techUK

techUK.org | @techUK

techUK response to Ofgem Call for Input: AI Assurance

August 2026 | techUK response

About techUK

techUK is a membership organisation launched in 2013 to champion the technology sector and prepare and empower the UK for what comes next, delivering a better future for people, society, the economy, and the planet.

It is the UK's leading technology membership organisation, with more than 1000 members spread across the UK. We are a network that enables our members to learn from each other and grow in a way which contributes to the country both socially and economically.

Question 1: Current AI assurance practices

> *a. How do organisations evidence that AI systems are operating as intended and delivering safe, fair and effective outcomes?*
> *b. What AI assurance approaches are currently used or under development, including in-house and third-party?*
> *c. What tools and technical capabilities are available to support AI assurance in practice, how mature and effective are they, and where are there gaps or opportunities for shared or sector-wide approaches?*
> *d. How are AI governance frameworks translated into operational practice?*
> *e. Which assurance or governance practices are most effective in supporting reliable outcomes?*
> *f. What skills, expertise and resources are required for effective AI assurance, and where are the main capability gaps?*

It is worth being clear at the outset what the UK Government's and industries working definition of "AI assurance" is: the process of measuring, evaluating and communicating the trustworthiness of AI systems, and of providing evidence of risk mitigation sufficient to build justified trust in them. Put simply, if regulation sets out what responsible AI looks like, assurance is how an organisation demonstrates it is meeting that in practice.

The UK has set an ambition to establish itself as a global leader in AI development and deployment, positioning the nation at the forefront of technological innovation. Central to this vision is the UK's pioneering work in AI assurance, developing the ecosystem designed to support justified trust in systems through evidenced based risk mitigation. This comprehensive approach to risk mitigation supports responsible AI adoption, ensures regulatory compliance, and facilitates access to capital for innovative ventures. The UK's AI assurance ecosystem is already taking shape. According to DSIT's "Assuring a Responsible Future for AI" (November 2024) around 524 firms are supplying AI assurance goods and services in the UK, generating an estimated £1.01bn in Gross Value Added and employing around 12,600 people, with the market potentially exceeding £6.5bn by 2035 if the barriers it identifies around demand, supply, and interoperability are addressed. These are exactly the barriers the AI Assurance Stakeholder Consortium, convened by DSIT and led by the British

Computer Society (BCS), has been set up to address; techUK leads the Consortium's Industry Advisory Group. The Consortium's membership also draws in bodies such as UKAS, BSI, NPL, the Chartered Quality Institute and the Ada Lovelace Institute. It is suggested that Ofgem and the energy sector should look to align with, and help inform, work of the Consortium particularly the skills and competency framework and code of ethics the Consortium is expected to produce, rather than developing a parallel energy-specific equivalent.

a. How do organisations evidence that AI systems are operating as intended and delivering safe, fair and effective outcomes?

Good governance of AI systems starts with organisations defining what trust means in their particular operational context, and building evidence, review and corrective action in from the design stage at the very start and throughout its lifecycle. This requires runtime governance to remain active during live operation, supported by continuous evidence generation rather than being confirmed once and assumed to hold. As AI systems become more operational, it could therefore be useful to distinguish three complementary forms of assurance:

- **Model Assurance:** provides confidence that an AI model performs appropriately under expected conditions.
- **Operational Assurance:** provides confidence that the deployed AI system continues to operate safely, reliably and in accordance with organisational governance throughout its operational lifecycle.
- **Execution Assurance:** provides confidence that AI-supported actions remain governed, observable, constrained and accountable throughout live operational execution.

Together, these extend assurance across the full 'Operational Execution Lifecycle' - from generation, through human review, to execution and audit – so that trust attaches to what the system actually does, not only to what the model recommends. These complementary assurance layers are strengthened when supported by common evidence relationships that provide traceability between governance objectives, operational controls, implementation artefacts, and assurance evidence throughout the operational lifecycle.

b. What AI assurance approaches are currently used or under development, including in-house and third-party?

Where the practice of AI assurance is strong within organisations (in-house assurance) this can mean the use of, or a combination of model cards, risk assessments, decision logs and ethics committees. Externally, independent audit is growing as a key AI assurance tool, including firms which have built frameworks to assure AI in critical energy infrastructure, and consultancy firms which now offer AI audits. Although it should be noted that organisations still lack a reliable way to judge which providers are credible - a possible barrier to increased use of the services already being provided, and something which is being looked at in the AI Assurance Consortium. It is also worth noting that the major Cloud Services and IT providers will have their own AI Governance and adoption frameworks that will provide advice, methodologies and tools, normally built-in that will support AI assurance.[1]

[1] For example, Amazon Web Services (AWS), has developed a version of its Cloud Adoption Framework methodology (CAF) that covers AI, Machine Learning and Generative AI.

c. What tools and technical capabilities are available to support AI assurance in practice, how mature and effective are they, and where are there gaps or opportunities for shared or sector-wide approaches?

While important work is underway, current AI assurance practices still lack maturity and right now are still patchily implemented. Where implementation is more mature, it can depend on individual Responsible AI leads piecing them together themselves, since no agreed template exists across the sector. It is also down to organisations who know what AI they use, including AI built into bought-in products, are aware of standards such as ISO/IEC 42001, ISO/IEC 23894 and The NIST Risk Management Framework (RMF)[2] and review systems on a fixed schedule, rather than just once. According to IAF CertSearch, ISO/IEC 42001 certification is accelerating rapidly with 134.31% growth in the first 6 months of 2026 with over 600 accredited certifications globally. However, it should be noted that, ISO 42001 certifies the organisation, not the product, but standards like EN 18286 are driving a product focused approach.

d. How are AI governance frameworks translated into operational practice?

