IMPERIAL
Centre for Sectoral Economic Performance

The Alan Turing Institute

CETaS Centre for Emerging Technology and Security
BRIEFING PAPER

Growing the UK's AI Assurance Market in Defence and Security

Natasha Karner
Anna Knack
Rupert Shute
Carolyn Ashurst

January 2026

Table of Contents

About CETaS

The Centre for Emerging Technology and Security (CETaS) is a research centre based at The Alan Turing Institute, the UK's national institute for data science and artificial intelligence. The Centre's mission is to inform UK security policy through evidence-based, interdisciplinary research on emerging technology issues. Connect with CETaS at cetas.turing.ac.uk.

This research was supported by the Centre for Sectoral Economic Performance (CSEP) at Imperial College.

About CSEP

The Centre for Sectoral Economic Performance (CSEP) is investigating ways of improving the competitiveness of the UK economy and driving economic growth. There are huge opportunities for the UK's key science and technology-based industries to generate real growth, but they also face challenges from rising global competition and counterproductive policy directions. We are researching the factors that will ensure they achieve their potential for global success and engaging with industry and government to develop actionable recommendations.

Acknowledgements

The authors are grateful to the interviewees for their time and perspectives, and for Rosamund Powell's contributions to the scoping, initial input and review of this project.

Cite this work as: Natasha Karner et al., "Growing the UK's AI Assurance Market in Defence and Security," *CETaS Briefing Papers* (January 2026).

Executive Summary

This CETaS Briefing Paper provides an evidence-based analysis of the UK's AI assurance market for Defence and National Security (D&S). A thriving AI assurance sector could enable AI adoption and become a key driver of UK economic growth. If organisations are confident that AI harms can be mitigated, this will help the UK Government achieve its aim of "fast, wide and safe" adoption of AI[1] and prevent AI capabilities from failing at the implementation stage. Effective assurance processes allow for the rapid integration of AI into existing business structures and processes, which can contribute to economic growth across multiple sectors. However, a range of factors currently limit both the supply of and demand for AI assurance services.

This Briefing Paper describes the current state of AI assurance in national security, defence and policing organisations, highlighting its strengths, challenges and possible mitigations. Drawing on this, the paper identifies lessons from D&S to support the growth of a robust AI assurance market for other sectors and advance AI innovation across the UK economy.

Key findings from this study are as follows:

- D&S is a diverse sector in AI assurance maturity, as it includes both early adopters and organisations at the start of their AI assurance journeys. This is due to a range of factors that vary across the sector, including: level of AI adoption; technical skills; infrastructure and testing capabilities; risk appetite; preference for in-house or external offerings; and level of engagement with external providers of AI assurance.
- Demand for AI assurance in D&S is driven by a desire to secure strategic and operational advantage from effective AI, the risk of high-consequence errors, policy requirements, and a need to assess AI providers' claims. Demand for *third-party* assurance is driven by skills shortages in government organisations, a lack of resources, a desire for independent testing and potential price advantages.
- Supply of AI assurance in D&S is limited by information asymmetries, skills gaps, unclear regulatory guidance and a lack of long-term funding. Demand for AI assurance in D&S is constrained by confusion over assurance offerings, information-sharing barriers, cultural barriers, a lack of funding and slow procurement processes.
- D&S provides a case study with broader lessons for the UK as it works to bolster its AI assurance market. This includes the need to: articulate sector-specific requirements; cultivate a market that caters for different levels of AI assurance maturity; develop initiatives to upskill key stakeholders; create mechanisms to disseminate best practice; and establish certification schemes for AI assurance providers.

[1] HM Government, *AI Opportunities Action Plan* (Department for Science, Innovation and Technology: January 2025).

1. Introduction

One of the most critical barriers to widespread AI adoption across UK D&S has been a lack of sufficient in-house knowledge, personnel and infrastructure to conduct AI assurance at scale. As Rosamund Powell and Marion Oswald explain, "AI assurance" is the portfolio of processes required to evaluate and communicate – iteratively throughout the AI lifecycle – the extent to which a given AI system:

- Does everything it says it is going to do, and nothing it should not do.
- Complies with the values of the deploying organisation and upholds established ethical principles.
- Is legally compliant and appropriate to the specific deployment context.[2]

AI assurance providers vary in the products and services that they offer, though two main areas of activity have emerged: consulting, training and procedural services and tools to help develop assurance strategies and processes; and technical tools to assess AI tools from a legal, ethical, technical or regulatory point of view. Specialised AI assurance companies, diversified firms and AI developers offer these services.[3]

- The UK Government has recognised the potential of third-party AI assurance to:
- Ensure compliance with legal and regulatory requirements.
- Enable AI adoption in D&S.
- Boost economic growth.

As the third-party AI assurance industry grows, the spillover effects could generate economic growth by: expanding the UK's AI assurance skills and supplier base; raising the value of the AI assurance sector more broadly; increasing confidence in and adoption of AI solutions; and improving productivity and security through the successful adoption of AI across the economy and national infrastructure.

Just as the development of the cybersecurity sector[4] continues to create jobs across the UK, the growth of the AI assurance sector would supply the confidence needed to enable rapid digitisation.[5] Modelling has suggested that the global market for AI assurance technologies will reach $276bn by 2030.[6] AI assurance activities contributed an estimated £1.01bn to the UK economy in 2024,[7] and this market could reach £18.8bn gross value added by 2035 if barriers to AI adoption are addressed.[8]

A recent study has shown that while the UK is ahead of comparable countries on AI assurance, both supply and demand for AI assurance services continue to lag behind their growth potential.[9] As a result, the UK Department for Science, Innovation and Technology (DSIT) has identified efforts to grow the UK's third-party AI assurance sector as a key priority in the AI Opportunities Action Plan.[10]

To date, research into growing the AI assurance market has been largely sector-agnostic. A sector-specific approach is needed to fully understand the requirements and challenges in particular domains, and to identify best practice that can be shared with other areas. This Briefing Paper provides such analysis of AI assurance in UK D&S, identifying key drivers of assurance services and limits to supply and demand. It makes a series of recommendations designed to enable safe and effective AI adoption in D&S, and to offer lessons for other sectors in growing a robust and thriving AI assurance market.

