Guidance

Ethical AI use in the energy sector (version 2)

Publication date:	13 May 2026
Contact:	Jonathan Thurlwell
Team:	Emerging Technologies Team
Email:	EmergingTechnologies@ofgem.gov.uk

Contents

Foreword

Artificial intelligence (AI) is an increasingly important part of Great Britain's energy system. Since the first edition of this guidance was published in May 2025, AI has moved from early pilot deployment to increasing operational use across the sector. It is supporting network planning and operation, improving forecasting and flexibility, enhancing asset management, and shaping how consumers interact with energy suppliers and networks through AI-enabled services. At the same time, rapid advances in AI capability have increased the importance of effective governance and risk management.

These developments are taking place alongside wider change within the energy system and its regulation. The transition to net zero is driving a more decentralised, digital and data-intensive system, while demand for electricity from digitally enabled infrastructure continues to grow. Following the Ofgem Review, as a regulator we are also changing our approach to strengthen protections for energy consumers, enhancing the focus on outcomes, and ensuring we are equipped to regulate a more complex, dynamic and innovative sector.

In this context, the independent review of AI deployment in the electricity networks highlights both the significant potential of AI to support a safe, efficient and flexible electricity system, and the need for clear accountability, proportionate assurance and effective oversight as deployment scales. This guidance provides sector-specific direction on the ethical adoption of AI within the existing energy regulatory framework.

The purpose of this second edition is to support stakeholders to adopt AI in ways that are safe, secure, fair and environmentally sustainable, while enabling innovation that benefits consumers. It reflects what we have learned since 2025, including the growing use of AI in areas that directly affect people, such as customer service, pricing and automated decision-making. In response, this edition strengthens expectations around transparency and explainability.

Our regulatory approach to AI remains outcomes-focused and proportionate. This guidance complements, and does not replace, existing legal and regulatory obligations. We will keep it under review as AI technologies and use cases evolve and consider whether our regulatory approach needs refining.

AI has a vital role to play in delivering a clean, secure, and affordable energy system. We look forward to continuing to engage and collaborate with our stakeholders to realise the huge opportunities AI presents, enabling growth, supporting the transition to net zero, and protecting consumers.

Tim Jarvis, Interim Chief Executive

1. Introduction

1.1 We are the independent regulator of Great Britain's (England, Scotland and Wales) energy markets and networks. We also administer a range of environmental and social schemes for the government. Our powers, duties and objectives come from statutes enacted by Parliament.

1.2 Our principal guiding objective is to protect the interests of current and future energy consumers. These interests are taken as a whole and include their interests by:

a. maintaining the security of gas and electricity supply

b. promoting sustainable economic growth

c. supporting the UK government in meeting the 2050 net zero target and other associated targets

1.3 We must carry out our functions in the manner that best fulfils our objectives. Wherever appropriate, we should do this by promoting effective competition. Before we do this, we must consider whether consumers' interests would be better protected if we acted in other ways.

1.4 We conduct our work by following the regulatory principles of transparency, accountability, proportionality, consistency and other principles that we consider represent best regulatory practice. Our regulatory approach is to identify and mitigate risks to our objectives, including the use of AI by licensees and other regulated persons, and the harm that this potentially creates for consumers and the energy sector.

1.5 We committed to publishing clear regulatory guidance on AI to maximise the potential benefits and to minimise potential harms to consumers. This guidance document fulfils that commitment. This second edition of the guidance contains additional detail throughout on transparency and explainability, including a new good practice at the end of the governance and policies section, reflecting the increasing use of more advanced and agentic AI systems across the sector. The areas of AI transparency and explainability were further explored through a series of stakeholder workshops in January and February 2026.

1.6 Overall, these updates sit alongside wider work by Ofgem to evolve our regulatory approach to AI, including engagement with national and international stakeholders on best practice, and the development of complementary regulatory tools to support safe, responsible innovation in the energy sector.

How to use this guidance

1.7 This document sets out good practice for the ethical adoption of AI across Britains' energy sector. Where stakeholders are developing novel, high-impact or

uncertain AI use cases, engagement with Ofgem's AI Reg Lab may be appropriate to explore regulatory considerations. The proposed AI technical sandbox is intended to be a safe, controlled environment where energy sector participants can test and evaluate AI systems, with proportionate regulatory oversight before any live deployment. These tools are complementary and should be used alongside this guidance where relevant.

Purpose

1.8 The National Cyber Security Centre (NCSC) defines artificial intelligence (AI) as computer systems which can perform tasks usually requiring human intelligence. Throughout this document, we adopt NCSC's definition.

1.9 AI has certain characteristics that sets it apart from non-AI systems. For example, AI systems cannot be fully tested, nor completely explained. This can be due to AI dependence on training data, the complexity of the task being performed, or the very large unstructured input space. Examples of this include neural networks and reinforcement learning. In contrast, non-AI systems may be fully tested and explained. They can have well-defined tasks and do not require judgement or reasoning. Also, they are not dependent on training and training data.

1.10 The purpose of this guidance is to encourage an ethical approach to AI adoption in the energy sector. Our approach to AI consists of four outcomes: safety, security, fairness and sustainability. Our ethical approach section explains our approach.

1.11 This guidance sets out good practice for stakeholders to consider when evaluating opportunities to use AI. These good practices are intended to supplement and support the existing regulatory regime that applies to the energy sector.

Scope

1.12 It is recognised a wide audience might require guidance on the deployment of AI in the energy sector. This guidance is therefore aimed at all stakeholders involved with AI in the sector which includes, but is not limited to, licensees, market participants, operators of essential services, dutyholders, technology companies, AI developers, consumer groups, other regulators and government.

1.13 In addition to complying with Ofgem's regulatory framework, stakeholders are also required to comply with other areas of legislation that apply to the energy sector and AI use, including data protection, equality, human rights, and health and safety.

1.14 This guidance is grounded in current AI capabilities and does not seek to anticipate all potential future system-wide or transformational changes to the energy system.

Pro-innovation

1.15 To promote our growth duty, we have aligned with the government's pro-innovation approach. The principle of proportionality informs our thinking and approach to AI, including any potential regulatory actions. As part of this, our good practice guidance is intended to support innovation through the use of AI and protect energy consumers from the risks. Our approach stresses the importance of risk management, that is the proportionate actions taken by stakeholders, informed by the identification and management of risks associated with the use of AI.

Effect

1.16 Based on our current understanding, Ofgem considers the regulatory framework to be appropriate to govern the use of AI. Accordingly, this good practice guidance aims to provide context specific to AI use and therefore complements Ofgem's existing regulatory framework. Licensees are already required to comply with their standard licence conditions, for example treating consumers fairly and operational capability. However, our regulatory approach may change if required as this technology and its outcomes evolve in the energy sector.

1.17 This guidance covers governance measures and policies to ensure effective oversight of AI, a risk approach to help stakeholders identify and manage risks associated with AI, and the competencies required for the ethical adoption of AI. The appendices provide information on the legal and regulatory obligations that stakeholders are required to comply with in the energy sector (Appendix 1), AI standards (Appendix 2), key legal and regulatory expectations around AI supply chain management (Appendix 3), data use and management (Appendix 4), AI and cyber security (Appendix 5), and transparency and explainability (Appendix 6). A glossary of key AI terminology is also provided.

Review

1.18 Ofgem will keep the existing regulatory framework and this guidance under review to ensure our approach continues to meet the needs of regulating AI effectively in the energy sector. Similarly, stakeholders should ensure that their approaches, policies, practices and controls remain up to date with relevant documentation cited in this guidance as AI technologies and associated risks evolve.

Background

1.19 In January 2025 the government set out its action plan to ensure the UK realises the opportunities that AI provides. The government's assessment of progress in January 2026 (further details set out in AI opportunities action plan: one year on) reinforces expectations on regulators to enable safe, trusted and scalable AI development through proportionate regulation, safety and assurance.

1.20 As the energy regulator for Great Britain, we have taken the following approach to developing this guidance and regulatory tools:

a. a call for input was published in April 2024 on the safe and responsible use of AI in the energy sector

b. our high-level strategic approach to AI was published in April 2024, which demonstrated our plan to have a robust method of regulation in the energy sector

c. conducted qualitative research into how people felt about AI in the energy sector

d. consultation to support the first edition of the guidance document which was published in December 2024

e. launched AI Reg Labs (pilot conducted in February 2025 and established on permanent basis every four months since October 2025), with high-level findings shared one month after each lab (October 2025, February 2026)

f. publication of the first edition of the guidance in May 2025

g. publication of an AI technical sandbox call for input in July 2025

h. publication of a consultation on a proposed 12-month pilot of an AI technical sandbox in January 2025

i. engagement across the sector to develop the guidance further in the areas of AI transparency and explainability

j. broader engagement with national and international stakeholders on best practice on guidance and regulatory tools

k. publication of the second edition of this guidance in May 2026

l. publish an AI technical sandbox decision in June 2026

m. publish an AI assurance call for input in Summer 2026

1.21 This guidance sits alongside wider government work on AI, including's DESNZ's vision for an AI-enabled energy system and the independent review of AI deployment in energy networks led by Lucy Yu, the DESNZ-appointed AI Champion for Clean Energy.