We welcome the work that Ofgem has already done and believe this is guidance that assurance practice can build on is, specifically the AI Reg Lab and AI technical sandbox. techUK has previously recommended in our white paper 'How can the energy sector leverage AI to deliver consumer benefits?' that industry builds on Ofgem's Reg Lab to create a central hub for testing and verifying algorithms used within the energy sector, combining outcomes-based assurance with meaningful human oversight.

Translating governance into practice also depends on the underlying data infrastructure being fit for purpose. We understand that Ofgem has mandated that all network companies in Great Britain (GB) will have to adopt the Common Information Model (CIM) – with BSI acting as the governance body. The GB CIM has enormous potential for data sharing, and therefore AI, provided it is developed to maximise flexibility and systems-level interoperability.[3] Gas networks are pursuing a parallel voluntary CIM effort that has not yet been formally mandated by Ofgem. In its Sector Digitalisation Plan, NESO advocates the move towards a cross-sector approach within the GB CIM, with a "robust, ontologically grounded foundation model that is supported by smart assets". For this to work, ontologies need to be developed, applied and maintained across the energy system at a deeper level than these will solutions will enable.

f. What skills, expertise and resources are required for effective AI assurance, and where are the main capability gaps?

On skills, there is no clear AI assurance role, no career path, and no standard training framework, as techUK set out in our paper, Mapping the Responsible AI Profession. The forthcoming European standard EN 18274 on competence requirements for professional AI ethicists should help establish common expectations here, alongside the AI Assurance Consortium's skills and competency framework noted above, though both remain in development and do not yet solve the immediate skills gap facing network companies. In energy, people with the right skills are often drawn into traditional OT (Operational

[2] NIST RMF is a voluntary (non-binding and not a certification) framework published by the US National Institute of Standards and Technology on 26 January 2023. It helps organisations identify, assess, and manage AI-related risks across four core functions: Govern, Map, Measure, and Manage.

[3] It should be noted that one of the areas within the GB CIM – IEC 62325, which covers energy market processes (forecasting, bidding, settlement) – was developed for the TSO markets, so this would need to be adapted for distribution networks.

Technology) jobs instead, and this shortage keeps holding back adoption. Rather than the energy sector standing up its own parallel skills initiative, however, we would recommend it engage with and help shape the AI Assurance Stakeholder Consortium's skills and competency framework as it develops. For example, energy-specific training needs are better addressed by feeding sector context into that cross-sector effort than by creating a separate track alongside it.

Question 2: Risks and challenges

a. What are the main barriers to implementing effective AI assurance?
b. What are the key risks associated with AI use in the energy system (including system reliability, market functioning and consumer outcomes)?
c. Which risks are most difficult to assess, evidence, or link to real-world outcomes?
d. Where are current AI assurance approaches most limited in practice?

a. What are the main barriers to implementing effective AI assurance?

The main barrier to implementing effective AI assurance is that there are two ecosystems that need to mature together but are currently not fully interacting. One is the accountability infrastructure (procurement, investor due diligence, insurance, incident reporting and liability) and the other is professional infrastructure (skills, certification, and standards). Neither works effectively without the other, so they must be better integrated.

This immaturity in professional infrastructure shows up in the difficulty organisations face identifying trusted assurance providers — there is often no reliable way to assess the competence, independence, and quality of assurance providers, which limits confidence in assurance markets more broadly. This is not a one-sided problem, providers often struggle to be found by their customers as well, due to discoverability issues. Linked to this is the lack of incentive for AI developers and deployers to meaningfully engage with AI assurance. General demand from investors and consumers for more responsible AI exists but is rarely specific enough to drive high-quality assurance processes. Government has a role in translating this general demand into specific initiatives — through insurance (so insurers in turn require testing) and public procurement requirements that set a clear benchmark for what "good" looks like. Many stakeholders report that they simply do not know what Government expects of them here; clarifying accountability would help address this uncertainty. For example, techUK's own research into AI procurement, found that model cards are rarely requested by buyers, vendor disclosures are frequently inadequate, and liability for AI embedded within bought-in products remains largely unresolved (see "Buying blind: Rethinking AI procurement as a governance function", May 2026; the IEEE 3119 standard on AI procurement is also relevant here).

Organisations also face two further interconnected challenges: ensuring AI systems operate safely and reliably (assurance), and managing the financial risk when they do not (insurance). The two are intrinsically linked through the broader challenge of risk allocation in an increasingly automated world. Insurers already evaluate an organisation's assurance practices when determining coverage and pricing, which means assurance stops being only a technical necessity or a compliance exercise and becomes a genuine economic advantage: safety and security become directly legible in the bottom line, giving developers and deployers a commercial reason to invest in assurance beyond regulatory compliance.

Insurance can also strengthen the evidence base described elsewhere in this response: as a condition of cover, insurers are well placed to require structured incident reporting, which would help close the current gap in energy-sector data on how AI systems actually fail in practice, and give both industry and Ofgem a fuller picture of emerging risk than assurance claims or self-reported audits alone can provide. Government, Ofgem and industry should therefore explore how insurance requirements could be used deliberately to drive both better assurance practice and more consistent incident reporting across the sector, rather than treating insurance purely as a risk-transfer mechanism sitting outside the assurance and governance conversation.