[2] Rosamund Powell and Marion Oswald, "Assurance of Third-Party AI Systems for UK National Security," CETaS Research Reports (January 2024): 3.
[3] Sarah Snelson and Vladislava Bar-Katz, *Economic Assessment of the AI Assurance Market* (Frontier Economics: May 2024), 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/.
[4] Bristol University and Imperial College London, *A UK cyber growth action plan – final report*, (DSIT: September 2025), https://www.gov.uk/government/publications/cyber-growth-action-plan-2025/a-uk-cyber-growth-action-plan-final-report.
[5] HM Government, Cyber Security Sectoral Analysis, (Department for Science, Innovation and Technology: May 2023), https://www.gov.uk/government/publications/cyber-security-sectoral-analysis-2024/cyber-security-sectoral-analysis-2024.
[6] Jam Kraprayoon, "Assuring Growth: Making the UK a Global Leader in AI Assurance Technology", IAPS, 11 July 2024, https://www.iaps.ai/research/assuring-growth.
[7] Sarah Snelson and Vladislava Bar-Katz, *Economic Assessment of the AI Assurance Market* (Frontier Economics: May 2024), 4, 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/.
[8] HM Government, *Trusted Third-Party AI Assurance Roadmap* (Department for Science, Innovation and Technology: September 2025).
[9] Sarah Snelson and Vladislava Bar-Katz, *Economic Assessment of the AI Assurance Market* (Frontier Economics: May 2024), 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/.
[10] HM Government, *AI Opportunities Action Plan* (Department for Science, Innovation and Technology: January 2025).

Research aims and methodology

This study builds on earlier research exploring third-party assurance for UK national security[11] and AI assurance for Defence.[12] The data were collected between August and October 2025 through a literature review of government guidance on AI assurance and publicly available information on the UK's AI assurance market, along with 15 targeted interviews with D&S government representatives and industry providers. The study is intended to lay the groundwork for a larger-scale sectoral plan to grow the UK AI assurance market in D&S and beyond.

[11] Rosamund Powell and Marion Oswald, "Assurance of Third-Party AI Systems for UK National Security," *CETaS Research Reports* (January 2024).
[12] Anna Knack *et al.*, "Assuring AI-enabled uncrewed systems: identifying promising practice for defence," The Alan Turing Institute Report (September 2025).

2. AI Assurance Approaches in Defence and Security

In this section, we provide a brief snapshot of AI assurance across D&S and the range of options for using third-party offerings. We show that AI assurance maturity varies across D&S and policing organisations. While there are some common factors in approaches to AI assurance, we have identified significant differences across and within these sectors.

The maturity of AI assurance varies greatly across D&S, ranging from early adopters to those starting to think about what AI assurance means for them. The factors contributing to this are differing levels of in-house skills, infrastructure and testing capabilities, as well as broader AI maturity. The level and role of external offerings also vary. Organisations with significant in-house expertise are more likely to rely on in-house processes and testing. Those with less technical expertise are more likely to use contractors or external offerings to perform the bulk of system testing.

For example, the police outsourced independent testing on accuracy and bias for the rollout of live facial recognition (LFR) to the National Physical Laboratory, which subcontracted a private company.[13] Conversely, an organisation in the United Kingdom Intelligence Community (UKIC) has developed a well-established set of practices to assist review processes, provenance and audits in-house, only using external tools for testing certain specialist tasks.[14] The Defence Artificial Intelligence Centre (DAIC) is working with the Front Line Commands and their Responsible AI Senior Officers to determine the balance between in-house and third-party assurance.[15] Although D&S includes some first movers on AI assurance, acceptance of the third-party AI assurance market varies across the sector.

2.1 National Security

One interviewee argued that AI assurance was "generally done well" in the national security sector, having rapidly developed in the last five years. This includes efforts to standardise assurance approaches by providing schemas of information and question sets for model deployment, by using model cards and building ethics panels.[16] Since security organisations were first movers in AI adoption, they chose to invest in building in-house assurance capacity because they had the requisite skills to do so, at a time when the external assurance market was in its infancy.[17] Although third-party offerings have emerged since then, assurance processes are still largely conducted in-house with limited use of external tools for specialised tasks. GCHQ has exported some of its own assurance tools to a wider community – by, for example, open-sourcing Bailo, which acts as a central repository for models and model cards.[18] For each model card, a two-stage review process (a technical assessment and a policy assessment) helps manage the AI lifecycle and any compliance requirements. One interviewee described the process of documenting this information as "incredibly helpful" in a highly regulated environment.[19]

2.2 Defence

Interviewees describe Defence as having "pockets of excellence"[20] for assessing safety, performance and the legal review of technical capabilities in general, but AI assurance capabilities that are "nascent,"[21] "not mature"[22] and "in an exploratory phase."[23] The large scale of Defence (relative to National Security) can create fragmented governance approaches. Moreover, Defence suffers from shortages of in-house technical expertise and limitations on its secure computing and data infrastructure.