2. Our ethical approach

2.1 AI technology is a continuation of the development of broader digital technologies. It builds on the foundations of hardware, software and data to create a novel capability.

2.2 As well as the expectations around the legal and regulatory framework and good practices outlined in this document, existing good practices covering hardware, storage (including cloud), software, supply chain, security, data, standards, and associated regulations remain applicable. This includes the full life cycle management for each of the constituent parts.

2.3 Stakeholders should ensure appropriate oversight and controls are in place that are relevant to both the AI and its intended use to make sure reasonable steps are taken to minimise any negative impact on safety, security, fairness and environmental sustainability. We summarise the risk around each of these in turn below, with key considerations of how to address them.

Safe AI

2.4 AI could be used in systems ranging from applications where failure will result in negligible impact on safety through to its use in decision making for interlinked critical national infrastructure which could, for example, result in power outages if not effectively managed. Assessments of AI use must therefore consider the context of operation across the energy value chain and encompass appropriate use, and account for potential misuse (both intentional and otherwise).

Secure AI

2.5 AI can present novel security risks alongside standard cyber security threats and risks due to the high pace of development and rate of change for this emerging technology. As a result, security should be considered at the outset and throughout the AI system life cycle.

Fair AI

2.6 AI can introduce unintentional bias that can result in direct or indirect discrimination. Ensuring fairness in AI outcomes is crucial for fostering consumer trust and increased confidence in AI use across the energy sector. Stakeholders must ensure AI systems serve energy consumers fairly by implementing robust governance, systems and processes.

Environmentally sustainable AI

2.7 AI is viewed as playing an important role in the UK's economic development, national security and improving our energy system efficiency and sustainability. However, the growth of AI and other data intensive technologies is predicted to

consume relatively large amounts of electricity due to the need for additional data centres and computing power.

2.8 Ofgem recognises that AI operates within a complex and often fragmented landscape. Policy regarding its environmental sustainability will continue to evolve. We are committed to encouraging innovation while raising awareness about environmentally sustainable practices through collaboration and where necessary the creation of guidance to help the UK to meet its net zero, and other associated, targets.

2.9 To deliver environmentally sustainable AI outcomes requires stakeholders to adopt good practice in the areas of policies, governance, and through life risk management.

2.10 In summary, delivering safe, secure, fair and environmentally sustainable AI outcomes requires stakeholders to comply with their regulatory obligations and to adopt good practice approaches. This includes governance and policies, risk and AI implementation throughout the AI life cycle including design, development, deployment, operations, monitoring, maintenance and decommissioning after use.

2.11 Through appropriate governance, robust risk management, and the right capability across the AI life cycle, stakeholders should be able to manage AI in line with this ethical approach. We outline good practices in relation to each of these in the following three sections.

3. Governance and policies

Good practice

3.1 Stakeholders have appropriate policies, processes and procedures in place to ensure proportionate and effective governance including oversight of the supply chain and use of AI systems, with clear lines of accountability established across the AI life cycle.

Description

3.2 In large and medium sized stakeholder organisations, strategies and organisational arrangements for the safe, secure, fair and environmentally sustainable use of AI are typically driven at board level, and senior management level, and take account of ethical standards. In small stakeholder organisations effective governance may be achieved through less formal arrangements. Either way, good practice governance arrangements should consider the proportionate application of the following good practices.

3.3 Arrangements for effective governance do not need to be AI-specific but need to account for the characteristics of AI and should consider what is proportionate given its context and associated risk to energy consumers. Strong governance arrangements should include robust independent challenge to help ensure risks are managed and opportunities are realised. Depending on the organisation this can be through, for example, board arrangements or effective management. Stakeholders' governance arrangements may need to take account of the following to ensure proportionate risk management is maintained by their:

a. risk appetite to systems containing AI and any associated data

b. approach to managing the AI system and associated data, including appropriate policies, standards (see Appendix 2 for AI standards), processes, procedures and, or practices which translate and embed the leadership's direction into business-as-usual activities

c. governance roles and responsibilities including risk ownership and risk mitigation

d. need for regular review to ensure governance arrangements are in line with the latest and developing AI risks and organisation's risk appetite

e. implementation of robust change control and data governance measures

Practice 1: clear strategy, with articulation of outcomes and associated risks

3.4 At the outset, stakeholders should define their strategy on the use of AI, which takes account of their operating environment, associated risks, and any positive

AI application implications. This strategy, if carried out effectively, can be used to ensure governance arrangements are proportionate to the risks associated with the use of AI. In some circumstances it may be necessary for the strategy to be agreed by the board (or equivalent governing body). The strategy should be communicated and reviewed periodically, in accordance with the organisation's policy, and amended due to any significant or relevant learning or change in operational environment.

Practice 2: effective accountability and governance

3.5 Effective accountability should ensure that the use of AI leads to positive outcomes. There is a strong link between accountability and trust in AI. Appropriate governance and accountability arrangements should lead to acceptable standards, quality, and performance and consequentially result in appropriate levels of trust in a specific application of AI or the technology itself.

3.6 Effective governance applies across the AI life cycle from concept to decommissioning. It should include aspects relating to the AI model, for example, specification, development, acquisition and the associated supply chain, training, and data management including data protection. Additionally, governance should cover the AI application including monitoring, suitable protection and mitigation against faults, consideration of the need for redress, and system modification. Further guidance around supply chain management is set out in Appendix 3, on data use and management in Appendix 4, AI and cyber security in Appendix 5, and transparency and explainability in Appendix 6.

Practice 3: clear guidelines and policies

3.7 Stakeholders are accountable for the consequences of AI use within their organisation and to their users. This requires robust risk management strategies and response plans for potential AI-related incidents. Where necessary, stakeholders employing AI in their operations should establish clear and proportionate guidelines and policies for its use.

Practice 4: clear leadership across the stakeholder's organisation

3.8 Effective AI governance is built on strong leadership to ensure energy consumers are protected from the most significant risk and is likely to include the following characteristics:

3.9 a. clear oversight: depending on organisational arrangements, oversight over AI can, for example, reside with the full board, an existing committee (for example, audit or technology, cyber security), or a newly formed committee dedicated to AI.

b. effective management: clear roles and responsibilities should be cascaded down through the stakeholder's management chain (including supply chain organisations) through clear and well-understood channels for communicating and escalating risks. These roles and responsibilities should be kept up-to-date

and may be informed using appropriate tools. Decision makers should be given the necessary authority to drive risk management activities and clear means to escalate AI risk management decision making.

c. access to competent persons: to help manage risks associated with AI use, stakeholders should ensure they have access to the necessary competence to ensure risks are identified and managed effectively and efficiently.

d. informed: to ensure effective management of AI deployment, decision makers should have access to information (for example, metrics and reports) to ensure uncertainties are taken into account, assumptions are challenged, and risks are managed. This information should focus on the potential impact the use of AI may have on safety, security, fairness, or environmental sustainability. Helpful information could include:

i. AI risk status and trajectory, including risk within the supply chain

ii. AI performance (supported by key performance indicators)

iii. operational experience gained from the application of AI

iv. outcome of any review (for example, risk audit, impact assessment, benchmarking)

v. outcome of any scenario planning to anticipate and mitigate potential risks (for example, data misuse or unintended AI consequences, including to the end users of the AI systems for example, consumers)

vi. effective AI governance arrangements should be communicated to those impacted. These communications may vary depending on the role of individuals within the organisation.

Case example: Board governance

3.10 Strategies and organisational arrangements for the use of AI are often driven at board and senior management levels.

3.11 Board members should be accountable, have the ultimate oversight, and own the role of delegating responsibilities. This may include top-down strategy and principles, and policies that take account of strategic context. To undertake this role effectively, a board should have the necessary competence, access to reporting metrics, opportunities to have meaningful discussion on AI and be aware of any benefits, assumptions and potential shortcomings of models and analyses. Competencies may include, for example, appointing board members with expertise in responsible AI (technical and non-technical), ethics, and consumer interests to effectively oversee AI opportunities and risks. These can be developed by co-opting specialist board members, undertaking training and coaching sessions.

