



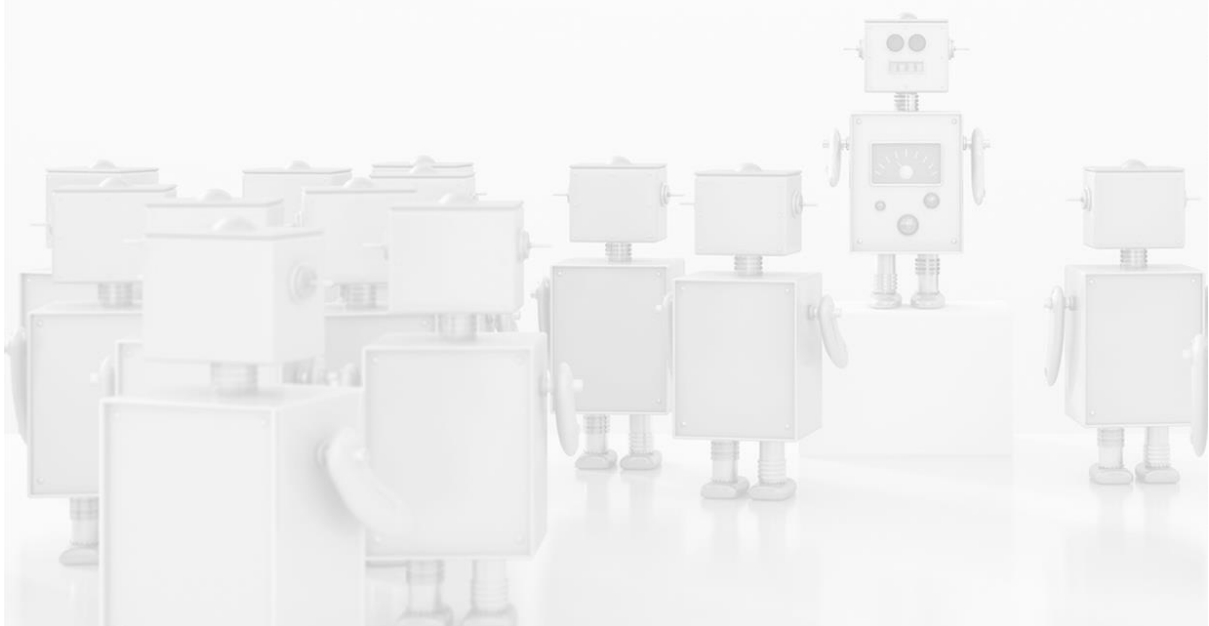
Department of
**Agriculture, Environment
and Rural Affairs**



DAERA Digital Services Division

Artificial Intelligence Policy:

“Securely Maximising the Benefits of Artificial Intelligence”



Document Information

Document details

Document name:	DAERA Artificial Intelligence Policy
Document version:	1.2
Document date:	12 th March 2026
Document status:	Updated
Author:	NAME REDACTED
HPRM reference:	REDACTED

Revision History

Version	Date	Summary of Changes	Author
1.0	30 October 2024	First finalised version	NAME REDACTED
1.1	4 April 2025	Figure 2 replaced for clarity	NAME REDACTED
1.2	12 March 2026	Annual review of policy: Clarity added regarding models, data residency, and life cycle management in sections 3.5, 3.8 & 6.3	NAME REDACTED

Document Review

Date of Next Scheduled Review
Q1 2027

Distribution

Name	Title / Department
Conor O'Donnell	Digital Director, DAERA
NAME REDACTED	Deputy Digital Director, DAERA
NAME REDACTED	Data Protection Officer, DAERA
AI Policy working group	DAERA
Enterprise Design Authority	Digital Services Division, DAERA
NAME REDACTED	TUS

Approval

Name	Position	Signature	Date
Conor O'Donnell	Digital Director, DAERA		12 March 2026