Attitudes towards AI assurance vary across Defence. One interviewee described assurance as perceived to be "standing still and thinking about the problem [and] isn't

[13] Home Office, "Live Facial Recognition technology to catch high-harm offenders," 13 August 2025, https://www.gov.uk/government/news/live-facial-recognition-technology-to-catch-high-harm-offenders.
[14] Interview with government participant, 7 October 2025.
[15] Interview with government participant, 8 September 2025.
[16] Interview with government participant, 7 October 2025.
[17] Interview with government participant, 14 October 2025.
[18] GCHQ/Bailo, "Bailo – managing the lifecycle of machine learning to support scalability, impact, collaboration, compliance and sharing," GitHub, https://github.com/gchq/Bailo.
[19] Interview with government participant, 7 October 2025.
[20] Interview with government participant, 4 September 2025.
[21] Interview with public sector focus group, 15 September 2025.
[22] Interview with government participant, 4 September 2025; Interview with government participant, 8 September 2025.
[23] Interview with government participant, 8 September 2025.

seen as valuable as running toward delivery."[24] However, the Ministry of Defence (MoD) and DAIC have appointed Responsible AI Senior Officers to help the various area-level budget holders develop bespoke assurance approaches based on AI policy guidelines (JSP 936), as well as an AI Practitioner's Handbook. DAIC also set up a Model Arena asking suppliers to competitively test and demonstrate their models for Defence use cases.[25] A mix of in-house and external services are being used by Defence, including an in-house Experimentation and Trials Unit that the MoD has adopted for tests in realistic environments.[26] Several interviewees described how third-party offerings were "critical" to Defence, which lacks the skills and resources to conduct assurance fully in-house.[27] One interviewee described an ambition to develop a model in which the Government sets the criteria and thresholds of assurance, contractors and other external providers assure to those standards, and the Government then validates the system using in-house or third-party testing.[28]

Defence has granted third parties access to certain facilities via Long-Term Partnering Agreements. For example, QinetiQ runs testing and evaluation on equipment and capabilities at 16 MoD sites, leading to Europe's first successful demonstration of teaming a piloted aircraft with an autonomous aerial vehicle.[29]

2.3 Policing

Levels of AI assurance maturity vary across policing, with many areas still at an early stage. Some policing organisations are in the process of establishing an AI advisory panel to assist their assurance leads and plan to trial an interim assurance process.[30] This process will determine if there is a need for internal upskilling or external expertise to grow assurance. As mentioned, policing organisations have experience in using external offerings for routine testing. For example, they have commissioned tests to measure the accuracy of LFR systems under certain operational conditions.[31] There are also plans to establish a national centre for AI and policing, complete with a lab element to manage testing and assurance alongside R&D.[32]

However, AI assurance has not yet been standardised across policing organisations. This has led to some fragmentation and differing risk appetites across forces.[33] There is also a concern about vendor lock-in with certain tech companies: the sensitivity of operations means they are unable to communicate their needs to the entire market, so revert to what one interviewee calls an "ad-hoc" [34] approach that quickly engages with the same handful of companies that are already vetted and have a track record of working across D&S.[35] Policing organisations are concerned not just about AI adoption in a general sense but also about the use of vendors for assurance activities.

2.4 Factors that contribute to AI assurance approaches in Defence and Security

All D&S organisations face common operating conditions for AI assurance – such as high-stakes and highly regulated environments – but, as discussed, they vary significantly in areas such as internal skills and the quality of data infrastructure. Here, we describe those common factors that have implications for AI assurance, before highlighting relevant differences.

2.4.1 Common factors

Factors that are common across D&S include: the high-stakes nature of the domain; the critical need to deploy systems in safe and explainable ways; a highly regulated landscape; information siloes; and the need for assurance to cater to different use cases and levels of risk. Interviewees stressed that reliable and robust AI was essential to safety-critical sectors, as AI failures could compromise operations and lead to miscarriages of justice, loss of life and an irreversible loss of public trust.[36] Furthermore, the highly regulated nature of their sectors points to a strong need for assurance and processes that can stand up to the necessary scrutiny. Engaging with external vendors is challenging due to the classified nature of data, capabilities and operations. This creates difficulties when trying to communicate requirements and share data, potentially restricting who can work on internal systems. Many organisations are unsure of how to work with external partners effectively.

Interviewees referred to a core dilemma in how to balance the pace of AI adoption with efforts to meet the robust requirements of regulation, legal review and the criminal justice system. They consistently expressed a desire for agile AI assurance offerings that can cater to different risk tiers and thereby drive innovation. Yet a pain point in some areas of the sector is underdeveloped mapping of levels of risks, which leads to confusion about what level of AI assurance would be most appropriate.[37] Finally, interviewees said that information siloes in Government left D&S organisations unaware of AI developments and assurance best practice among their peers, leaving them unable to benefit from existing efforts.[38]

2.4.2 Differing factors

Key differences across D&S organisations include their approaches to building end-to-end assurance processes, access to in-house technical skills, levels of AI adoption, quality of data infrastructure and risk appetites. There are also nuances within skill sets. For instance, some national security organisations have high levels of AI skill. Defence has deep, historical experience with technology and testing in high-risk environments, suggesting that it has great *potential* in AI assurance. However, Defence's AI expertise is still developing.

While some organisations prefer to internally develop and own their assurance process,[39] others are comfortable with adopting external assurance designs, especially for low-risk AI applications.[40] Some prefer in-house assurance to avoid the time-consuming vetting of new companies and to leverage what one interviewee calls the "free resources" of existing internal processes.[41] Finally, policing has a relatively risk-averse approach to AI adoption and assurance,[42] whereas Defence's risk appetite is influenced by broader geopolitical considerations, such as the strategic need for advantage in wartime conditions.[43]

2.5 Summary of Approaches to AI Assurance across Defence and Security

Table 1 summarises different approaches to AI Assurance across D&S and outlines the key benefits and limitations of each approach.