Case example: Project governance

3.12 There should be clear and agreed justification for the use of AI with goals and outcomes defined. Within a project, arrangements should be in place to enable adherence to relevant good practice through planning, documenting, monitoring, and escalating to ensure ethical use of AI and governance goals are embedded, including:

a. change control

b. data governance and risk management

c. testing models and procedures

d. regular review and internal audit

e. recordkeeping, including but not limited to board and other meeting minutes and associated supporting materials

f. ways to educate users about the use of AI and identifying errors and mitigations

Practice 5: proportionate organisational transparency and explainability

3.13 To enable organisations to be transparent about the use of AI to individuals potentially affected by AI systems, organisations that use AI should establish an organisational framework that supports AI transparency and explainability in a proportionate way throughout the AI lifecycle by:

a. establishing clear processes for proportionate transparency and explainability, including clear roles and responsibilities to ensure teams and individuals are accountable for monitoring, validating and interpreting any AI-generated results

b. ensuring that AI transparency and explainability requirements are met throughout the lifecycle (for example, design, development, deployment, operations and decommissioning)

c. maintaining records that explain the basis of AI decisions in a proportionate manner. Where necessary, the basis of the decisions should have sufficient detail for stakeholders to review input and assure processing

d. creating governance mechanisms that allow users to escalate AI decisions when they appear incorrect, biased or opaque

e. being transparent about the use of AI by providing meaningful and easy-to understand information that is context specific

4. Risk

Good practice

4.1 Stakeholders should evaluate the risks associated with the use of AI in the energy sector to help them effectively identify and implement measures necessary to manage those risks.

Description

4.2 Our regulatory approach generally focuses on outcomes in managing risk rather than setting prescriptive rules on the application of AI. The aim is that this approach allows stakeholders to have flexibility in how they deliver desired outcomes and manage the risk while empowering stakeholders to innovate. The following risk framework is intended to assist stakeholders in understanding the energy regulator's view on AI risk and proportionality. The alternative to outcome-based regulation is a prescriptive approach, regulating through the application of rules. These rules would need to account for all eventualities which would be very difficult to implement effectively for rapidly developing technologies such as AI.

4.3 Stakeholders are expected to identify, assess and prioritise AI-related risks within their organisations, implementing risk management measures proportionate to the likelihood, potential impact and materiality of harm to consumers, the energy system or market integrity. These arrangements should form part of stakeholders' wider risk management and governance processes and be kept under review to ensure they remain appropriate as AI use and associated risks evolve.

4.4 The use of AI can present novel risks including the risk of bias in training data, model inaccuracy and accounting for shifts in model due to adaptation or changes in the application environment.

4.5 From the outset it is important to consider the use of AI as a component within a larger application system. For example, this may be an AI component within a wider engineered and, or technical system or persons overseeing the use of a large language model. This is important because the technical and, or human systems that surround the AI component play an important part in managing the risk associated with the use of AI, including any uncertainty associated with the AI component and its operating environment.

4.6 The rapidly developing nature of AI is resulting in several principles-based approaches to the management of the risks associated with the use of AI. These approaches are maturing over time. We continue to review a number of these frameworks, including:

a. National Risk Register 2025 on GOV.UK

b. Considerations for developing artificial intelligence systems in nuclear applications, Office for Nuclear Regulation

c. Assurance of machine learning for use in autonomous systems (AMLAS) tool v1 user guide, University of York

d. ISO/IEC 23894 (AI Risk Management)

e. ISO/IEC 27001 (Information Security Management Systems)

f. ISO/IEC 42001 (AI Management System)

g. AI Risk Management Framework, National Institute of Standards and Technology, US Department of Commerce

h. Microsoft Responsible AI Standard v2, General Requirements

i. The AI Security Institute

4.7 It is good practice that the potential users of AI in the energy sector consider the application of these framework practices in a proportionate manner in line with the safety, security, fairness and environmental sustainability risks associated with its application.

Practice 1: ensure AI is the most appropriate technology to use

4.8 Stakeholders considering using AI clearly articulate the benefits of AI, its use and any associated risk compared with alternative or traditional technologies. As part of this, a clear plan for the entire AI project life cycle, from planning, development, deployment to monitoring, auditing and decommissioning is developed.

Practice 2: risk assessment and management

4.9 Use of AI is accompanied by proportionate management of risk established through an effective evidence-led risk assessment. The purpose of the risk assessment is to guide stakeholders considering using AI towards proportionate actions and ensure the risks of failure are avoided or mitigated or both. Risk is commonly expressed as the combination of the probability of something adverse happening and the consequence of the adverse event. Depending on level of risk associated with the use of AI it may be beneficial for potential users of the technology to use risk matrix frameworks, for example Machine Learning Principles, on the NCSC website, and keep a record of the assessment to aid future use. Risk areas might include, but not be limited to, operational, legal and reputational risks.

Practice 3: adopt good practice in specification and development

4.10 In a similar manner to conventional software, robust AI components, and wider systems intended to protect against any uncertainty associated with the use of AI, are developed using established good practice over the life cycle of the systems

(that is, from concept to system end of operational life). This includes clear requirements (for example, input specification, output requirements, assumptions) and good practice in software development, data governance and management, and AI component training (including AI ethics for responsible AI use).

Practice 4: understand the characteristics of the AI component within the broader system

4.11 The system containing AI is tested in a proportionate manner to develop confidence in the performance characteristics of the AI and the surrounding system (human or physical). Given the difficulty determining the reliability of any AI component it is likely that the broader system, not just the AI component, will make a large contribution towards the overall reliability. The AI component and its operational environment will likely change with time due to factors including:

a. ageing

b. environmental changes

c. evolution of organisational culture, for example perceived trust in the AI component

d. system changes, for example changes in signal sampling (known as quantisation) and timing changes

Practice 5: identify and address potential failure modes that could impact safety, security or fairness

4.12 Users of AI assess potential failure modes and maloperation in a proportionate way to make sure that the broader systems can control and mitigate the consequences of potential failures and maloperation. The assessment of failure modes and maloperation may very well drive the consideration of engineering protection (functional safety) or human intervention. This assessment considers an understanding of any unintended consequences, such as loss of skill base due to the introduction of AI or overreporting of positive outcomes. Monitoring can also be used to help identify any drift in behaviour of the AI system or its operating environment. This could be via the use of independent systems such as diversity in AI components, the use of digital twin comparators, or conventional systems (for example, functional safety). Arrangements for mitigating the consequences of AI component failure and recovery should be implemented to support, if necessary, the overall recovery of operations.

Practice 6: develop confidence in the performance of the AI component within the broader system

4.13 Rigorous testing is likely to be essential in building confidence in the use of AI prior to its application. However, given the complexity and uncertainty associated with AI systems, and the complexity of the AI components, determining the overall level of uncertainty of an AI component may be difficult, and potentially impossible. Therefore, arrangements are needed to make sure the output of AI components is used appropriately. It may be beneficial to develop metrics to evaluate the performance of the AI component, and the suitability of the controls and mitigation of the broader system or systems through monitoring and evaluation.

Practice 7: access to competent persons

4.14 Ensuring the necessary skills and experience needed to deploy AI effectively and safely is available and developed. This includes:

a. operational application knowledge including understanding the consequences of failure and maloperation

b. behaviour and culture to ensure the deployment of the system or systems containing AI is done in such a way as to reduce any associated risk

4.15 It is also important stakeholders provide appropriate training to staff generally, depending on their role, to ensure they understand AI technologies, their use and impact on their activities.

Practice 8: human and AI interaction

4.16 It is important to take account of the complexity of the interaction between humans and the systems containing AI. As such, human oversight from the earliest stage is recommended as a risk control and mitigation measure. It is important to consider both overconfidence and a lack of trust in the AI components especially where human oversight is used as a risk control and mitigation measure.

Practice 9: monitor and review

4.17 Arrangements are reviewed on a regular basis and in a proportionate manner to ensure they remain effective. Again, metrics may be helpful to monitor system performance.

5. Competencies

Good practice

5.1 Stakeholders have the right knowledge, skills and capability to understand how AI opportunities can be realised, and any associated challenges are clearly understood and appropriately mitigated. This includes the need for resilience, scalability and robust management of the associated vulnerabilities, risks, and threats.

Description

5.2 For traditional software services and components, that are rules based and deterministic in nature, suitable assurance can be achieved by established testing approaches and good software engineering practices over its life cycle. Read more about this in IEC 61508, on the International Electrotechnical Commission's website. With AI based technologies, that are probabilistic in nature, new challenges, issues and risks arise. This probabilistic behaviour presents challenges for the resilience of developed services but also creates risks around untested outputs that could cause components to fail in unexpected ways, creating potential vulnerabilities and opportunities for attack.

5.3 AI based components and services, whether developed in-house or externally, with hosted services to be integrated into the organisation, require specialised knowledge and skills due to both AI's unique behaviour and characteristics, as well as the rapid pace of development and changing capabilities.