Table of Contents

1	EXECUTIVE SUMMARY	4
1.1	Key Take-Aways	4
2	INTRODUCTION	5
2.1	Artificial Intelligence Definition	5
3	PRINCIPLES AND GUIDANCE	7
3.1	Knowledgeable: Know what AI is and what its limitations are	7
3.2	Ethical Responsibility: Use AI lawfully, ethically, and responsibly	7
3.3	Data Security: Know how to keep AI tools secure	8
3.4	Accountability: Have meaningful human control at the right stage	9
3.5	Lifecycle Management: Understand how to manage the full AI lifecycle	9
3.6	Tooling Choice: Use the right tool for the job	10
3.7	Collaboration: Collaborate with a variety of others from the start	10
3.8	Assurance: Use principles alongside organisation's policies and assurance in place	11
4	AI PRODUCT USAGE POLICY	11
5	EXCEPTIONS POLICY	12
6	SOLUTION AND TOOLING GOVERNANCE	12
6.1	Inhouse Development	12
6.2	AI Products on Service Catalogue	13
6.3	Third Party Solutions and Procurement	13
7	APPENDICES	15
7.1	Governance Context	15
7.2	Risk Assessment	17
7.3	Benefits Analysis	24
7.4	Glossary of Terms	26
7.5	References	28

1 Executive Summary

This policy provides high-level principles and guidance to enable the consistent, effective, and secure use of Artificial Intelligence (AI) across DAERA. Given recent advances in both the use and abilities of AI, it is important to ensure that the Department can leverage the benefits of AI, while ensuring responsible and ethical use of this technology. At a high-level this policy covers the following four points that all DAERA staff and contractors must follow.

1. **Seek and Leverage AI Opportunities.** DAERA's policy is to encourage the responsible use of AI where this can improve delivery against our strategic objectives, policy development and regulatory responsibilities. Guidance on the potential benefits of AI is provided in Section 7.3.
2. **Adhere to AI Principles.** Section 3 provides high-level guidance and principles that must be followed for all use of AI within the department. These are based on UK Government guidelines and are designed to ensure that the Department can realise the benefits of AI in an ethical and responsible way.
3. **Follow Department Standards.** Adhere to the Department's IT procurement, design, implementation, and operation processes and standards for AI-enabled systems. These enable effective and responsible use of AI in line with the Department's risk appetite, while safeguarding our data and meeting our obligations to the public. This policy will also inform how we maintain and update our processes and standards so that these remain effective and relevant as advances in AI technology develop. *Any AI enabled technology that is used or introduced without the appropriate ICT governance will be deemed in breach of this policy.*
4. **Stay Informed on Legislation.** As the use and abilities of AI continue to grow, legislation and regulation in relation to AI is an evolving landscape and therefore it is important to stay informed on changes to law as these develop, and abide by all the regulatory and legislative obligations, including any foreign or international law where these apply to DAERA. Section 7.1, outlines existing key standards and legislation relevant to AI, but note that this is not exhaustive and will evolve.

1.1 Key Take-Aways



Figure 1 - Young robot reading
(Image generated by GenAI from text prompt)

- AI is not one thing but a suite of possibilities.
- AI is not the answer to every, or even most, challenges especially in the public sector. When addressing a need, both AI and non-AI solutions should be considered, and the most appropriate solution should be selected.
- Familiarise yourself with what can be done, what can't be done, and carry out a risk assessment.
- Treat AI like a new entrant graduate – eager, well-read, somewhat naïve, but not trusted to make important decisions.

2 Introduction

With the rapid growth of AI in all sectors, especially with Large Language Models (LLM) and Generative AI (GenAI), DAERA needs a policy which provides guardrails for the use of all forms of AI, including (but not limited to) Generative AI.

The challenge and purpose of the policy is to enable the use of AI in a manner that is safe, ethical and keeps data secure.

Policy objectives:

- assist in decision making
- provide guidance and references for research
- inform users of acceptable use principles for AI
- Provide a governance framework to manage AI risks

This policy was developed following extensive engagement with industry, multi-disciplinary workshops, conferences, and online research. Given the fast pace of AI advancements, it will be reviewed regularly.