[24] Interview with government participant, 4 September 2025.
[25] Ministry of Defence, *Launching the AI Model Arena* (MOD: 10 November 2025), https://www.gov.uk/government/news/launching-the-ai-model-arena.
[26] Interview with government participant, 3 September 2025.
[27] Interview with public sector participant, 8 October 2025; Interview with government participant, 4 September 2025; Focus group with public sector, 15 September 2025.
[28] Interview with government participant, 7 October 2025.
[29] Interview with government participant, 4 September 2025; "Major £1.5 billion defence contract with British firm ensures world-class equipment testing for UK forces and secures 1,200 jobs," (Ministry of Defence: 11 May 2025), https://www.gov.uk/government/news/major-15-billion-defence-contract-with-british-firm-ensures-world-class-equipment-testing-for-uk-forces-and-secures-1200-jobs.
[30] Interview with government participant, 15 October 2025.
[31] See the original report here: Dr Tony Mansfield, *Facial Recognition Technology in Law Enforcement: Equitability Study*, (National Physical Laboratory: March 2023), https://science.police.uk/site/assets/files/3396/frt-equitability-study_mar2023.pdf.
[32] Interview with government participant, 14 October 2025.
[33] Interview with government participant, 15 October 2025; Interview with government participant (2), 15 October 2025.
[34] Interview with government participant, 14 October 2025.
[35] Interview with government participant, 15 October 2025.
[36] Interview with government participant, 14 October 2025.
[37] Interview with government participant, 14 October 2025.
[38] Interview with government participant, 14 October 2025; Interview with government participant, 15 October 2025; Interview with government participant (2), 15 October 2025.
[39] Interview with government participant, 15 October 2025.
[40] Interview with government participant, 7 October 2025.
[41] Interview with government participant, 15 October 2025; Interview with government participant (2), 15 October 2025.
[42] Interview with government participant, 15 October 2025; Interview with government participant (2), 15 October 2025.
[43] Interview with government participant, 4 September 2025.

Table 1. Current and potential approaches to AI assurance across Defence and Security

Approach	Examples	Potential benefits	Potential limitations
Establish public AI assurance testing and evaluation labs	MITRE's AI Sandbox and AI Assurance Lab[44] NATO AI Labs[45]	Ability to conduct assurance quickly, with D&S-specific personnel readily available and developing expertise on D&S use cases. Different government organisations could rely on these labs to complete the bulk of testing, complementing this with their own additional testing where they have bespoke requirements.[46]	If the lab cannot meet demand, it can become a bottleneck. If national AI labs do not explicitly focus on D&S AI use cases, they may be limited to technical assurance of individual commercial-oriented models and miss integration risks.
In-house assurance by government personnel	Core UKIC processes[47] Defence approach of appointing Responsible AI Senior Officers helps develop tailored assurance approaches according to users' needs, with regulations and policy guiding compliance requirements. Internal model and system cards[48]	Can speed up assurance once established. Assurance approach can be tailored to the organisational context of each service and use case. Helps organisations build up internal expertise in the strengths and limitations of AI systems.	Resource-intensive to establish. Potential reinforcement of siloes of knowledge. Regulation and policy that have common compliance criteria for all types of AI may miss some risks or place excessive constraints on relatively benign use cases.[49]
Third-party assurance	Commercial industry practice and emergence of third-party AI assurance providers.[50] Outsourcing testing for LFR in policing	Available immediately, especially for dual-use models,[51] and would not need to be delayed by a need to build in-house infrastructure.[52] Costs and risks of training, hiring and retaining personnel would be borne by industry instead of Government.[53] Diverse assurance methodologies may increase AI system robustness. Market competition will stimulate continuous AI assurance innovation.	Lack of certification mechanisms may make it unclear which third-party AI assurance providers are providing consistent and reliable advice aligned with D&S requirements. Lack of consistency. Skills are not developed in-house. Approach may be less tailored to the organisation.
Mixed in-house and third-party assurance approach	Validation model in which Government sets criteria and thresholds; industry builds and assures to those standards; and government validates the system against the criteria (using in-house or third-party testing).[54] Developing a tiered system of risks specific to sectors and use cases. Largely in-house assurance processes complemented by niche testing products developed within industry.	Many benefits of third-party assurance retained, such as market competition, outsourcing risks[55] and the development of diverse methodologies. Stimulates continuous AI assurance innovation. Could support parts of D&S that lack technical skill or are relatively large.[56] Could allow for adaptable procurement process, preventing overspend on external providers and overinvestment in in-house capacity when the risk is low.[57]	Practitioners may not trust external assurance providers.
Model testing and demonstration competitions	DAIC's Model Arena[58]	Could enable triage of best models by asking developers to compete in testing demonstrations. Could stimulate AI testing innovation.	Challenging to test AI integrated into platforms or systems of systems in this way.

