic Based on Case Studies of High-Risk Industries' (*arXiv*, 3 September 2024) http://arxiv.org/abs/2407.20847 accessed 28 March 2025.

156 Friederike Albersmeier and others, 'The Reliability of Third-Party Certification in the Food Chain: From Checklists to Risk-Oriented Auditing' (2009) 20 Food Control 927 https://www.sciencedirect.com/science/article/pii/S0956713509000218 accessed 5 June 2025.

157 Inioluwa Deborah Raji and others, 'Outsider Oversight' (n 6).

On the other hand, many of our interviewees felt that professionalisation focused on binary verification risks 'compliance-washing'. On the other hand, they recognised that a more granular approach to assurance presents challenges around scaling and standardising assurance that verifies divergent processes and structures. Both kinds of assurance will likely be valuable. However, the field must disambiguate the goals of binary versus more specific assurance, and what each can achieve.

Methodology

While there are growing calls to professionalise the AI assurance industry,[158] [159] [160] there is limited empirical evidence on how this process should unfold in practice. The Ada Lovelace Institute partnered with the Center for Democracy & Technology to conduct qualitative research to explore the conditions for, and potential impacts of, professionalising the AI assurance industry.

Our study addresses the current evidence gap by offering rich context-specific insights drawn directly from practitioners in the field. By grounding our findings in practitioner perspectives, we aim to provide actionable guidance for policy audiences – particularly those with an international focus – on how to support the development of a mature, professionalised AI assurance ecosystem.

From November 2024 to February 2025, we conducted 15 semi-structured interviews with a range of practitioners and experts, including:

- Firms providing third- and second-party audits for algorithms or AI products
- Third-party model evaluations technical professionals[161]
- Professionals and experts from technical standards bodies
- Firms offering AI governance training or certification.

158 UK Government, 'Six Lessons for an AI Assurance Profession to Learn from Other Domains - Part One: How Can Certification Support Trustworthy AI? - Centre for Data Ethics and Innovation Blog' (12 July 2023) https://cdei.blog.gov.uk/2023/07/12/six-lessons-for-an-ai-assurance-profession-to-learn-from-other-domains-part-one-how-can-certification-support-trustworthy-ai/ accessed 18 August 2023.

159 ICAEW, 'The Necessary Foundations for Good AI Assurance | ICAEW' https://www.icaew.com/insights/viewpoints-on-the-news/2023/may-2023/the-necessary-foundations-for-good-ai-assurance accessed 5 June 2025.

160 techUK, 'Mapping the Responsible AI Profession, A Field in Formation' https://www.techuk.org/resource/techuk-paper-mapping-the-responsible-ai-profession-a-field-in-formation.html accessed 5 June 2025.

161 Like assurance and audit, there is no consensus on a precise definition for an 'evaluation' of a model. A narrower view focuses on tests of model outputs or behaviours in a pre-deployment setting, and a broader view might incorporate tests of downstream real world impacts on people and society. See: Ada Lovelace Institute, *Under the radar?* (n 150)

We explored the following three research questions:

1. What is AI assurance and what is it setting out to achieve?
2. What role can a third-party professionalised industry play in ensuring AI assurance?
3. What is needed to ensure assurance works well?

We recruited widely for interviews, also targeting professionals and experts in AI assurance with previous experience in other fields such as privacy, for example. We also examined AI assurance and its professionalisation in a variety of contexts and AI systems, from hiring algorithms, to generative AI systems like large language models (LLMs), to capture a diversity of knowledge and experience in assurance across the ecosystem.

Participant IDs and expertise

- P1: AI / algorithm auditing
- P2: Technical standards development
- P3: AI auditing
- P4: AI assurance certification
- P5: AI model evaluations
- P6: AI / algorithm auditing
- P7: Anonymous
- P8: AI assurance
- P9: Anonymous
- P10: AI and privacy
- P11: AI model evaluations
- P12: A I / algorithm auditing
- P13: AI assurance
- P14: AI assurance
- P15: AI assurance

Acknowledgements

This paper was lead-authored by Lara Groves, Amy Winecoff and Miranda Bogen, with substantive contributions from Michael Birtwistle, Nuala Polo, Andrew Strait and Samir Jain.

We would like to thank our interview participants for their time and expert contributions to this study, both as individuals and representatives of organisations:

- Ryan Carrier, Executive Director, ForHumanity
- Gemma Galdon Clavell, CEO and Founder, Eticas AI
- Adam Leon Smith, Chair, AIQI Consortium
- Shea Brown, CEO, Babl AI
- The International Association of Privacy Professionals
- Karim Ginena, Co-Chair, AI Policy Committee, IEEE
- Herbie Bradley
- Borhane Blili-Hamelin, Officer, AI Risk and Vulnerability Alliance
- Jacob Appel, Algorithm Auditor, ORCAA
- Heather Frase, PhD, Principal, VerAltech
- Maria Axente
- Christopher Painter, Head of Policy, METR

As well as those who preferred not to be named.

About the Ada Lovelace Institute

The Ada Lovelace Institute was established by the Nuffield Foundation in early 2018, in collaboration with the Alan Turing Institute, the Royal Society, the British Academy, the Royal Statistical Society, the Wellcome Trust, Luminate, techUK and the Nuffield Council on Bioethics.

The mission of the Ada Lovelace Institute is to ensure that data and AI work for people and society. We believe that a world where data and AI work for people and society is a world in which the opportunities, benefits and privileges generated by data and AI are justly and equitably distributed and experienced.

We recognise the power asymmetries that exist in ethical and legal debates around the development of data-driven technologies, and will represent people in those conversations. We focus not on the types of technologies we want to build, but on the types of societies we want to build. Through research, policy and practice, we aim to ensure that the transformative power of data and AI is used and harnessed in ways that maximise social wellbeing and put technology at the service of humanity.

We are funded by the Nuffield Foundation, an independent charitable trust with a mission to advance social well-being. The Foundation funds research that informs social policy, primarily in education, welfare and justice. In addition to the Ada Lovelace Institute, the Foundation is also the founder and co-funder of the Nuffield Council on Bioethics and the Nuffield Family Justice Observatory.

Find out more:

Website: Adalovelaceinstitute.org
Bluesky: @adalovelaceinst.bsky.social
LinkedIn: Ada Lovelace Institute
Email: hello@adalovelaceinstitute.org

About the Center for Democracy & Technology

The Center for Democracy & Technology (CDT) is the leading nonpartisan, nonprofit organisation fighting to advance civil rights and civil liberties in the digital age.

We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the earliest days of the internet. The organisation is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.

CDT's AI Governance Lab develops and promotes adoption of robust, technically informed solutions for the effective regulation and governance of AI systems. The Lab provides public interest expertise in rapidly developing policy and technical conversations, to advance the interests of individuals whose lives and rights are impacted by AI.

Website: cdt.org/
Bluesky: @cdt.org
LinkedIn: Center for Democracy & Technology

Preferred citation: Lara Groves, Amy Winecoff and Miranda Bogen, *Going pro? Considerations for the emerging field of AI assurance* (Ada Lovelace Institute and Center for Democracy & Technology, 2025) https://www.adalovelaceinstitute.org/report/going-pro/

ISBN: 978-1-0684261-1-7