In the future, it is likely that many ICT solutions will contain some elements of AI and therefore this policy should be taken into consideration when any new ICT solution is being planned or procured, and even for browser-based tools that do not require procurement. This policy will still apply for end user responsibility.

2.1 Artificial Intelligence Definition

According to the Gov.uk white paper, there isn't a universally agreed-upon definition of AI. Instead, AI is defined based on two key characteristics:

1. **Adaptivity:** AI systems learn by identifying patterns in data, which may not be obvious to humans. They can develop new ways of making inferences that their programmers didn't explicitly design.
2. **Autonomy:** AI systems can make decisions independently, without ongoing human control, which can complicate assigning responsibility for their outcomes.

2.1.1 Industry Scope

This policy covers all use of AI by or on-behalf of DAERA and its Arm's Length Bodies with the following exceptions:

- Use of AI in CAFRE for educational purposes. While the guidelines set out in this policy may be used to inform use of AI in CAFRE, use of AI for educational purposes should abide by all relevant education regulation and policy.
- The use of statistical techniques for the purposes of creating scientific reports, publications, and other official statistics.
- Any other exceptions should follow the exceptions process outlined in Section 5.

2.1.2 Technology Scope

The diagram below attempts to highlight, from a technology point of view, what is in scope for AI and what is not in scope. This is not exhaustive as the field of AI is fast changing at the time of writing. However, within the scope of this policy, AI refers to any machine-based system that can automatically generate predictions, content, recommendations or decisions using any of the following techniques.

1. **Logic and knowledge-based systems** of AI encode knowledge and solve problems using logical statements and rules. Example applications include proving mathematical theorems, fault diagnosis and playing chess.
2. **Machine Learning systems** learn and improve from examples without all its instructions being explicitly programmed and can extrapolate knowledge to unfamiliar situations. Examples of this include computer vision, pattern detection, LLMs and GenAI.
3. **Optimization and Statistical Techniques** includes predictive statistics, time series analysis, and probability theory. However, simple descriptive statistics that summarize data without making predictions are not considered AI and are excluded from this policy.

Any software system that does not leverage any of the above techniques and is not used to automate or recommend a decision, classify data, or predict an outcome is generally not considered AI and is therefore not covered by this policy.