[44] "MITRE to Establish New AI Experimentation and Prototyping Capability for US Government Agencies," MITRE, 7 May 2024, https://www.mitre.org/news-insights/news-release/mitre-establish-new-ai-experimentation-and-prototyping-capability-us.
[45] "AI Laboratory," NATO Strategic Communications Centre of Excellence, n.d., https://stratcomcoe.org/projects/ai-laboratory/5.
[46] Interview with government participant, 14 October 2025.
[47] Interview with government participant, 7 October 2025; Interview with industry participant, 8 September 2025.
[48] GCHQ/Bailo, "Bailo – managing the lifecycle of machine learning to support scalability, impact, collaboration, compliance and sharing," GitHub, https://github.com/gchq/Bailo; Anna Knack et al., "Assuring AI-enabled uncrewed systems: identifying promising practice for defence," The Alan Turing Institute Report (September 2025); Advai "Assurance-Engineered AI Adoption Playbook," May 2024, https://www.advai.co.uk/ai-adoption-playbook/; Hugging Face, "Model Cards," https://huggingface.co/docs/hub/model-cards.
[49] Interview with industry participant, 8 September 2025.
[50] Sarah Snelson and Vladislava Bar-Katz, *Economic Assessment of the AI Assurance Market* (Frontier Economics: May 2024), 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/.
[51] Focus group with industry, 15 September 2025.
[52] Interview with industry participant, 6 October 2025
[53] Interview with industry participant, 6 October 2025
[54] Interview with government participant, 7 October 2025.
[55] Interview with industry participant, 6 October 2025
[56] Interview with government participant, 16 September 2025; Focus group with industry, 15 September 2025.
[57] Interview with government participant, 14 October 2025.
[58] Focus group with industry, 15 September 2025.

3. Drivers of Demand for AI Assurance

This section outlines the key drivers for AI assurance in D&S, followed by drivers of external, *third-party* AI assurance specifically. The former include:

- **Growth in demand for AI applications:**[59] increasing numbers of AI applications provide new opportunities to transform D&S organisations. In policing, AI is viewed as potentially capable of driving reform by increasing the efficiency of routine tasks and contributing to better criminal justice outcomes.[60] In Defence, increased demand comes from the geopolitical context and the potential strategic advantage to be gained from AI systems such as those for battlefield command decision-making and uncrewed systems.[61] AI applications that triage and analyse data to aid intelligence analysts are seen as offering a speed advantage over adversaries.[62] As AI ambitions grow in D&S, assurance is needed to ensure that AI systems provide the envisaged benefits.

- **Policy and compliance:** legislation and policy can mandate assurance requirements for AI use. AI assurance can help demonstrate compliance with these requirements – which, in turn, can indirectly promote assurance by driving AI demand. For example, interviewees said that the MoD's JSP 936: Dependable AI in Defence has been useful in "anchoring and orientating" conversations about AI assurance with the MoD and the Army, even if further granularity is needed.[63]

- **Unique properties of AI:** unlike traditional software, some types of AI may be probabilistic or have an opaque or 'black box' nature that requires new or adapted assurance processes. Model drift,[64] distribution shift[65] and software updates can require processes that allow for continuous monitoring and testing.[66]

- **Potential high-consequence errors:** errors in D&S contexts can lead to reputational damage, loss of public trust, miscarriages of justice and a loss of life.[67]

- **Accountability and liability:** interviewees see assurance as addressing concerns about accountability, liability and reputational damage, helping senior decision-makers understand AI systems' limitations before deployment.[68] D&S leaders are used to making decisions under uncertainty but, as one interviewee noted, "on the battlefield, we have to take huge risks, but we have a process to deal with that. I don't know how we apply that to models."[69] Assurance can contribute to a better understanding of the limitations and risks of systems, and can help ensure that due process is followed.

- **Distinguishing between AI systems' true capabilities and AI marketing hype:** although there are many open-source benchmarks and frameworks for AI testing, there is no authoritative, independent organisation that can help users distinguish between, as one interviewee puts it, "what you hope to get from AI and where your AI use case or application is at"[70] with the legitimacy to inspect and approve AI systems in D&S contexts. Robust assurance processes can help challenge or validate AI developers' claims.

- **Return on investment:** one interviewee noted a recent report from MIT that revealed that despite over $30bn–40bn in corporate spending, 95% of generative AI pilots failed due to breakdowns in AI integration.[71] Given resource constraints in D&S, there is a high evidential requirement in showing that investments in AI generate returns. Processes that normalise third-party AI assurance send a clear demand signal to industry that they will be asked for evidence of compliance with mission, legal and policy requirements.

[59] Interview with government participant, 3 September 2025.
[60] Interview with government participant, 14 October 2025.
[61] Interview with government participant, 4 September 2025.
[62] Interview with government participant, 3 September 2025.
[63] Interview with public sector participant, 8 October 2025.
[64] Model drift refers to the decline of a machine learning (ML) model's predictive performance due to changes in data or the relationship between input and output variables, resulting in faulty decision-making and bad predictions. See more here: Jim Holdsworth, Ivan Belcic and Cole Stryker, "What is model drift," IBM, n.d., https://www.ibm.com/think/topics/model-drift.
[65] Distribution drift is when the properties of the data on which a ML model was trained changes and the historical data no longer represents the current environment, resulting in inaccurate predictions. See more here: Rajan Adhikari, "Distribution Drift: A silent model killer in production," Medium, 17 November 2024, https://medium.com/@jugalraj/distribution-drift-db25ff80495e.
[66] Interview with public sector participant, 8 October 2025.
[67] Anna Knack *et al.*, "Assuring AI-enabled uncrewed systems: identifying promising practice for defence," The Alan Turing Institute Report (September 2025); Interview with academic participant, 1 September 2025; Interview with government participant, 14 October 2025.
[68] Focus group with public sector, 15 September 2025.
[69] Interview with government participant, 3 September 2025.
[70] Interview with international expert, 17 September 2025.
[71] Aditya Challapally et al., *State of AI in Business 2025* (MIT NANDA: 2025), https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf; Interview with public sector participant, 8 October 2025.

The study also identified the following drivers of demand unique to *external, third-party* AI assurance offerings:

- **Skills shortages:** a lack of in-house skills and talent across government D&S organisations drives demand for external AI assurance, as government teams may be unable to develop processes or conduct the necessary testing.[72] To quickly scale AI use in the near term, it will be critical to bring in third-party offerings – pending the long-term investment needed for hiring, training and retaining staff who can oversee the assurance case once AI systems have been acquired.[73]

- **Resource shortages:** many D&S organisations are limited by a lack of resources (money and time) to build and scale in-house assurance teams and capabilities.