A third barrier sits at the level of data infrastructure. Assurance tools and evidence are only as good as the data feeding them, yet that data layer has considerable gaps. The Data Sharing Infrastructure (DSI) will be a vital mechanism for sharing sensitive data but will not be fully rolled out until 2028. The Utilities Act still prevents network operators from sharing some data even under Ofgem's "Presumed Open" principle. There is also no systems-wide ontology or common language for describing energy assets, and the sector also lacks the broader semantic layer and knowledge management infrastructure needed to maintain and share that common understanding across organisations. This has been widely identified in other sectors and countries as an important enabler for successful and responsible AI deployment but the UK is lagging here.

b/c. What are the key risks associated with AI use in the energy system, and which are most difficult to assess, evidence, or link to real-world outcomes?

Energy regulation is largely outcome-based, but AI-assisted operational decisions introduce difficulty in evidencing how those outcomes were reached. The issue is not whether AI can be used, but how AI-informed decisions can be evidenced and explained against evolving expectations for auditability, fairness, safety and accountability. This matters most where operational choices must be made quickly, under constrained resources, with clear responsibility after the event.

Several risks flow from this. The hardest to evidence are those where compromised behaviour is indistinguishable from legitimate operation at the input/output level. For example, in a Proof of Technology conducted with Cisco through the UK's Laboratory for AI Security Research programme (with NCSC/GCHQ), attacks embedded in routine operational data were engineered to read as legitimate technical instructions, defeating keyword and blacklist-style filters entirely; only examining AI model-internal reasoning achieved 100% detection, including attack types never seen in training. In relation to energy, AI-assisted dispatch, balancing and forecasting systems ingest external feeds (market data, sensor telemetry, control signals) that could be compromised the same way, with the risk remaining invisible to output-only monitoring until it had already affected the network.

Agentic AI is now beginning to be trialled in decision-support roles adjacent to dispatch and balancing, yet current methods struggle to test and verify it or understand how it is functioning. As techUK's agentic AI brief sets out, this is partly because standard model benchmarking is not sufficient for agentic systems. Assurance depends on being able to test multi-step workflows, agent-to-agent handoffs and performance under changing conditions, and on observability, meaning visibility into the chain of decisions and actions throughout a workflow rather than only at its start and end points. Endpoint-only monitoring can miss errors, manipulation or unexpected behaviour at intermediate steps, particularly where agents act on

one another's outputs at speed, and many deployments would struggle to reconstruct a full decision trail if an auditor or regulator asked for one. The wider standards and assurance ecosystem is also not yet mature for these systems, since frameworks such as ISO/IEC 42001 were not designed with multi-agent coordination and autonomous operational workflows in mind. Once AI sits inside forecasting, optimisation or network control, risk cascades can spread across gas and electricity infrastructure.

d. Where are current AI assurance approaches most limited in practice?

The weakest point in the whole process can come after deployment, as assurance is often strong at proof-of-concept stage and weaker once a system runs live. Traditional assurance systems often assumed relatively static systems. Indeed, in DSIT's AI adoption research, businesses in agriculture, mining, manufacturing, energy and construction all named the physical safety of AI-run machinery as their primary concern.

One reason post-deployment assurance remains difficult is the distinction between decision time and execution time. An AI recommendation may be appropriate when generated but no longer appropriate by the time it is executed, since operational conditions, network topology, asset availability, demand, permissions, or cyber posture may have changed in the interim. This is particularly relevant in energy systems, where operating conditions evolve continuously and AI-supported decisions increasingly influence interconnected physical infrastructure.

Assurance should therefore validate not only the model and its original recommendation, but the execution context immediately before an action is applied, with the ability to re-evaluate, constrain, pause, or reject a proposed action where material changes have occurred since it was generated. This execution-time 'process' can, for example, provide explicit revalidation of whether a previously generated recommendation remains appropriate under the live operational conditions that exist immediately before action.

The humans overseeing, approving, or acting on AI outputs are also subject to change, through shifts in roles, workload, expertise, or trust in the system over time. Addressing these various issues can be achieved by considering how and what sign-off itself records. Sign-off of an AI system should capture: the scope of what was assessed, the period for which the assessment is expected to hold, and the material changes that would prompt reassessment — such as changes to deployment context, input data, oversight arrangements or system capability. Recording these at the point of sign-off could reduce the need to reconstruct assumptions after an incident, and gives both the organisation and the regulator a clearer basis for judging whether an earlier assurance claim still stands. A favourable result does not by itself demonstrate that the underlying process was safe, particularly where success may have depended on favourable circumstances or undocumented human intervention that may not recur. Assurance should therefore evaluate the governed process behind a decision, not only the outcome it happened to produce.

Question 3: Critical infrastructure considerations

> *a. How should AI assurance reflect the criticality of energy systems as national infrastructure?*
> *b. What level of rigour is appropriate for high-impact or safety-critical AI use cases?*

a. How should AI assurance reflect the criticality of energy systems as national infrastructure?

For AI used in or adjacent to critical energy operations, assurance should extend beyond whether the system performs as intended to whether operations remain safe when it does not. That requires evidence of a tested fallback or safe degraded mode; clear authority to override or withdraw the system, with response times proportionate to the operational risk; and arrangements for recovery and, where necessary, orderly withdrawal. Underpinning all of this should be a system of record for inference: every AI-supported decision should generate an immutable, queryable record of the data, model state and reasoning behind it, not just the final output; without this, replaying or reconstructing a significant operational event after the fact remains aspirational rather than achievable.