Practice 1: robust training plans

5.4 Dependent on application and proportionate to the risk, stakeholders have a robust plan, not only to develop, but also maintain the appropriate knowledge, skills and capability in AI based technologies to ensure a robust approach to selecting, designing, developing, operating and governing the appropriate AI based solutions. Our good practices are to:

a. define a baseline level of foundational knowledge needed around AI for all staff, including ensuring appropriate training and testing of internal knowledge for established policies and procedures around the safe, secure, fair and environmentally sustainable use of AI

b. define and agree the appropriate level of additional AI knowledge and skills needed by role across the organisation, including governance, management, technical and operational areas

Practice 2: suitably qualified decision makers and staff

5.5 Roles are assigned to those with the appropriate knowledge and capabilities for AI including alignment with legal, regulatory and standards, managing AI risks, and

fostering transparency. Consideration should be given to the benefits of designating a person (such as an AI officer) to oversee the ethical, responsible and effective deployment of AI technologies. This person may be responsible for ensuring AI not only drives value, but does so in a responsible, transparent and compliant manner.

5.6 AI decision makers should have appropriate skills, knowledge, tools, and authority. Those responsible for AI risk management and treatment should be:

a. suitably experienced in the design, development, use and operation of AI

b. familiar with the operational model of the stakeholder and in relation to AI use

c. knowledgeable in the assets, as well as dependencies on any third parties used to deliver and support the AI services

5.7 Stakeholders implement a training and development plan to upskill staff around AI appropriate to their roles and responsibilities, including management, project, technical and operational roles.

Practice 3: knowledge management policies and procedures

5.8 Stakeholders ensure knowledge management policies and procedures have been updated to cover AI knowledge and skills, including managing the rapidly changing landscape of this emerging technology. This should include cyber security teams having an ongoing responsibility to monitor existing and emerging threats and vulnerabilities associated with the adoption and use of AI, both within the organisation and the external threat landscape. In addition, establishing an approach to share experiences and lessons learnt from the use of AI, such as forums and communities of practice can assist in knowledge management.

5.9 Stakeholders may consider using appropriate proof of concept, research and development, learning projects, collaboration and partnerships to develop and test organisational capabilities to safely design, deliver and operate AI-based technologies

Practice 4: horizon scanning

5.10 Stakeholders monitor and track developments in AI, including in cyber security, to identify important areas where learning and development plans require updating and access to competent people may need to be arranged.

6. Sector-specific examples

6.1 The following case examples are targeted at supporting adoption of AI within the energy sector, in line with this guidance. Appropriate good practice actions are highlighted but these should not be considered exhaustive.

AI in consumer interactions

6.2 AI can be used in a range of ways to support service agent interactions with a customer. For example, it could provide case history summary including key points of previous interactions and therefore reduce the time to enter the context of a case. With increased consumer insight AI could also assist in drawing the attention of the customer interaction agent to the key issues, expectations, key policy documentation, and processes.

6.3 However, if ill-conceived or implemented poorly, this could result in inaccurate or irrelevant information being used to inform the customer, with the potential for them to be treated unfairly. Transparent communication of appropriate information to consumers is therefore important, as is the ability of those involved in delivering the services, whether supported by systems, technology or both, to interpret and use that information correctly.

6.4 Considerations include:

a. the role of governance in developing, implementing and overseeing the effectiveness of AI

b. undertaking a risk assessment that identifies the necessary control measures throughout the life cycle, ensuring that the AI system functions as defined, treats consumers fairly, and cyber-physical systems comply with health and safety legislation

c. ensuring the AI system embeds proportionate explainability throughout its design that is consistent with the risk profile, including AI uncertainty and the potential impact on human users

d. implementation of the identified control measures

e. implementation of mitigation measures including access to any necessary redress should consumers be treated unfairly

f. testing and monitoring the effectiveness of the AI model, the implemented controls, and mitigation measures

g. training agents to help identify any incidents where consumers may be being treated unfairly

h. any other measures necessary to reduce the risk to a tolerable level and ensure consumers are treated fairly

i. ensuring ongoing transparency to consumers and stakeholders about the use of AI in a proportionate way. This might include clear communication that AI is being used, the purpose of its use, and any implications for consumer or stakeholder decision-making or service delivery

AI used to identify and assist vulnerable or excluded consumers

6.5 AI could help to identify excluded consumers and highlight relevant information for service decision making, such as during engineering works, or for the priority service registers. In addition, AI could be used to identify and adapt processes to maximise the potential for consumers to be treated fairly, and if necessary, highlighting procedures for redress. For example, this could apply to consumers in vulnerable circumstances or consumers who are digitally excluded.

6.6 In this scenario, identifying where consumers may be treated unfairly could be difficult. For example, consumers may be digitally excluded and therefore poorly represented in digital data used, for example, in AI training sets. Care should be taken to ensure that any intervention does not reinforce any existing bias and that consumers' data is protected according to existing legal requirements.

6.7 Considerations include:

a. the role of governance in developing, implementing, and overseeing the application of AI to ensure it is effective and not reinforcing bias

b. undertaking a risk assessment which identifies areas where the existing system may fail to support excluded consumers and identify measures that results in the consumer being treated fairly

c. implementation of the identified measures

d. use of scenario planning and trials with robust evaluation measures to ensure arrangements are effective and not reinforcing bias

e. alternative and diverse means of reaching excluded consumers including collaboration with other stakeholders to maximise transparency

f. implementation of mitigation measures including access to any necessary redress should consumers be treated unfairly

g. ensuring robust arrangements are in place to establish and maintain data privacy and compliance with the relevant regulation

h. testing and monitoring the effectiveness of the AI model, the implemented controls, and mitigation measures

i. any other measures necessary to reduce the risk and ensure consumers are treated fairly

AI in predictions and forecasting

6.8 AI can be used in creating predictions and has been notably used in forecasts, where the models can be used to complement existing models and improve predictions. Within the energy sector, examples include:

a. weather forecasting, such as renewable generation considering granular weather data and satellite imagery, or to predict storm damage to networks for use in planning responses to extreme weather events

b. predicting time to failure of equipment and expected maintenance requirements

c. predicting electricity usage at different granularities and on different time scales, supporting operations and planning

6.9 These predictions can provide useful insight into what is going to happen in the future. In a similar way to existing methods of prediction, AI also has its limitations around accuracy and uncertainty. However, AI can be used in conjunction with existing methods to manage these limitations.

6.10 Considerations include:

a. the role of governance in developing, implementing and overseeing the effectiveness of AI

b. undertaking a risk assessment which identifies control measures needed through the life cycle to ensure the AI system functions as defined and results in usable predictions

c. identifying the likelihood and consequences of any potential failure modes and maloperation, and any controls or mitigations necessary

d. implementation of the identified control measures

e. understanding of required accuracy, biases, and the available data to support the creation of the necessary models

f. appropriateness of the AI models in comparison with traditional models, and using a combination of multiple models and methods for prediction to explain AI model outputs: for example, situational use of AI models for specific conditions, or in combination with traditional models

g. providing clear explanations of the underlying methodology and factors influencing predictions to ensure the outputs of AI are proportionately explainable, transparent and trustworthy

AI in cyber-physical systems

6.11 AI can be integrated into the control of physical systems, and may provide opportunities to mitigate risk, or reduce operational costs. Examples include:

a. using robots and autonomous systems in hazardous environments, such as drones for inspections, can potentially reduce health and safety risks to individuals

b. AI automation controlling devices, such as battery storage, aligning generation and demand forecasts

c. AI in a cyber security context being used by network operators to monitor for potential cyber-attacks, launch risk and vulnerability assessments, and deploy automated defences. This includes situational awareness and triggering actions such as energy dispatch to mitigate risks and maintain system stability

6.12 If implemented poorly, these applications can result in harm to the systems they are supporting, the automation leading to maloperation that could damage both the device itself, and the systems it is connected to or operating with. This could include inappropriate shutdowns, potentially causing a cascading effect, or autonomous drones colliding with and damaging infrastructure.

6.13 Considerations include:

a. the role of governance in developing, implementing and overseeing the effectiveness of AI

b. undertaking a risk assessment which identifies control measures needed through the life cycle to ensure the AI system functions as defined and that in its operation, outcomes are as required

c. implementation of the identified control measures, for example using frameworks such as functional safety. This may include wraparound systems and guardrails, intended to ensure the system remains in a safe state

d. appropriate consideration of the potential impact of failure on a broader system to ensure any consequential impact of the failure is mitigated

e. ensuring the system containing AI is designed to take into account its complexities and interactions with humans

f. testing and monitoring the effectiveness of the AI model, the implemented controls, and mitigation measures

g. appropriate training of operators and overseers to identify incidents where the system is entering maloperation, and how to intervene

h. identification and implementation of any other measures necessary to reduce the risk

AI in pricing and trading

6.14 AI enables energy traders to make informed and profitable trading decisions by detecting market opportunities and risks. AI can analyse complex market

dynamics in energy trading by processing real-time data on pricing, demand and supply trends. AI can also assist in conducting risk management, by proactively assessing market volatility and uncertainty. Energy portfolios can be optimised by simulating market scenarios, analysing sentiment, automating tasks and continually adapting to changing market conditions.