- **Filling AI assurance capability gaps:** relatedly, third-party offerings may be capable of meeting different levels of needs in ways that in-house teams cannot. For example, narrow technical tools can be used for specialist testing – such as model scanning or automated detection of vulnerabilities – while other development teams could bring in contractors to help build in-house testing centres.[74]

- **Desire for independent testing:**[75] tools and processes developed in-house can benefit from external review and testing. Equally, interviewees see the claims that external vendors make about the safety of their products as an effort to mark their own homework.[76]

- **Potential price advantage:** some interviewees feel that outsourcing assurance via third parties is a more pragmatic, cost-effective and agile approach than building and maintaining expensive in-house AI assurance testing and simulation facilities, particularly in the short term.[77]

- **A variety of suppliers can test different AI assurance approaches:** growing the third-party AI assurance sector could allow D&S customers to learn about different assurance methodologies and improve their best practice.

[72] Interview with government participant, 4 September 2025.
[73] Interview with public sector participant, 8 October 2025.
[74] Interview with government participant, 7 October 2025.
[75] Interview with government participant, 3 September 2025.
[76] Interview with government participant, 16 September 2025; Interview with government participant, 15 October 2025; Focus group with public sector, 15 September 2025.
[77] Interview with international participant, 17 September 2025.

4. Challenges to Growing the AI Assurance Market in Defence and Security

Despite clear demand for AI assurance, there are several challenges to overcome before third-party AI assurance will achieve broad adoption in D&S. This section outlines supply- and demand-side challenges.

4.1 Supply-side challenges

There are several constraints on the supply of high-quality AI assurance:

1. **Information asymmetries and misaligned demand signals:** AI assurance suppliers often have limited access to training data, operational use cases or information to help tailor their services to the needs of D&S. In addition, as interviewees noted, there is a mismatch between AI assurance service providers focused on frontier AI and the need to assure a wider range of AI systems.[78]

2. **Interpreting guidance:** high-level guidance that lacks specific requirements for assurance providers makes it difficult for companies to determine what criteria they are marking AI systems against. While "safety" and "responsible AI" are broad terms, different use cases will have different minimum requirements.[79] This is particularly relevant for D&S sectors that operate under different risk thresholds depending on the need to deploy AI systems (such as during a terrorist attack or the outbreak of a conflict) and the level of risk (such as in monitoring online content or using kinetic force).

3. **Difficulties in demonstrating the quality of assurance offerings:** a lack of standardisation and certification processes can make it challenging for companies to demonstrate the quality of their services or ensure their employees have the requisite skills and qualifications.[80] Conversely, accreditation bodies may struggle to acquire the domain knowledge necessary to certify third-party assurance.[81]

4. **Risk exposure**: the makeup of the assurance market may be limited, as companies of varying sizes differ in their ability to take on risk. A small startup may lack the capacity to take on D&S projects that are high-risk, prevent it from retaining the IP for a tool it develops, or involve uncertainty about funding or a return on investment required to sustain the business.[82]

5. **Lack of funding and short-term contracts:** external providers are not incentivised to make AI assurance offerings if they are unsure whether a project can survive its pilot stage and scale.

6. **Skills and talent shortages:** AI assurance requires a mix of expertise, including the technical skill to understand models and performance, and the legal and ethical knowledge to evaluate compliance and domain-specific information (such as that in the defence sector).

4.2 Demand-side challenges

This study also identified the following key challenges to demand for AI-assurance:

1. **Absence of certification and AI supplier standards:** this has led to a fragmented approach to assurance in practice.[83] Government organisations struggle to judge whether companies provide high-quality offerings and to identify which types of assurance they should procure, often reverting to adaptation of in-house processes to accelerate adoption.

2. **Low appreciation of AI assurance:** a lack of organisational knowledge about AI risks and vulnerabilities can lead to misperceptions about the need for and benefits of AI assurance. There may be varied perceptions of assurance as a barrier to innovation or an onerous "box-ticking" compliance exercise – or an assumption that all procured systems are already assured. [84]

3. **Lack of funding and slow procurement processes:** there is a growing gap between the hopes of leveraging AI in D&S and the funding and resources required to implement AI assurance. For example, much of the funding for autonomy in UK Defence will not be received until 2027, and procurement processes will not keep pace with AI developments. As one interviewee pointed out, "you can't develop and buy models the same way as tanks: you'll get one model a decade, which is too slow."[85] These factors make it harder for D&S to send strong demand signals to industry for corresponding assurance offerings.[86]

4. **Confusion over AI assurance offerings:** across our interviews, it was clear that assurance means "different things to different people, ranging from narrow technical testing, developing compliance processes, or more holistic definitions."[87] In practice, assurance may require several mechanisms working in tandem and may utilise personnel across internal teams, third-party auditors and regulators.[88] Many organisations lack mechanisms to determine which assurance technique suits their needs or which provider offers the best solution.

5. **Skills gaps:** there is a lack of knowledge necessary to evaluate the AI assurance offerings that are available. Government buyers may not receive adequate technical advice and struggle to articulate the "what" and "how" of their AI assurance needs. Senior decision-makers and specialists in procurement may be unaware of assurance evidence they can ask their suppliers to produce.

6. **Infrastructure shortages:** some organisations lack the infrastructure necessary for adequate and continuous AI testing, particularly in sufficiently representative environments.[89]

7. **Information sharing barriers:** information sharing is particularly difficult in D&S due to the prevalence of highly confidential data and other forms of information. Organisations in the sector may need to work with List X contractors[90] or personnel who have security clearances – which narrows their options for suppliers.