Rigour should scale with the criticality of the function the AI supports, with the operational authority, consequence, and reversibility of the action the system can influence or execute. An AI system providing advisory input, forecasting or decision-support presents a materially different assurance challenge from one capable of directly reconfiguring network assets, dispatch resources, alter operating parameters or initiate automated recovery actions. For the most critical uses require higher operational authority as well as stronger governance, evidence and intervention mechanisms, evidence of safe degradation and tested recovery.

This is consistent with early findings from techUK's ongoing work on AI assurance in critical national infrastructure, run with RAND and others, which point to a persistent accountability gap: it is often unclear who owns AI risk end-to-end in CNI, because safety functions sit in silos, and accountability can weaken the further an AI output moves from where it was built to where it becomes a real-world action. Closing that gap should be treated as a precondition for effective assurance, not a secondary governance issue.

Please note techUK held an expert roundtable on the subject in May, followed by a half day conference in June and are set to share a paper if of interest in November.

b. What level of rigour is appropriate for high-impact or safety-critical AI use cases?

CNI has an exceptionally low tolerance for hallucination or deployment failure. This makes generative-AI-derived agentic architectures poorly suited to genuinely critical core functions, with traditional, deterministic ML generally preferred for high-consequence work. Where a generative or agentic component is used in such a role, it should be expected to meet an equivalently high standard of predictability and control.

CNI also demands a systems-of-systems perspective: assurance should consider not only the behaviour of individual AI components but how AI-supported actions interact across interconnected operational, physical and organisational systems — including the humans who operate and oversee them — recognising that risk, authority and consequences may propagate well beyond the originating AI system.

For high-impact or safety-critical use cases describe in 3a, assurance should go beyond model performance and point-in-time testing to demonstrate that operational authority evidence of:

- clearly defined and technically enforced operational boundaries;
- independent validation of the execution context immediately before action;
- governance controls capable of allowing, constraining, pausing, denying or escalating actions;
- proportionate human intervention and override capabilities;
- comprehensive operational evidence supporting significant decisions;

- a queryable system of record for inference (as highlighted above in answer to question 3a);
- deterministic replay where technically feasible, or sufficiently detailed operational reconstruction where full deterministic replay is not possible;
- tested degraded-mode, fallback and safe-state behaviour; and
- reassessment following material changes to models, data, software, operational authority or operating conditions.

The objective should not be to demonstrate that failure is impossible, but that unsafe behaviour can be detected, contained, investigated, recovered from and continuously improved, with clear accountability maintained throughout.

Question 4: Consumer protection and fairness

> *a. How can AI assurance support fair treatment of consumers, including vulnerable groups?*
> *b. What risks arise from AI-driven pricing, segmentation or prioritisation, e.g. fairness, transparency, consumer outcomes?*

a. How can AI assurance support fair treatment of consumers, including vulnerable groups?

Using AI to help improve customer service and identify vulnerable energy customers is already delivering value, for example, mapping fuel-poverty risk to target Priority Services Register (PSR) support. However, this requires assurance of the underlying model's fairness and of the safeguards around sensitive inferences drawn from customer data. Fairness should not only be assessed at the point of deployment as consumer outcomes can change over time as models evolve, data quality shifts, customer behaviour changes or operational policies are updated. Assurance should therefore include ongoing monitoring for unintended bias, disproportionate impacts and changes in decision-making behaviour, particularly where systems influence pricing, prioritisation or access to services.

In the case of fuel poverty, it is equally important to ensure that fuel poverty not captured by these models is still identified and addressed through other means, and that there are no unwanted or unfair consequences arising from how the models classify people. For example, an asset-rich but cash-poor pensioner may be in genuine fuel poverty but not be identified by the model. On the other hand, individuals who are captured by the model may be profiled in ways that have other unintended consequences, such as being targeted for marketing of particular tariffs on the basis of their circumstances. Assurance thus needs to consider both under-inclusion (who is missed) and the downstream uses of the profiling itself (what happens to those who are captured). For higher-impact decisions of this kind, organisations should be able to demonstrate not just the outcome reached, but the governed process behind it: which data sources, policies and controls materially influenced the decision, and how it can be reviewed, challenged or independently verified. This does not require disclosing proprietary algorithms, but it does require enough visibility for organisations, regulators and consumers to ensure there is appropriate accountability.

b. What risks arise from AI-driven pricing, segmentation or prioritisation, e.g. fairness, transparency, consumer outcomes?

An area to consider is how dynamic and "smart" tariffs, though providing customers with savings on bills, can also raise transparency and consent questions, particularly where suppliers gain direct device control. Where AI systems are permitted to influence or directly control customer devices, safeguards should ensure that consent, operational boundaries and override mechanisms remain clear and effective for as long as the system operates, not only at the point consent was first given.

This is further complicated by the emergence of agentic AI. The SEC Privacy Sub-Committee has begun considering how agentic AI would fit within existing consent assurance processes, though this work has not yet been published.

Delegating authority to an AI agent to act on a consumer's behalf is arguably quite different from a consumer giving direct consent, and it is not yet clear what "transparency" means in practice when people configure agentic tools to pursue outcomes on their behalf — they may not see the relevant privacy notices, or even know which services the agent has selected and approved, raising a genuine question over whether consent given in this way can be considered valid. This mirrors the shift in oversight set out in techUK's industry brief on scaling the responsible adoption of agentic AI. As agents operate at greater speed and across more complex workflows, the traditional model in which a person checks every action evolves towards one where a person supervises and intervenes when needed, and oversight no longer means reviewing every individual action. Consent is under the same pressure. A single, point-in-time agreement is the consumer-facing equivalent of checking every action, and it does not map cleanly onto a system that will go on to take many decisions on a person's behalf. For consent to remain meaningful, it should attach not only to individual actions but to the boundaries set on what the agent is permitted to do by design, and the consumer should be able to see, adjust, constrain and withdraw that authority for as long as the agent operates, not only at the point it was first configured. As with the authority-based scaling discussed above, a system that simply recommends a tariff or action poses a different risk to consumers than one capable of automatically adjusting tariffs, demand-side flexibility or customer-owned assets.