6.15 The use of AI in pricing and trading has the potential to adversely impact competition. Where AI is used to fix bids, prices or margins, or to exchange commercially sensitive information between competitors, particularly pricing information, it could breach competition law. The use of AI, in certain circumstances, might also amount to the abuse of a dominant market position in breach of competition law. It is also necessary to be transparent to prevent any potential market abuse resulting from the use of AI. Further information on competition law compliance obligations is set out in Appendix 1.

6.16 The use of AI in grid management has the potential to influence the procurement and flexible distribution of energy to meet local needs. Stakeholders have a duty to be transparent about adjustments that are made. This requires the influence of AI to be explainable in a manner that is proportionate to the risk.

6.17 Therefore, it is important to consider:

a. the role of governance in developing, implementing, and overseeing the application of AI to ensure the system containing it is not operating in an anti-competitive manner

b. undertaking a risk assessment which identifies areas where the system may fail and identify measures that results in fair market outcomes

c. implementation of the identified measures

d. having appropriate oversight, monitoring and audit trail in place confirming that the AI system has been tested with a view to identifying potential breaches of competition law, and ensuring it is proportionately explainable to demonstrate compliance

AI in data analytics and consumer privacy

6.18 AI can process highly granular energy consumption data. This could provide tailored services, such as personalised energy-saving advice or automatically switching a consumer to another tariff. However, this data could reveal patterns of life, for example, when residents are home and their routines. Data could be classified as personal data under UK GDPR regulations and where this is the case controllers of the data, including for the purposes of AI, need to include the following in the privacy information:

a. the purposes for processing the data

b. the retention periods for the data

c. who the data is shared with

6.19 These risks are not new: existing frameworks including UK GDPR and associated regulatory guidance, already place clear limits on the collection, sharing and use of granular energy consumption data. The use of AI can increase the speed and scale at which such data analysed, increasing the importance of effective compliance and oversight.

6.20 Lack of transparency in AI systems can create information asymmetries between energy suppliers and consumers. In particular, opaque AI models may enable suppliers to derive detailed insights into household energy consumption patterns or infer domestic and socio-economic characteristics that are not readily visible to consumers themselves. Where consumers are not adequately informed about how such inferences are generated or used, there is a risk that these insights could influence pricing, tariff eligibility, or access to services without meaningful consumer awareness or understanding of the underlying decision-making logic.

6.21 Considerations should include:

a. organisations processing data using AI should provide an appropriate level of explanation of the AI they use. In line with ICO guidance on explaining decisions made with AI, organisations should also ensure that consumers are informed about use of data relating to them, which could include setting out the main factors informing AI-enabled decisions or outputs where these have a material impact on the consumer.

b. stakeholders should ensure that AI serves energy consumers fairly and transparently. This includes identifying and mitigating unintentional bias that could lead to direct or indirect discrimination

c. AI requires training data, and relevant data management and governance practices need to be in place to ensure existing data regulations are complied with. Data should be fit for purpose, accurate, and complete

d. strategies for the safe and fair use of AI should be driven at board and senior management levels. This includes undertaking a risk assessment to identify proportionate measures that result in fair and transparent outcomes

e. where legally required organisations must have appropriate measures in place to enable stakeholders to opt-in, opt-out and have the right to erase the use of their data

Use of black boxes

6.22 The terminology black box is often used to describe systems where it is not possible to understand and quantify how the system generates its output. Algorithmic transparency is defined as the degree to which the factors informing

the AI output, such as recommendations or decisions, are knowable by various stakeholders. These factors may include:

a. The model type, logic and decision rules

b. how the AI model has been trained and validation performed

c. what data and inputs the AI model is trained on

d. which key features or factors most influenced the decision of the AI's output

e. what decisions the AI would have made under different scenarios

6.23 AI, as with other forms of technology during their development, can be difficult to explain or understand. Such technologies can provide powerful capabilities, as seen with large language models (LLMs), and can be successfully adopted with the appropriate approach. However, they often have associated with them higher levels of uncertainty, which needs to be accounted for.

6.24 When using black box technology, the levels of increased uncertainty require users to mitigate the additional risk this creates. In addition, maloperation can occur, and be indistinguishable from normal operation without additional verification and validation, such as hallucinations in LLMs.

6.25 Considerations should include:

a. the role of governance in developing, implementing and overseeing the effectiveness of AI

b. undertaking a risk assessment which identifies control measures needed through the life cycle to ensure the AI system functions as defined and that in its operation, outcomes are as required

c. implementation of the identified control measures

d. testing can be used to increase the understanding of the characteristics of the AI system, including empirical assessments of bias and accuracy. Sufficient assurance of the system performance should be combined with appropriate control measures

e. it may be necessary to validate or corroborate the AI output against known good data, research or expertise

f. whether there is a proportionate level of explanation of the AI output that is understandable and transparency about when and where it is used

g. the residual uncertainty and its potential impact after mitigations must be considered by the stakeholder, and the organisation must decide as to whether it can tolerate the risk, and act accordingly

Glossary

A

Adaptivity: the ability to identify patterns, reason, and make decisions in contexts and ways not directly envisioned by human programmers or outside the context of a system's training data. (DSIT, 2024)

AI agents or agent, autonomous agent: AI systems that are capable of accomplishing multi-step tasks in pursuit of a high-level goal with little or no human oversight. AI agents may do things like browsing the internet, sending emails, or sending instructions to physical equipment. (DSIT, 2024)

AI deployers: any individual or organisation that supplies or uses an AI system to provide a product or service. Deployment can be internal, where a system is only used by the developers, or external, allowing the public or other non-developer entities to use it. (DSIT, 2024)

AI developers: organisations or individuals who design, build, train, adapt, or combine AI models and applications. (DSIT, 2024)

AI end user: any intended or actual individual or organisation that uses or consumes an AI-based product or service as it is deployed. (DSIT, 2024)

AI life cycle or **AI product life cycle**: all events and processes that relate to an AI system's lifespan, from inception to decommissioning, including its design, research, training, development, deployment, integration, operation, maintenance, sale, use, and governance. (DSIT, 2024)

AI risks: the combination of the probability of an occurrence of harm arising from the development or deployment of AI models or systems, and the severity of that harm. (DSIT, 2024)

Algorithm: a set of instructions used to perform tasks, such as calculations and data analysis, usually using a computer or another smart device. (UK Parliament POST, 2024)

Algorithmic bias: AI systems can have bias embedded in them, which can manifest through various pathways including biased training datasets or biased decisions made by humans in the design of algorithms. (UK Parliament POST, 2024)

Algorithmic transparency: the degree to which the factors informing general-purpose AI output, for example recommendations or decisions, are knowable by various stakeholders. Such factors might include the inner workings of the AI model, how it has been trained, what data it is trained on, what features of the input affected its output, and what decisions it would have made under different circumstances. (DSIT, 2024)

Alignment: the process of ensuring an AI system's goals and behaviours is in line with its developer's values and intentions. (DSIT, 2024)

Artificial intelligence (AI): describes computer systems which can perform tasks usually requiring human intelligence. This could include visual perception, speech recognition or translation between languages. (NCSC, n.d.)

Artificial general intelligence (AGI): a potential future AI system that equals or surpasses human performance on all or almost all cognitive tasks. A few AI companies have publicly stated their aim to build AGI. However, the term AGI has no universally precisely agreed definition. (DSIT, 2024)

Autonomy or autonomous: capable of operating, taking actions, or making decisions without the express intent or oversight of a human. (DSIT, 2024)

Automated decision-making: automated decision-making is the process of making a decision by automated means without any human involvement. These decisions can be based on factual data, as well as on digitally created profiles or inferred data. (ICO, 2022)

B

Black box: a system, device or object that can be viewed in terms of its inputs and outputs, without any knowledge of its internal workings. (ICO, 2023)

C

Capabilities: the range of tasks or functions that an AI system can perform and the proficiency with which it can perform them. (DSIT, 2024)

Cognitive tasks: tasks involving a combination of information processing, memory, information recall, planning, reasoning, organisation, problem solving, learning, and goal-oriented decision making. (DSIT, 2024)

Compute: computational resources, required in large amounts to train and run general-purpose AI models. Mostly provided through clusters of graphics processing units (GPUs). (DSIT, 2024)

D

Deep learning: a set of methods for AI development that leverages very large amounts of data and compute. (DSIT, 2024)

Deployment: the process of releasing an AI system into a real-world environment, such as a consumer-facing AI system. (DSIT, 2024)

Disinformation: deliberately false information generated or spread with the intent to deceive or mislead. (DSIT, 2024)

Guidance Ethical AI use in the energy sector (version 2)

E

Explainability: clear, accessible explanations tailored to audience and context, covering outcomes, processes, and ethical practices, enabling individuals to understand and challenge outputs. (The Alan Turing Institute, 2025)