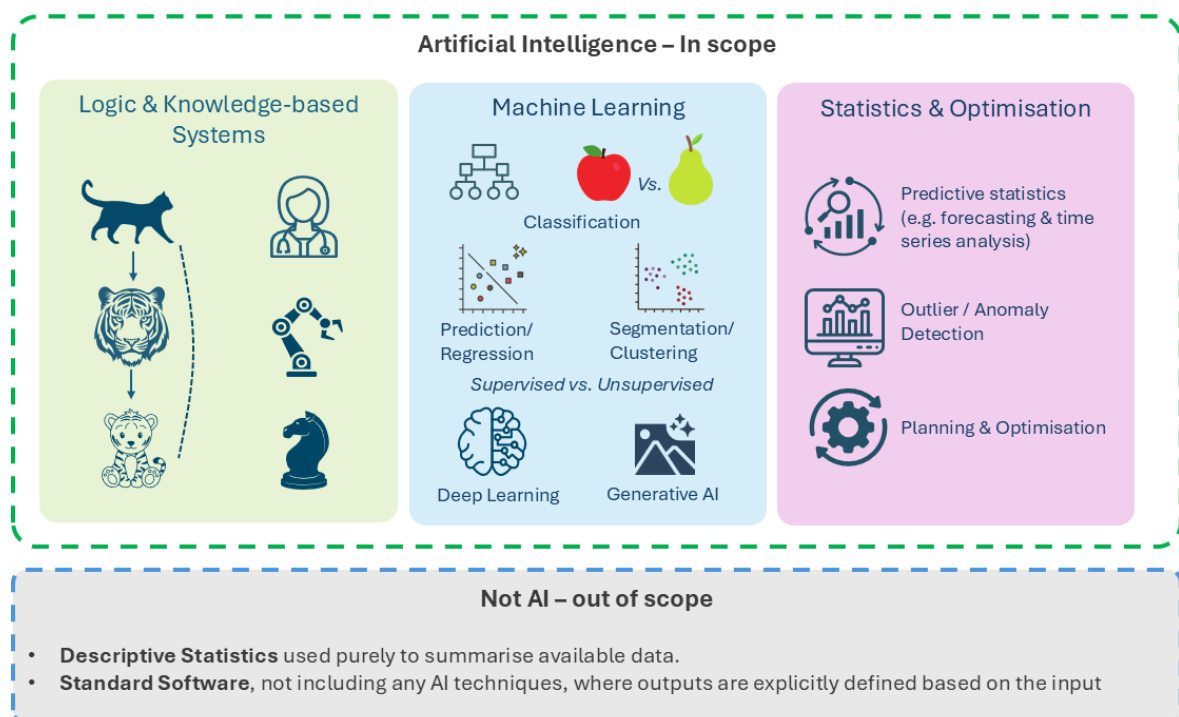


Figure 2 -AI In & Out of Scope

3 Principles and Guidance

The following principles are based on UK Government Generative AI Framework and are used to provide guidance when it is difficult to say in advance which way to go. They help you make decisions when there is more than one viable option, unlike rules that are followed as the right direction is already understood. The purpose of this policy is to empower you make the right decisions and guide behaviour in pursuit of your AI objectives.

Enterprise Architecture Principles apply for AI as to any other ICT system design – <http://nics.intranet.nigov.net/daera/articles/enterprise-architecture-principles>

3.1 Knowledgeable: Know what AI is and what its limitations are



Description: There are a number of types and uses of AI, and the technologies and limitations need to be understood to ensure effective choices are made and how it is put to use. You should know that AI requires an understanding of new skills such as prompt engineering¹ and you, or your team, should have the necessary skill set.

Impact: This impacts decisions on readiness to proceed with a project or end-user readiness to interact with specific AI tools. It is applicable whether your role involves designing, building, procuring, or using solutions containing AI. End users should understand how to use and interpret the results of the system correctly.

Actions:

- Actively learn about AI technologies, their capabilities, benefits, and risks.
- Participate in civil service training courses on AI.
- Stay updated on the latest AI developments and guidelines.

Governance: Show evidence of due diligence to understand the effects and limitations on new projects or products containing AI functionality. Show evidence of end-user training being completed prior to active use. Complete specified AI awareness and security training with adoption rates/performance tracked using the MetaCompliance training platform.

3.2 Ethical Responsibility: Use AI lawfully, ethically, and responsibly



Description: AI models are trained on large data sets, which may include biased or harmful material, as well as personal data. Biases can be introduced throughout the entire lifecycle, and you need to consider testing and minimising bias in the data at all stages. Provide transparency to the end-user by stating that the solution includes AI and how it works. AI solutions should augment human capabilities and provide beneficial outcomes for people and the planet.

Impact: This impacts decisions on how operational processes and procedures will be carried out during the lifetime of the solution.

Actions:

¹ Prompt engineering is the process of structuring an instruction that can be interpreted and understood by a generative AI model.

- Engage with compliance professionals, such as data protection, privacy, and legal experts in your organisation early in your journey;
- Seek legal advice on intellectual property equalities implications, and fairness and data protection implications for your use of AI;
- Ensure the use of high-quality training data, which is as accurate and representative as possible, and that biases and distortions are understood and addressed both on a technical and business level;
- Put in place procedures to address ethical concerns and check the model still complies with objectives and responsible use throughout its lifetime;
- Provide meaningful information to appropriate stakeholders on the product accuracy, limitations, decision making, and how to override results, appropriate to the context of the AI solution.