Ultimately, consumer trust depends not only on the fairness of outcomes but on confidence that decisions are reached through a transparent, governed and accountable process.

Question 5: Cyber security and resilience

> *a. How should AI assurance align with existing cyber and operational security frameworks, e.g. NIS Regulations?*
>
> *b. How can AI assurance support system resilience, including identifying and mitigating cyber, operational and AI-specific risks?*

a. How should AI assurance align with existing cyber and operational security frameworks?

AI assurance should align with existing cyber and operational security frameworks by being embedded into the live operation of critical services and extended across the full delivery chain that supports them. This delivery chain should be understood in socio-technical terms — not only the technical components and organisations involved, but the humans who operate, oversee and interact with these systems day to day, since assurance that stops at the technical boundary will miss risks that only emerge at the human interface.

In practice, this means treating AI-enabled systems under the same principles already used for high-impact services, including continuous monitoring, change management, access control, logging, red-teaming, business continuity testing and incident escalation, while also ensuring that these controls apply not only to operators but to model providers, cloud platforms, software vendors, integrators and managed service partners. Existing frameworks such as the NIS Regulations provide the baseline for security and resilience, but AI assurance should add controls for risks that are specific to AI, including performance drift, adversarial manipulation, opaque model updates, external dependency risk and unpredictable behaviour after deployment.

Cyber assurance and execution assurance should be understood as complementary rather than overlapping disciplines: cyber security establishes who is permitted to perform an action, through identity, authentication and authorisation controls; execution assurance establishes whether an authorised action remains appropriate within the current operational context at the moment it is carried out. Both are necessary for resilient operation, so that actions remain appropriate even where conditions have changed since authorisation and so that some protection remains even if credentials or trusted systems are compromised.

This is especially important where AI capability may be embedded inside third-party software without full buyer visibility or contractual protection. The goal should therefore be a single end-to-end assurance model in which cyber, operational and AI risks are governed together across the whole service chain, with clear accountability for who monitors, reports and acts when risk emerges. This approach is consistent with the NCSC/CISA-led Guidelines for Secure AI System Development, which frame AI security across secure design, development, deployment and operation, and explicitly includes supply chain security, logging, monitoring and incident management. It is also reinforced by the ETSI Securing AI baseline standard, which sets minimum security requirements across the AI lifecycle for developers, vendors, integrators and operators.

Finally, AI assurance considerations should not be crowded out by security concerns alone – both are distinct and essential dimensions of resilience for critical infrastructure, and neither should be treated as a subset or by-product of the other.

b. How can AI assurance support system resilience, including identifying and mitigating cyber, operational and AI-specific risks?

AI assurance supports resilience when it provides continuous operational evidence about how an AI-enabled service is performing in the real world and how failure could propagate across interconnected systems, suppliers and operational processes. Organisations should monitor AI in production for degradation, anomalous outputs, adversarial interference, unsafe automation, hidden model changes and unexpected interactions with surrounding systems, and recognise that these failure modes are rarely purely technical - they surface at the human interface, in how operators interpret and act on AI outputs, how trust and over-reliance build up over time, and how human oversight can itself be degraded or bypassed.

Socio-technical assurance approaches (such as the work of Careful Industries and Lloyd's Register in this area) offer a useful reference point for capturing these human-in-the-loop dynamics rather than treating AI failure modes as isolated technical events. Organisations should combine that monitoring with wider cyber and operational telemetry so compound risks can be detected early.

Within “drift” specifically, guidance would be more useful if it distinguished three situations, since they call for different responses. Sometimes the system’s behaviour has changed, and the appropriate response is correction, restriction or rollback. Sometimes the system behaves as it did when assessed, but deployment conditions have moved away from what the assessment assumed; the appropriate response there is reassessment or revalidation rather than a change to the system. And sometimes the monitoring evidence is not, or is no longer, diagnostic of the risk it is meant to cover, in which case the monitor itself needs validation or replacement. Treating these as one category may create false confidence, because an organisation facing the third situation while applying controls suited to the first can appear covered when it is not.

Assurance should also test dependency chains by mapping where AI is used, where third parties influence outcomes, and how disruption in one component could cascade across critical infrastructure. Scenario-based exercises, fallback procedures and incident response plans should explicitly include AI-driven failure modes, and incident reporting should make clear when AI was the actual or contributing cause. Organisations should also be able to demonstrate that governance mechanisms — the ability to constrain, pause, deny or escalate AI-supported actions — continue to function correctly during degraded conditions, including cyber incidents, infrastructure failures and communications loss, not only during normal operation.

It is also important to make the positive case: AI is not only a source of cyber risk but an important enabler of cyber defence when adopted securely. As the NCSC’s Frontier AI guidance states, AI can “significantly improve and accelerate cyber defence”, highlighting practical uses including threat detection, enhanced endpoint detection and response, vulnerability discovery, scanning, penetration testing and red teaming. A separate NCSC assessment on the impact of AI on cyber threat to 2027, judged that AI-enabled attacker automation will make identification, tracking and mitigation more challenging without effective AI assistance for defence. The NCSC has also noted that AI is already being used to improve alert triage, log and file analysis, phishing detection, secure code development and threat intelligence. That said, the NCSC is clear that AI is not a substitute for core cyber hygiene — the defensive upside depends on secure-by-design deployment, continuous monitoring and strong baseline security controls.