F

Foundation models: machine learning models trained on very large amounts of data that can be adapted to a wide range of tasks. (DSIT, 2025)

Frontier AI: highly capable general-purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models. (UK Parliament POST, 2024)

G

Generative AI: an AI model that generates text, images, audio, video or other media in response to user prompts. It uses machine learning techniques to create new data that has similar characteristics to the data it was trained on. Generative AI applications include chatbots, photo and video filters, and virtual assistants. (UK Parliament POST, 2024)

I

Input (to an AI system): the data or prompt fed into an AI system, often text or an image, which the AI system processes before producing an output. (DSIT, 2024)

L

Large language model (LLMs): machine learning models trained on large datasets that can recognise, understand, and generate text and other content. (DSIT, 2024)

M

Machine learning (ML): the set of techniques and tools that allow computers to 'think' by creating mathematical algorithms based on accumulated data. (ICO, 2023)

Massive multitask language understanding (MMLU): a widely used AI research benchmark that assesses a general-purpose AI model's performance across a broad range of tasks and subject areas. (DSIT, 2024)

Memorisation: a phenomenon in which AI tends to memorise specific details from examples rather than learning general patterns, affecting model generalisation, security, and privacy. (Jiaheng Wei, 2024)

Misinformation: incorrect or misleading information, potentially generated and spread without harmful intent. (DSIT, 2024)

Model drift: where the domain in which an AI system is used changes over time in unforeseen ways leading to the outputs becoming less statistically accurate. (ICO, 2023)

Model poisoning: attacks in which the model parameters are under the control of the adversary. Model poisoning attacks attempt to directly modify the trained machine learning model to inject malicious functionality into the model. (NIST, 2024)

N

Narrow AI: an AI system that only performs well on a single task or narrow set of tasks, like sentiment analysis or playing chess. (DSIT, 2024)

O

Open-ended domains: scenarios or environments that have a very large set of possible states and inputs to an AI system, so that developers cannot anticipate all contexts of use, and thus cannot test the AI's behaviour in all possible situations. (DSIT, 2024)

Open source: open source often means the underlying code used to run AI models is freely available for testing, scrutiny and improvement. (UK Parliament POST, 2024)

P

Pre-training: the first stage of developing a modern general-purpose AI model, in which models learn from large amounts of data. Pre-training is the part of general-purpose AI training that requires the most data and computational resources. (DSIT, 2024)

Prompt injection: attacker technique in which a hacker enters a text prompt into a large language model or chatbot designed to enable the user to perform unintended or unauthorised actions. (NIST, 2024)

R

Risk factors: elements or conditions that can increase downstream risks. For example, weak guardrails constitute a risk factor that could enable an actor to malicious use an AI system to perform a cyber attack (downstream risk). (DSIT, 2024)

Responsible AI: often refers to the practice of designing, developing, and deploying AI with certain values, such as being trustworthy, ethical, transparent, explainable, fair, robust and upholding privacy rights. (UK Parliament POST, 2024)

S

Safety and security: the protection, wellbeing, and autonomy of civil society and the population. In this publication, safety is often used to describe prevention of or protection against AI-related harms. AI security refers to protecting AI systems from technical interference such as cyber-attacks or leaks of the code and weights of the AI model. (DSIT, 2024)

Shadow AI: a form of **Shadow IT**. Unknown AI assets, for example services and components, used by individuals within an organisation for business purposes. These assets are not authorised, governed, accounted for by asset management, nor aligned with corporate IT processes or policy. (NCSC, n.d.)

Shadow IT: refers to the unknown IT assets that are used within an organisation for business purposes. These assets are not accounted for by asset management, nor aligned with corporate IT processes or policy. (NCSC, n.d.)

Synthetic data: data, such as text and images, that has been generated artificially, for instance by general-purpose AI models. Synthetic data might be used for training general-purpose AI models, such as in cases of scarcity of high quality natural data. (DSIT, 2024)

T

Transfer learning: a machine learning technique in which a model's completed training on one task or subject area is used as a starting point for training or using the model on another subject area. (DSIT, 2024)

Transformer architecture: a deep learning architecture at the heart of most modern general-purpose AI models. The transformer architecture has proven particularly efficient at converting increasingly large amounts of training data and computational power into better model performance. (DSIT, 2024)

Training datasets: the set of data used to train an AI system. Training datasets can be labelled, for example pictures of cats and dogs labelled 'cat' or 'dog' accordingly or unlabelled. (UK Parliament POST, 2024)

Transparency in use: responsible, context-appropriate disclosure about the use of AI systems, including relevant data sources, decision logic, capabilities and limitations, without requiring disclosure of proprietary code or datasets. (OECD, 2024)

Use case: an AI application or problem an AI system intends to solve. (ICO, 2023)

W

Weights: parameters in a model that are akin to adjustable dials in the algorithm. Training a model means adjusting its parameters to help it make accurate predictions or decisions based on input data, ensuring it learns from patterns it has seen. (DSIT, 2024)

White box: a system deployed without restrictions such that a user can access or analyse its inner workings. (DSIT, 2024)

Glossary references

DSIT. (2024, May 17). *International scientific report on the safety of advanced AI: interim report*. Retrieved from https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai/international-scientific-report-on-the-safety-of-advanced-ai-interim-report

DSIT. (2025, April 28). *Frontier AI: capabilities and risks – discussion paper*. Retrieved from https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper/frontier-ai-capabilities-and-risks-discussion-paper

ICO. (2022, October 17). *What is automated individual decision-making and profiling*. Retrieved from https://ico.org.uk/media2/bibhz1gs/existing-guidance-automated-decision-making-and-profiling.pdf

ICO. (2023, March 15). *AI and Data Protection Glossary*. Retrieved from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/glossary/

Jiaheng Wei, Y. Z.-L. (2024, June 6). *Memorization in deep learning: A survey*. Retrieved from https://arxiv.org/pdf/2406.03880

NCSC. (n.d.). *Artificial Intelligence*. Retrieved from https://www.ncsc.gov.uk/section/advice-guidance/all-topics/artificial-intelligence

NCSC. (n.d.). *Shadow IT guidance*. Retrieved from https://www.ncsc.gov.uk/guidance/shadow-it

NIST. (2024, Jan). *Adversarial Machine Learning, A Taxonomy and Terminology of Attacks and Mitigations*. Retrieved from https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2023.pdf

OECD. (2024, May). *Transparency and explainability (Principle 1.3)*. Retrieved from OECD AI Policy Observatory: https://oecd.ai/en/dashboards/ai-principles/P7

The Alan Turing Institute. (2025). *AI Explainability in Practice*. Retrieved from AI Ethics and Governance in Practice: https://aiethics.turing.ac.uk/modules/explainability/?modulepage=part-one-introduction-to-ai-explainability

UK Parliament POST. (2024, Jan 23). *Artificial Intelliegence (AI) Glossary*. Retrieved from https://post.parliament.uk/artificial-intelligence-ai-glossary/

Send us your feedback

We are keen to receive your feedback about this guidance. We would also like to get your answers to these questions:

- Do you have any comments about the quality of this guidance?
- Do you have any comments about its tone and content?
- Was it easy to read and understand? Or could it have been better written?
- Do you have any further comments?

Please send your feedback to stakeholders@ofgem.gov.uk.

Appendices

Index

Appendix 1 Legal and regulatory obligations

Introduction

A1.1 Great Britain's energy sector is governed by a layered regulatory framework with components that reside in binding legislation, regulations, licences and energy industry codes. Complementing these are codes of practice and guidance documents. Currently, we consider the existing regulatory framework to be appropriate to govern the use of AI, complemented and supported by this guidance document, which sets out good practices for stakeholders to consider when deploying or procuring AI.

A1.2 There are competition law risks associated with the deployment of AI and stakeholders are responsible for ensuring their use of AI complies with the Competition Act 1998.

A1.3 Our regulatory toolkit includes various enforcement powers and actions where the use of AI breaches relevant licence conditions and other relevant requirements, and the competition rules.

Ofgem's statutory remit

A1.4 Ofgem's key statutory functions are underpinned by:

a. Gas Act 1986

b. Electricity Act 1989

c. Competition Act 1998

d. Utilities Act 2000

e. Enterprise Act 2002

f. Energy Acts of 2004, 2008, 2010, 2011, 2013, 2016 and 2023

g. The Gas and Electricity (Consumer Complaints Handling Standards) Regulations 2008

h. Electricity and Gas (Market Integrity and Transparency) (Enforcement etc.) Regulations 2013, known as REMIT – the Regulation on Wholesale Energy Market Integrity and Transparency

i. Network and Information Systems Regulations 2018

j. Domestic Gas and Electricity (Tariff Cap) Act 2018

A1.5 Stakeholders are also required to comply with other areas of legislation that apply to the energy sector and AI use, including data protection, equality, human rights, and health and safety.