8. **Lack of trust and cultural factors:** some in D&S view third-party products (both AI systems and AI assurance offerings) as requiring more risk management than those developed in-house. Their concerns relate to supply chains, data provenance, system robustness and the sites that host government data.[91] One interviewee reflected that an assurance process in which all developments happened "outside of Crown control" would not be desirable or trustworthy.

[78] Interview with public sector focus group, 15 September 2025.
[79] Interview with government participant, 8 October 2025.
[80] HM Government, *Trusted Third-Party AI Assurance Roadmap* (Department for Science, Innovation and Technology: September 2025): 7.
[81] Interview with public sector participant, 8 September 2025.
[82] Interview with public sector participant, 2 September 2025.
[83] Interview with government participant, 15 October 2025.
[84] Interview with public sector participant, 2 September 2025.
[85] Interview with government participant, 3 September 2025.
[86] Interview with government participant, 3 September 2025.
[87] Interview with academic participant, 1 September 2025.
[88] Interview with academic participant, 1 September 2025.
[89] Interview with government participant, 4 September 2025.
[90] "Security requirements for List X contractors," (Cabinet Office, National Security and Intelligence, and Government Security Profession: April 2014), https://www.gov.uk/government/publications/security-requirements-for-list-x-contractors.
[91] Interview with public sector focus group, 15 September 2025; Interview with government participant, 7 October 2025.

5. Lessons for Accelerating AI Assurance Adoption across the UK Economy

Given the challenges described above, there are several lessons from experiences in D&S that could support the growth of a thriving and robust AI assurance market across the UK economy. These include:

- **Understanding sector-specific requirements:** to ensure policy interventions and industry offerings meet the unique requirements of certain sectors, it is important to consider the level of AI maturity, technical skills, attitudes to risk, the regulatory environment and unique requirements such as clearances for D&S. Government interventions and consortiums such as DSIT's "consortium of stakeholders"[92] should reflect these variances. For example, they should seek consortium representatives from different industries or establish different implementing consortiums for different sectors.

- **Upskilling key stakeholders:** senior decision-makers need to understand what assurance is needed to set organisational direction and funds. Likewise, upskilling those involved in the procurement of AI and assurance services would empower them to know what to look for. Mechanisms to upskill industry would also ensure that external companies (such as startups) understood the requirements and realities of D&S.

- **Developing solutions to information asymmetries:** this can involve establishing dialogue channels, and sharing requirements and datasets between sector representatives and representatives from the assurance sector.[93]

- **Centralised support and the dissemination of best practice:** when those at the forefront of different areas of a sector adopt the same best practice, this can help address duplication of effort and disparities in maturity. To reduce fragmentation across sectors – particularly among government bodies – a central assurance hub or 'Front Door' resource can assist organisations in determining what assurance is needed and what ecosystem exists to service this.[94]

- **Refining policy for implementation:** high-level guidance (such as JSP 936: Dependable AI in Defence) has driven demand by indicating the direction of travel.[95] In addition, granular benchmarks enhance industry confidence, as they can demonstrate the quality of assurance offerings. Sectors other than D&S should set high-level requirements to motivate action, and should then translate them into granular benchmarks.

- **Certifying AI assurance providers:** sectors should support current efforts by DSIT to develop a professional certification scheme for AI assurance.[96] This approach should account for the needs of specific sectors. Accreditation bodies should also have sufficient domain knowledge to asses the assurers.[97] Several interviewees also noted the role of establishing standards (as seen in early testing and evaluation of aviation),[98] which could be addressed by bodies such as the British Standards Institution and the UK Accreditation Service.

Table 2 summarises the lessons from D&S and offers possible policy interventions.

[92]HM Government, *Trusted Third-Party AI Assurance Roadmap* (Department for Science, Innovation and Technology: September 2025).
[93]Interview with government participant, 7 October 2025.
[94]Interview with government participant, 14 October 2025.
[95]Interview with public sector participant, 8 October 2025.
[96]HM Government, *Trusted Third-Party AI Assurance Roadmap* (Department for Science, Innovation and Technology: September 2025).
[97]Interview with industry participant, 8 September 025
[98]Interview with public sector focus group, 15 September 2025; Interview with public sector participant, 8 October 2025.

Table 2. Lessons from Defence and Security on AI assurance

	Challenge to growing the D&S AI Assurance Market	Possible policy interventions
Supply side	Information asymmetries and D&S requirements	Provide public information on AI assurance requirements.[99] Establish public–private dialogue channels.[100] Replicate current processes (such as the AI Verify Foundation's assurance of general-purpose models)[101] to focus on specific sectors or specific technology or assurance problems.
	Interpreting guidance	Developing simplified worksheets, with evidence required of suppliers to prove the model is assured. Introduce training for SMEs on AI assurance requirements.
	Risk exposure and short-term contracts	Multi-year public–private partnerships.[102] Subsidies for implementation of assurance.[103]
	Skills shortages	Developing a skills and competencies framework for AI assurance.[104]
Demand side	Lack of trust	Certification, codes of conduct and/or registration of AI assurance providers.[105] Trial third-party AI assurance providers for individual model testing in low-risk use cases. Bespoke demonstrations by third-party AI assurance providers that focus on D&S use cases.[106]
	Lack of appreciation of AI assurance standards	Communicate value of AI assurance to end users, senior responsible owners, venture capital investors and AI insurance providers.[107] Develop shareable materials illustrating types of AI assurance offerings available in the market and in-house, to illustrate how in-house AI assurance compares to the market standard.[108] Wargame After Action Reviews with and without AI assurance.