Question 6: Proportionality

> *a. What does proportionate AI assurance look like across different use cases and risk levels?*
>
> *b. How can assurance approaches be tailored while remaining effective and practical, including for smaller organisations?*

a. What does proportionate AI assurance look like across different use cases and risk levels?

Proportionality should not be based on organisational size, since size alone does not necessarily track to the potential risk of the AI use case, especially for high tech organisations. For example, a small business using AI in a health and safety critical task may be doing something riskier than a larger business running a basic customer service chatbot. The same principle holds within a single organisation, as a company may run AI systems that need very different levels of assurance, from low-risk administrative tools to safety-critical operational

systems, so proportionality should attach to the use case rather than the organisation as a whole.

In addition, proportionality should also be assessed against markers such as the volume and variety of data processed (and the risk of combining it), the purpose of the processing, and the number and characteristics of the individuals affected (e.g. vulnerable groups), alongside the potential consequences of failure, the degree of autonomy exercised, the reversibility of actions and the criticality of the affected infrastructure.

In practice, this suggests a scaling approach: for lower-risk use cases, proportionate assurance may consist of documented governance, periodic review, transparency of purpose and appropriate human oversight. As systems gain greater operational authority or begin influencing safety-critical processes, assurance should progressively incorporate stronger evidence, continuous monitoring, execution-time validation, independent review and demonstrated intervention and recovery capabilities.

b. How can assurance approaches be tailored while remaining effective and practical, including for smaller organisations?

Compliance cost should still be proportionate to an organisation's ability to bear it, but this is a separate question from risk, and the two should not be conflated. Proportionality of assurance should attach to the risk of the use case, not to the size of the organisation undertaking it: a small organisation performing a high-risk function should face the same assurance expectations as a large one performing that function, and organisational size should not be used as grounds to reduce the rigour applied to genuinely high-risk work.

What can and should be reduced for smaller organisations is the cost and complexity of demonstrating assurance, not the standard itself. In practice, this points toward a lighter self-assessment for lower-risk work, along the lines of AI Management Essentials (AIME),[4] rather than full ISO 42001 certification. Practical guidance, reusable templates, reference assurance cases and shared assessment frameworks could meaningfully reduce the burden of implementation while promoting greater consistency across the sector.

Sector-specific guidance should build on the cross-sector foundations that already exist and would benefit from including practical, worked examples of how proportionate assurance applies to common energy-sector use cases, so organisations can adopt an evidence-based approach that is both achievable and proportionate to the level of operational risk. There is also a role for industry-level facilitation, and for larger organisations to support smaller suppliers in their supply chains — particularly SMEs relied on for niche but important inputs — rather than leaving proportionality to be solved at the level of each small organisation alone.

Question 7: External assurance and standards

a. Are existing frameworks and standards sufficient, or is sector-specific AI assurance guidance needed?

b. What role should independent assurance (e.g. audits, certification) play?

[4] AIME (AI Management Essentials) is a free, voluntary self-assessment tool from DSIT to help organisations build sound governance practices around developing and using AI. Crucially, it assesses your organisational *processes*, not the AI products themselves.

c. What are the benefits and risks of external assurance approaches?

a. Are existing frameworks and standards sufficient, or is sector-specific AI assurance guidance needed?

Cross-sector frameworks can provide a useful baseline, but conformity with them does not by itself establish that a particular energy deployment, or the interconnected operational system it sits within, is adequately assured; sector-specific guidance should set the evidence expectations for energy-critical use cases and the role of independent assurance in verifying them.

Rather than creating new assurance frameworks, future guidance should focus on strengthening how existing standards are evidenced within operational environments: standards establish governance expectations, while assurance demonstrates that those expectations continue to be met throughout the operational lifecycle.

This is also the direction of travel set out in DSIT's Trusted Third-Party AI Assurance Roadmap, which sets out Government's ambitions for the third-party assurance market and the near-term actions it will take to support it. Ofgem's sector guidance should build on, rather than duplicate, this wider roadmap.

b. What role should independent assurance (e.g. audits, certification) play?

As mentioned above, independent assurance already has a growing role: the market is genuine and growing, however, organisations currently have no reliable way to judge which providers are credible. This is a barrier that the AI Assurance Consortium will be exploring.

Right now, independent assurance tools and approaches is a most valuable tool when it verifies energy-specific evidence: that a system has been tested against sector-relevant scenarios, that monitoring evidence remains diagnostic over time, and that human oversight is working as intended, rather than simply confirming conformity with generic standards.

Building on this, independent assurance should increasingly evaluate a structured record, linking claims about safety, resilience, fairness and governance to objective supporting evidence generated continuously across the 'Operational Execution Lifecycle,' rather than relying solely on documentation or periodic assessment. This gives independent reviewers a consistent basis for judging whether governance is actually being applied in live operation, not just whether it has been documented on paper.

c. What are the benefits and risks of external assurance approaches?

The main benefit is comparability. An independent verdict against a common sector standard is more portable and trustworthy than organisations self-certifying against frameworks they've each interpreted differently, reinforcing the case for shared sector guidance rather than fragmented approaches.