A1.6 Following the publication of the Ofgem Review in April 2026, Ofgem is looking into how it can strengthen the application of its existing statutory duties, with a greater focus on outcomes and regulating an increasingly complex and innovative energy sector.

Regulatory framework components

A1.7 Certain activities within the energy sector cannot be carried out unless a person has a licence. Ofgem has powers to grant those licences. Licences may be granted to suitably qualified operators for the purposes of engaging in specified activities within the energy sector, such as supply, smart meter communication, distribution, transportation, transmission, generation and interconnection. The licences list the conditions that all licensees must abide by to engage in the specified activities. Licence conditions can be prescriptive or principles-based and licences may contain both. Prescriptive conditions tend to be detailed and specific, identifying how licensees must achieve a certain outcome. Principles-based conditions have more general requirements, such as 'to treat customers fairly', which places the onus on licensees to determine how compliance should be achieved. This approach also provides opportunities for the licensee to innovate so long as the risks are appropriately managed.

A1.8 Under REMIT, which is a regulatory framework specific to wholesale energy markets, market participants are not required to be licensed but are required to register before trading wholesale energy products.

A1.9 Under their respective licences, licensees are required to maintain, become party to, and, or comply with energy industry codes. These are detailed multilateral agreements that define the terms under which licensees can access the electricity and gas networks, and the rules for operating in energy markets.

A1.10 We also publish other information to help licensees understand and comply with their obligations. This includes guidance documents and decisions from investigations.

A1.11 Licensees and regulated persons are reminded that they are required to comply with their standard licence conditions (SLCs), energy industry codes and other obligations. For licensees this includes Treating Consumers Fairly (for example SLCs 0 and 0A of the supply licence, SLC 10AA of the distribution licence) and Operational Capability (for example, SLC 4A of the supply licence). Similarly, market participants are required to comply with their REMIT obligations and designated operators of essential services are required to comply with their obligations under the Network and Information System Regulations 2018.

Ofgem's competition law functions

A1.12 Ofgem is a competition authority, and the use of AI has the potential to adversely impact competition. Ofgem is required to act in a manner we consider will best

further our guiding principal objective of protecting consumers' interests by promoting effective competition in the activities we regulate wherever appropriate. Ofgem also has Competition Act 1998 duties and powers including the power to conduct dawn raids and to issue notices requiring the recipient to provide specific documents or information.

A1.13 The Competition Act 1998 contains two prohibitions: the Chapter I prohibition relates to anti-competitive agreements such as price fixing, bid rigging or market sharing. The Chapter II prohibition relates to the abuse of a dominant market position. The inappropriate use of AI may breach the Chapter I prohibition if the effect of the AI programme is to fix bids or prices or margins or to exchange commercially sensitive information, particularly pricing information. The inappropriate use of AI may, in certain circumstances, also breach the Chapter II prohibition.

A1.14 It is stakeholders' responsibility to ensure that their use of AI technology complies with the Competition Act 1998. Stakeholders are responsible for the AI tools they choose to deploy or procure. Stakeholders, including IT Directors and IT Managers, can bear personal responsibility for a breach of competition law and will therefore need to ensure that their AI system does not produce anti-competitive effects. Stakeholders should have an appropriate audit trail in place confirming that they have checked their use of AI for compliance with competition law. In this context it is vital stakeholders can explain how they have minimised the likelihood of the system containing the AI resulting in an adverse impact on competition. In addition, stakeholders should be able to assure themselves, and competition authorities, that the system containing AI is not operating in an anti-competitive manner.

Enforcement powers

A1.15 This guidance has been developed to enable stakeholders to use AI in an appropriate and risk informed manner. However, under existing enforcement powers Ofgem can impose financial penalties of up to 10% of a regulated person's turnover in sectoral matters. Ofgem can also make consumer redress orders and issue provisional and final orders, where appropriate, for breaches of relevant licence conditions and other relevant requirements (and certain other provisions) under the Gas Act 1986, the Electricity Act 1989 and certain other regulations or legislation with which licensees or other regulated persons must comply.

A1.16 Under the Competition Act 1998 we can impose penalties up to 10% of global turnover on companies for competition law infringements. We can also apply to the court to disqualify directors for a maximum of 15 years who are involved in breaches of competition law where their conduct would make them unfit to be a director in a company.

A1.17 There should be nothing in the guidance to excuse or prevent licensees or other regulated persons from complying with their obligations under licence conditions and other rules. Where licensees use AI, they must nevertheless comply with these obligations. Any breach of those obligations derived from the use of AI does not excuse or mitigate breach. Therefore, licensees or other regulated persons must ensure they have and maintain appropriate processes, systems and governance to ensure AI does not cause them to breach their regulatory obligations.

A1.18 Further details on our enforcement policies and processes are set out in our Enforcement Guidelines.

Conflict and hierarchy

A1.19 In the event of any conflict between this guidance and:

a. Ofgem's statutory remit

b. regulatory framework components

c. Ofgem's competition law functions

A1.20 the hierarchy of legal frameworks referred to above shall prevail. Also, it is important to note that this guidance does not constitute legal advice, and therefore stakeholders are assumed to seek their own advice as required.

Appendix 2 AI standards

A2.1 AI standards and frameworks continue to be developed to support the ethical use of AI. Standards are agreed ways of doing something. These cover, for example, how to make a product, manage a process, or deliver a service. Standards are open, consensus-based guidance that represent good practice. Standards are produced by national and international standards bodies (such as the British Standards Institute in the UK and the International Organisation for Standardisation). These product and quality standards are outside of Ofgem's remit but can be used by stakeholders to manage risks associated with AI use. Ofgem may take adherence to standards and frameworks into account when assessing compliance with legal and regulatory obligations.

A2.2 Read the standards listed on the AI Standards Hub.

Appendix 3 AI supply chain management

A3.1 There is a well-established principle that, regardless of the arrangements for procured services and components, an organisation cannot outsource responsibility or accountability for safety, security (including data privacy and confidentiality), fairness or sustainability.

A3.2 Stakeholders should ensure effective supply chain management is in place to deliver safe, secure, fair, and sustainable use of AI. Risks associated with the use of AI should be appropriately allocated in contractual arrangements along the supply chain, with legal responsibility and liability clearly articulated. These arrangements may also include the following measures to help manage the risk:

a. ensuring that AI developers have robust arrangements in place to build confidence in the development of safe and responsible AI

b. effective and auditable training of AI systems

c. appropriate testing

d. arrangements for anonymisation of data, governance, triangulation of data risk management (for example, drawing correlations between data sets that are misleading) and incident reporting

e. data provenance records and adherence to commercial data assurance

A3.3 Users of AI may have limited input into the AI system design and development and be heavily reliant on guidance, policies, copyright statements and other materials and communication issued by AI developers. The supply chain, for both external AI services and internal AI-based components and systems, should be understood and should have sufficient transparency, traceability, validation and verification processes in place to ensure the desired outcome is achieved. The supply chain has an important role to play to assist users by explaining the output of AI systems at a level proportionate to the risk. Users should be able to access information regarding what and where the AI is used, its characteristics, any failure modes and necessary risk control measures needed (including cyber security risk control) and how to identify issues such as incorrect or misleading results (often referred to as hallucinations).

A3.4 The supply chain for AI systems adds additional complexity to developing secure systems, for example exacerbating data issues. NCSC identify securing this supply chain as the first point in their secure development section of guidelines for secure AI system development. This challenge is similar to that faced with the cloud. Due to the supply chain challenges, NCSC produced guidance, stating 'the service provider should ensure that its supply chain meets the same security standards that the organisation sets for itself. If this principle is not implemented, supply chain compromise can undermine the security of the service and affect the implementation of other security principles.

A3.5 AI can also present new vulnerabilities and threat vectors due to their complexity, rapid development and the potential opacity of the supply chain. Novel vulnerabilities and cyber security threats with AI include:

a. training data and model poisoning: injection of malicious or corrupt data

b. prompt injection: attackers feed the AI malicious prompts

c. model memorisation: models remembering specific detail

A3.6 Stakeholders should ensure that:

a. where data is exchanged with externally developed, hosted and operated services there is a clear and thoroughly documented understanding of the data flows, data use, data storage and retention

b. supply chain managers address relevant technical aspects of procured services and components, and these are reflected in commercial arrangements, including data ownership, privacy, confidentiality and cyber security

c. the supply chain management arrangements are continually reviewed and updated to reflect the pace of development

Appendix 4 Data use and management

A4.1 This appendix does not cover cyber security concerns (see Appendix 5).

A4.2 In addition to the potential uncertainties associated with AI, the accuracy, availability and completeness of data used to train AI system components is likely to impact the usefulness of the AI materially.