[99]Interview with industry participant, 1 September 2025.
[100]Interview with government participant, 7 October 2025.
[101]"Global Assurance Sandbox," AI Verify Foundation, 7 July 2025, https://aiverifyfoundation.sg/ai-assurance/.
[102]Interview with public sector participant, 8 October 2025; Interview with public sector participant, 17 September 2025.
[103]Interview with academic participant, 1 September 2025.
[104]HM Government, *Trusted Third-Party AI Assurance Roadmap* (Department for Science, Innovation and Technology: September 2025), 17.
[105]Interview with industry participant, 8 September 2025; Interview with industry participant, 6 October 2025.
[106]Interview with government participant, 8 September 2025.
[107]Interview with academic participant, 1 September 2025.
[108]Interview with government participant, 3 September 2025.

	Lack of funding	Coordinating across the sector to pool resources. Accessing new funds such as the DSIT AI Assurance Innovation Fund (£11m).[109]
	Slow procurement processes	Use innovation accelerator initiatives such as the UK Defence and Security Accelerator to focus on AI assurance.[110]
	Confusion over AI assurance offerings	Establish sandboxes[111] and demonstration competitions, ensuring end users and senior responsible owners are exposed to AI testing demonstrations – thereby increasing their AI assurance literacy. Commission external independent research to assess AI assurance offerings, including by gathering feedback from customers.[112] Provide open-source information on various AI assurance techniques.[113]

[109] This fund will issue £11m to support the development of "innovative and novel AI assurance mechanisms" with applications opening in Spring 2026. See more here: HM Government, *Trusted Third-Party AI Assurance Roadmap* (Department for Science, Innovation and Technology: September 2025).
[110] HM Government, "The Defence and Security Accelerator (DASA) finds and funds exploitable innovation for a safer future," https://www.gov.uk/government/organisations/defence-and-security-accelerator/about.
[111] Interview with academic participant, 1 September 2025.
[112] Interview with government participant, 4 September 2025.
[113] Interview with international participant, 17 September 2025.

6. Conclusion

A thriving and robust AI assurance sector will greatly enhance the UK's efforts to implement its AI Opportunities Action Plan and become a global leader in AI. Within D&S, there is strong recognition of the need for AI assurance and significant pockets of excellence where good practice is well established. Yet there is a need to ensure that this is expanded to other parts of the sector and beyond.

Addressing the supply- and demand-side challenges discussed in this paper would boost confidence in the market among both providers and buyers of AI assurance. Given that Defence organisations have large technology budgets, they will be dominant in shaping the demand signal. So, it will be especially important to prioritise efforts to match supply and demand for assurance here. Most significantly, external offerings could be improved by considering the unique characteristics of not only the sector as a whole but also different parts of the sector that have vastly different requirements. Overall, D&S organisations and the wider ecosystem need to think carefully about where third-party offerings are best placed to contribute, and where in-house expertise and capabilities should be built up.

This project has provided a snapshot of AI assurance across D&S. It has laid the groundwork for more substantial research to address the challenges identified, grow and access AI assurance across the UK, and accelerate effective AI adoption. Such research should also develop an action plan to drive growth in the D&S sector and beyond. This plan could set out key moves by government bodies, AI developers and AI assurers to operationalise the lessons of this paper. In this way, the UK could develop its assurance market, strengthen its position in the global AI landscape and ensure that it benefits economically and strategically from advancements in AI.

About the Authors

Natasha Karner is a Research Associate in the Turing's Defence and National Security Grand Challenge. She is based within CETaS and the AIDA defence policy workstream. Her recent and ongoing research projects include AI assurance in defence and national security, integrating AI into command decision-making, and national security and AI industry collaboration. Natasha's PhD dissertation explores the governance of Autonomous Weapons Systems (AWS), including deliberations under the auspices of the United Nations Convention on Certain Conventional Weapons (CCW). She is a Junior Associate Fellow at the NATO Defense College and a member of BASIC's Emerging Voices Network for nuclear weapons policy.

Anna Knack is a Senior Research Associate in the Turing's Defence and National Security Grand Challenge and Lead Researcher at the Centre for Emerging Technology and Security. She currently leads the AI for Data-Driven Advantage (AIDA) defence-focused policy research at the Turing, which focuses on identifying solutions to the technical and policy challenges Defence encounters as it drives towards AI adoption for strategic advantage. Her recent and ongoing research focuses on AI assurance in Defence, AI in intelligence analysis and decisions and AI for cyber defence.

Rupert Shute is Professor of Practice in emerging technology governance and regulation at Imperial College London. Prior to entering academia, he held senior roles in UK Government as the Forensic Science Regulator and Deputy Chief Scientific Adviser to the Home Office.

He is a Chartered Engineer and a Fellow of the Institution of Engineering & Technology. His technical career includes the development of robotics and Artificial Intelligence systems for high-consequence applications in defence and national security. He is active in AI research, standards setting and the startup ecosystem.

Dr Carolyn Ashurst is a Senior Research Fellow at The Alan Turing Institute, the UK's national institute for data science and AI. She is based within the Centre for Emerging Security and Technology (CETaS), where she leads the centre's work on Trust in AI for UK National Security.

Her research and engagement are motivated by the question: how do we ensure AI and other digital technologies are researched, developed and used responsibly? Her research interests include trustworthy AI, algorithmic fairness, transparency and responsible research practices. Carolyn sits on a range of advisory boards, including with OECD, the FBI and the Turing Research Ethics Process, and has worked with organisations such as PAI, DSIT, NICE and the ICO to convene multi-stakeholder events on a range of pressing topics.