Risks include false confidence in certification: certification against ISO 42001 assures governance, not deployment risk, and treating the two as equivalent can mask real gaps. There is also false confidence in point-in-time audit, since a one-off audit can be read as demonstrating continuing operational safety even after the supporting evidence has become outdated; independent assurance should therefore combine periodic review with continuous operational evidence, so confidence is maintained throughout the operational lifecycle rather than established only once. Finally, there is a proportionality risk, since external assurance

requirements sized to organisational size rather than risk could burden smaller organisations doing low-risk work while under-scrutinising higher-risk use cases at larger organisations.

techUK's own cross-sector research points the same way: our December 2025 paper, "A Maturing AI Assurance Ecosystem: Sector Specific Applications", draws on qualitative research across health, finance, defence, education and justice and identifies three priorities that we believe apply equally to energy, consolidating around established frameworks rather than creating new ones, adapting ethical principles to sector-specific obligations, and establishing formal mechanisms for cross-sector knowledge transfer.

Sector-specific guidance should also encourage interoperability between assurance frameworks themselves, through common evidence structures, terminology and reporting practices. This would allow assurance to be demonstrated consistently and shared, reviewed and compared across operators, regulators, suppliers and assurance providers, without mandating a single technology or implementation approach.

Question 8: Future guidance

a. What would be most useful in future AI assurance guidance for the energy sector?

b. What types of evidence are most useful in demonstrating outcomes in practice?

c. What examples or case studies would be valuable?

a. What would be most useful in future AI assurance guidance for the energy sector?

Existing standards and governance frameworks already provide an essential foundation. The next evolution is to extend these approaches so organisations can demonstrate that AI-supported decisions remain governed, transparent, and accountable throughout their operational lifecycle.

The most useful future guidance would be guidance that creates comparability across the sector rather than forcing every organisation to invent its own approach. Ofgem should build on existing cross-sector foundations, but translate them into a shared sector template for energy, with common terminology, baseline control expectations, and reusable assurance artefacts (ideally, this would be machine readable ready for AI system use). Guidance would be particularly valuable if it supported a common testing and verification capability, building on the Reg Lab and other collaborative mechanisms, so that firms can assess AI systems against common expectations rather than in isolation.

techUK's November 2024 paper mapping the UK's ethical principles to the assurance techniques and standards used to achieve them offers a ready-made template Ofgem could adapt: it includes a RAG framework (pages 31–34) showing, for example, how industry demonstrates the principle of fairness in practice by running a bias audit. An energy-specific version of this kind of mapping — showing which techniques and standards are being used against which principles, sector by sector — could be a useful building block for the shared template described above. Industry is always looking for more evidence and examples of what good looks like.

To support this, Ofgem may also wish to consider publishing a repository of use cases, drawing together the approaches already emerging across the market into shared reference

points that organisations can adapt to their own AI systems. This would sit alongside, not replace, existing assurance frameworks and standards.

This approach would also be consistent with wider UK Government's shift toward shared infrastructure, reusable tools and common implementation pathways rather than fragmented local experimentation. The Government's AI Opportunities Action Plan response backs a cross-Government "Scan > Pilot > Scale" approach, including a rapid prototyping capability, a data-rich experimentation environment, a scaling service for successful pilots, an AI Knowledge Hub, and mission-focused national AI tenders. In parallel, the Incubator for Artificial Intelligence has been established to pioneer transformative AI applications for the public good and to turn pilots into reusable public-sector capabilities, while the AIME tool was developed by DSIT to distil key governance practices into an accessible self-assessment framework.

For Ofgem, the lesson is that energy-sector guidance should do the same: create common templates, shared evidence structures and sector-wide learning loops that make adoption faster, more coherent and easier to compare across firms. Guidance could also encourage the use of machine-readable operational artefacts—such as process models, decision models, software configurations and structured operational evidence—to improve consistency, interoperability and evidence reuse across organisation. For example, representing standards in structured, machine-readable forms could help to link individual clauses to assurance objectives, controls and operational evidence. This allows evidence generated through engineering, security, testing and operational activities to be reused across multiple standards and regulatory obligations while improving traceability.

Future guidance should focus on strengthening the operational infrastructure that sits around existing standards, rather than treating the standards themselves as the point in need of reform. In particular, much of current assurance practice still relies on traditional, human-led audit, and guidance should address how evidencing can keep pace with AI systems that change and operate faster than a periodic review cycle can track. Whether automated, AI-on-AI monitoring can meaningfully substitute for this – and whether test and evaluation environments are yet realistic enough for such monitoring to be trusted – remains an open question that future guidance should help resolve, rather than assume. This is a question of strengthening how existing frameworks are evidenced and operationalised in practice – through continuous monitoring, reusable assurance artefacts and shared infrastructure of the kind set out above — not a case for replacing the standards model itself.

b. What types of evidence are most useful in demonstrating outcomes in practice?

On the types of evidence most useful in practice, two properties are most relevant. First, evidence is most useful to a later reader when it records what was and was not established, and the conditions under which it should be re-checked. That applies to model cards, test reports, third-party audit statements and the assurance material supplied with bought-in products alike. Second, guidance should look for some evidence that the assurance method itself is sensitive to representative failures it is intended to detect, for example performance against controlled cases or past incidents with known outcomes, where such evidence is available.

Beyond this, the same comparability principle set out in 8a applies to evidence formats, which includes: standardised metrics, common reporting templates, audit trails mapped to

recognised frameworks, incident taxonomies, and evidence structured in a way that allows regulators, operators and suppliers to compare like with like. In a sector where systems are interconnected and assurance maturity is uneven, evidence becomes more valuable when it supports coordination, benchmarking and shared learning, not just internal assurance.

For higher-impact operational uses, this evidence should also preserve the relationship between the AI-supported recommendation, the live execution context, the governance controls applied, any human intervention, the final action taken and the resulting operational outcome. This would allow assurance evidence to support not only compliance and comparison, but also incident investigation and operational reconstruction.