A4.3 AI is also likely to exacerbate issues currently experienced regarding the collection, storage and use of data. AI requires training data, whether purchased from a third-party supplier, tuned, or trained internally. It is essential that the relevant data management and governance practices are in place to ensure that this can be carried out appropriately, and existing data regulations are complied with. For example, during the creation of any AI, it is necessary to consider GDPR and security classification of the data. Currently, practical and robust methods for removing data from AI are not available. Due to this, it is necessary to consider data removal prior to the creation, or purchase of any AI. The implications of the right to erasure will need appropriate measures in place to ensure that this obligation can be met. This could be achieved by using data not subject to GDPR erasure considerations or putting in place appropriate processes for safely and securely deleting the AI and entries from the base dataset, before training the AI on the new dataset. More information about right to erasure can be found on the ICO's website.

A4.4 Ofgem's Data Best Practice Guidance aims to create interoperability in data across the energy sector to maximise its value. Best practice is intended to be used to mitigate risks associated with data use and management. This relates to input data, the output of AI and data used for training purposes. It is important to remember that the AI model itself is a data asset in its own right. Stakeholders should ensure:

a. appropriate data governance is in place, covering the entire data life cycle from collection to disposal. Data used to train and operate the AI is fit for use for the application, accurate and complete

b. data used to train AI models is appropriately prepared. Relevant steps are taken such as to remove biases, and address unfairness where it is captured in the data. This should include versioning datasets, recording weaknesses and assumptions

c. training and operational data is appropriately managed (including the AI), and secured, with appropriate consideration to data accumulation risks, for example large or combined datasets becoming more sensitive than their components. This should include access to and use of AI models, with arrangements in place to protect the AI as both a software and data asset

d. data collection methods and associated data generated are appropriately secured against intentional and unintentional compromise, through the relevant measures, for example physical and technical security at weather stations

e. appropriately secure methods are in place for the disposal of both AI systems, and the data processed, with the relevant governance and assurance in place

f. appropriate monitoring and validation arrangements are implemented to identify shifts in behaviour, for example model drift

Appendix 5 AI and cyber security

The use of AI and its impact on cyber security

A5.1 Given the rapidly evolving nature of AI-related cyber security threats, stakeholders should ensure that their cyber security practices remain under review and take account of relevant standards, principles, guidance, frameworks and threat publications cited in this appendix as AI technologies and associated risks evolve.

A5.2 The application of AI has the potential to improve resilience to cyber attack. For example, intruder detection by monitoring for subtle changes in system performance. However, the uncontrolled, unauthorised and ungoverned risk of any technology, particularly emerging technologies, can cause exposure to an unknown number of existing and emerging cyber security threats and vulnerabilities. The rapid proliferation of AI technologies, the ease of availability of AI services particularly Generative AI, the widespread interest in their use, and that AI has such a broad array of uses may lead to AI being used without knowledge or authorisation. This rapid proliferation of emerging technology is not a new risk and has been observed before when, for example, internet access and cloud computing were made commonly available at the office desktop.

A5.3 To maintain effective cyber security hygiene, stakeholders should ensure AI is used in a controlled manner. A key element for robust cyber security hygiene is that the stakeholder is always aware what technology is being used, for what purpose and where it is being used, and that it is being used in a controlled manner by appropriately authorised staff and systems.

A5.4 AI technology is a continuation of development of broader digital technologies including hardware, software and data, each of which have associated cyber guidance and good practice. Good practice should be applied with the same consideration and rigour, based on applied risk analysis and assurance methodologies. This includes:

a. NCSC Secure Design Principles

b. NCSC Secure Development and Deployment Guidance

c. NCSC Cloud Security Guidance

d. NCSC Cyber Assessment Framework

e. ICO GDPR Guidance

f. Ofgem Data Best Practice Guidance

A5.5 In addition to established guidance and good practice, it is necessary to consider the novel and exacerbated security risks that are AI-specific due to the unique behaviours and characteristics of AI based systems and processes. This

good practice should be applied in a manner appropriate to the role of the stakeholder within the energy sector, the nature and criticality of the systems and processes where the technology is being used, and the risk and threat assessment and other non-functional needs of the use case being developed.

A5.6 NCSC continues to assess the cyber security impact of AI. NCSC has captured this in a case study, issued specific Machine Learning Principles, and guidelines on secure AI system development. In addition, the ICO has published how security and data minimisation should be assessed in AI systems.

A5.7 The suitability of each stakeholder's established cyber risk management approach should be assessed. This includes identifying, assessing, mitigating and managing threats and risks to their organisation and customers associated with AI use.

A5.8 To manage cyber security risks associated with the use of AI, stakeholders are also expected to consider:

a. AI's impact on existing cyber security, risk management and incident response arrangements

b. the need for any additional controls to prevent the unauthorised or uncontrolled use of AI, for example, preventing the use of shadow AI through additional education and technical controls

c. the effectiveness of governance and accountability arrangements (see Governance and policies)

d. competency within the stakeholder's organisation (see Competencies)

e. supply chain management (see Appendix 3)

A5.9 Regarding the AI supply chain, stakeholders should apply appropriate mitigations based on the identified and assessed risk, including potential emerging threats and vulnerabilities, and assessed impact on services and assets, such as recommended in the following guidance:

a. NCSC Basic Risk Assessment and Management Method

b. NCSC Cloud Security Guidelines

c. NCSC Supply Chain Security Principles

d. NCSC Machine learning principles: Securing your supply chain principles

A5.10 It should be noted that standards continue to be developed to assist in the secure application of AI, such as ETSI TS 104 223.

Managing offensive AI threats

A5.11 There is evidence that the emergence of AI is increasing the volume, sophistication and effectiveness of existing cyber attack tactics. This can also change the threat landscape with new AI-based attacks, vulnerabilities and opportunities for cyber threat actors.

A5.12 NCSC has published analyses of the impact of AI on the cyber threat landscape, including an assessment of threat to 2027. These assessments provide an overview of AI's potential uses and impact as a tool for cyber attack.

A5.13 Stakeholders should update their threat intelligence gathering to include AI and consider increasing the frequency of review and breadth of intelligence sources. This will enable an organisation's cyber risk assessment and associated controls to take account of AI-related threats.

A5.14 It will also be necessary for stakeholders to ensure that cyber security management and the technical team's skills, knowledge and capability are kept up to date with the latest developments in AI and associated threat intelligence.

A5.15 NCSC offers guidance on intelligent security tools for stakeholders including defensive AI-based security tools in security posture. The consequential impact of the loss of internal security knowledge and skills from the introduction of such tools should be considered.

Appendix 6 Transparency and explainability

A6.1 The introduction of AI in the energy sector should not impact negatively on the protection of the interests of current and future energy consumers. Stakeholders need to maintain compliance with their existing obligations by protecting consumers' interests and treating them fairly. Transparency and explainability are vital to fulfilling this objective.

A6.2 Transparency in AI refers to a commitment by AI actors to responsible disclosure regarding their use of AI. This includes the proportionate means of:

a. providing meaningful information, appropriate to the context and consistent with the state of the art

b. fostering a general understanding of AI systems, including their capabilities and limitations

c. making stakeholders aware of their interactions with AI systems (for example, chatbots, automated decisions) in a proportionate way

d. offering plain and easy-to-understand information about:

i. the sources of data and, or input

ii. the factors, processes, and logic that led to a prediction, recommendation, or decision

e. enabling those affected by an AI system to understand and challenge its output

f. disclosing when AI is being used in line with the importance of the interaction

g. explaining how an AI system is developed, trained, operates, and deployed

h. facilitating public, multi-stakeholder discourse to build awareness and trust

A6.3 Transparency does not necessarily require disclosure of proprietary code or datasets, which may be too complex or protected by intellectual property rights.

A6.4 Explainability is not just about technical transparency but also includes communicability – providing clear and accessible explanations that are appropriate to the audience and context. This includes proportionate means of:

a. explaining outcomes of algorithmic models (for example, decisions or recommendations)

b. explaining processes by which models and systems are designed, developed, deployed, and decommissioned

c. ensuring explanations reflect ethical and responsible practices, including data integrity and protection

A6.5 Therefore, stakeholders should ensure transparency and explainability regarding the use of AI in a proportionate way including:

a. being transparent about the processing of personal data in an AI system, and complying with any obligations towards consumers whose personal data that they plan to process

b. ensuring that individuals are informed from the outset about who is collecting their data, as well as the reasons and methods for processing it. This clarity enables individuals to make informed choices about whether to engage

c. ensuring that AI systems are explainable by providing clear and accessible explanations appropriate to the audience and context. This includes communicating how decisions or recommendations are reached, outlining the processes by which systems are designed and deployed, and evidencing that ethical and responsible practices have been followed which result in outcomes that are safe, secure and fair

d. explainability supports transparency but goes further by enabling those affected to understand and, where necessary, challenge system outputs