Future guidance could draw on the same logic behind the Government's emerging AI adoption architecture: publish reusable methods, capture lessons centrally and make successful patterns easier to replicate. The AI Knowledge Hub is intended to publish best-practice guidance, results, case studies and open-source solutions in one place, while i.AI describes its model as building reusable tools and making code and methods available across Government. Applied to energy, that suggests evidence formats should be designed not only for compliance, but for transferability across operators, suppliers and regulators.

c. What examples or case studies would be valuable?

The most valuable case studies would be collaborative or sector-level examples rather than isolated organisational success stories. For example, case studies showing how multiple actors used a sandbox or Reg Lab environment, how common information models or shared ontologies improved assurance, or how data-sharing and interoperability barriers were addressed in practice would be highly valuable. These examples would help Ofgem design guidance that is scalable, repeatable and relevant to the interconnected nature of the energy system.

Case studies would also be valuable if they illustrated how assurance should scale across different levels of operational authority — from low-risk decision support through to more autonomous operational control — showing how governance, evidence collection, intervention and incident investigation work together in practice, rather than focusing only on model evaluation or compliance activity in isolation.

It would also be useful to reference wider public-sector examples of AI exploitation and acceleration where Government has deliberately created common assets that others can build on. Relevant examples include i.AI's Consult, which processes consultation responses at scale; Redbox, Lex and Parlex, which provide reusable AI tools for civil servants; and the Government's use of an AI scan function, AI Commercial Strategy and challenge-led procurement to speed up procurement and scaling. The value of citing these is not that energy should copy the use cases directly, but that Ofgem can borrow the model: shared testbeds, common platforms, documented playbooks and centrally curated evidence that reduce duplication and accelerate safe uptake.

Internationally, Singapore's AI Verify Foundation and its Moonshot project offer a useful comparator for open-sourcing assurance best practice across a testing community, and Ofgem may wish to draw on that model when designing shared sector tools.

Closer to home, techUK helped launch the Portfolio of AI Assurance Techniques with what was DSIT's Centre for Data Ethics and Innovation in 2023. It collects real-world case studies from multiple sectors covering technical, procedural and educational assurance approaches,

mapped against the principles in the UK Government's AI regulation white paper, and is aimed at anyone designing, developing, deploying or procuring AI-enabled systems; DSIT is clear that inclusion of a case study is not a Government endorsement of the technique or organisation, only an illustration of the range of options available. The OECD's catalogue of tools and metrics for trustworthy AI is also a useful complementary resource for the same purpose. Ofgem could point energy-sector organisations to both as an immediate, ready-made source of case studies rather than commissioning new ones from scratch, while also encouraging energy firms to contribute their own examples back into the Portfolio.

Dedicated funding would help too. techUK has recommended that funding from existing AI pools, such as the Advanced Research and Invention Agency (ARIA), could be extended to the energy sector and ringfenced for projects that evaluate, mitigate and resolve the risks of AI.

Question 9: Communication

a. How should organisations communicate AI assurance to different audiences?

b. What information is most useful for consumers, boards, senior management and affected groups?

Communication needs different tiers for different audiences, and be proportionate to both the audience and the level of operational risk as different stakeholders need different levels of technical detail, but all should be able to understand how assurance has been achieved, how accountability is maintained, and what evidence supports that confidence.

Boards, senior managers, investors and insurers require risk and liability set out in business terms, such as what could go wrong, who is liable, and how it is insured.[5] Operational teams are a distinct audience in their own right, requiring more detailed information to support safe operation, incident response and continuous improvement. This includes clear visibility of operational boundaries, execution constraints, intervention mechanisms, escalation procedures, governance decisions, operational evidence and lessons learned from assurance activities and operational events.

Regulators need an evidenced audit trail, such as standards mapping and incident history. This should focus on structured evidence rather than narrative assurance alone — demonstrating not only compliance with recognised standards, but how governance has been applied throughout operation, in a way that supports independent review of significant recommendations, interventions and operational changes without requiring disclosure of commercially sensitive information.

For consumers and affected groups in particular, there needs to be plain language about outcomes - what the system does, what safeguards exist, and how to raise a concern. Given how complex AI assurance is, expecting most consumers to understand its underlying mechanics is unrealistic and arguably the wrong goal. One example of a current assurance mechanism in use that provides the market with details of an AI model's intended uses, architecture, training data sources and performance boundaries to evidence safety and performance is a system model card. These systems incur model drift, which means assurance is iterative rather than a one-off certification. A static "safe" stamp is therefore not

[5] This should include the operational authority granted to AI systems, residual risks, governance effectiveness, significant incidents, assurance findings and areas requiring management attention, so that AI assurance becomes an ongoing governance activity rather than a periodic compliance exercise.

possible, but stronger external communication of the actions taken to evaluate and mitigate risk could still prove valuable, including when they last took place and by whom.

Moreover, trust is strengthened when consumers understand both the protections that exist and the routes available should concerns arise, including when AI has materially influenced a decision affecting them and how that decision can be reviewed or challenged. Disclosure and science communication should therefore be layered: accessible plain-language summaries for the public as standard, with the option for investors, procurement teams, or any other interested party to request the fuller technical detail and assurance standards that sit behind it.

Across all audiences, communication should increasingly distinguish between governance claims and assurance evidence: organisations should be able to explain not only what governance policies exist, but how those policies are evidenced during live operation, moving assurance communication from statements of intent towards demonstrable operational confidence.