Governance: Show evidence of operational processes and procedures in place to test and regularly retest for unethical or other inappropriate responses. Information Asset Owners need to be aware of the types of AI deployed in their business area and ensure that the required Data Protection Impact Assessments <http://nics.intranet.nigov.net/daera/documents/data-protection-impact-assessment-dpia> are completed. This will ensure that personal data processed will have a lawful basis and compliant with DAERA's Privacy Statement <https://www.daera-ni.gov.uk/daera-privacy-statement>.

3.3 Data Security: Know how to keep AI tools secure



Description: AI tools can consume and store sensitive government information and personal identifiable information if the proper assurances are not in place. When using AI tools, you need to be confident that your organisation's data is held securely, and that the AI tool can only access the parts of your organisation's data that it needs for its task.

If sensitive data is provided as input to an AI system, or used to train an AI model, there is a risk that sensitive information may be provided by the system's output. Input and training data can also be manipulated maliciously to influence the system's behaviour in an unethical way. Safeguards should therefore include appropriate controls over access to AI inputs, outputs, and training data.

Impact: This impacts decisions on the data that is used to train a model and/or that it has access to during its working life.

Actions:

- Do not provide AI models with personally identifiable information (PII) without an explicit lawful basis for processing. In addition, do not provide AI models with any PII, non-public, proprietary, or confidential information, without the knowledge or consent of the Information Asset Owner of the data in question.
- Ensure all work to develop and integrate AI tools follows the DAERA Repeatable Design Process which addresses security by design and ensures all relevant and necessary security controls, including access controls, are built in.

- Perform thorough 3rd party product due diligence before implementing AI solutions, ensuring that any security and ethical implications are identified and evaluated. You should ensure adherence to best practice and industry standards. See section 6.3.
- Report any unusual or unexpected AI output, or suspected security risks, immediately through standard incident reporting procedures.²

Governance: Follow existing DAERA assurance/governance processes (Refer section 3.8). The Information Asset Owner (IAO) is the approver and must be aware of security, technical and organisational measures, as well as how data is being used.

3.4 Accountability: Have meaningful human control at the right stage



Description: Organisations and individuals developing, deploying, or operating AI systems are held accountable for their proper functioning and use. This requires quality assurance processes with trained and qualified personnel reviewing AI outputs and validating decision-making the AI outputs have fed into. In cases where human review isn't possible, such as with a chatbot, the product must be fully tested prior to deployment and provide mechanisms for end-user feedback to trigger a human review.

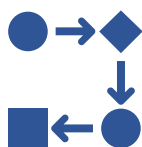
Impact: This impacts decisions on ownership from project initiation and throughout the life of the AI solution.

Actions:

- Identify an accountable person from the start, such as the Information Asset Owner (IAO).
- Implement a human-in-the-loop to assess outputs from AI.
- Ensure all users understand their responsibilities in using AI.

Governance: Show evidence of the accountable person taking responsibility for assurance of outcomes.

3.5 Lifecycle Management: Understand how to manage the full AI lifecycle



Description: You and your team must have appropriate knowledge of the AI tool's full lifecycle from setup to operational maintenance and upgrade through to offboarding. As the input or request to an AI system can vary over time, as can the patterns learned by the system and used to generate output, risks can develop and accumulate during use. Model suitability and update cadence must be periodically reviewed and documented in line with guidance.

Impact: This impacts decisions on requirements gathering and include not just functionality but also non-functional requirements such as for security, maintenance, upgrades and offboarding.

Actions:

- Provide comprehensive requirements for the full lifecycle.

² <http://nics.intranet.nigov.net/daera/articles/how-report-security-incident>

- Include requirements for tracking behaviour and performance of the AI to evaluate the impact on individuals and society and make changes if data or decisions become unrepresentative.
- Establish a robust testing and monitoring process to catch problems.
- Carry out appropriate “in life” testing by thoroughly assessing the ethical implications of the AI system with simulated scenarios to ensure fair and non-discriminatory outcomes.
- Perform risk assessments and regular risk reviews to ensure continued compliance.

Governance: Show evidence of non-functional requirements provided, and procedures being followed for setup, maintenance, upgrades, testing, and monitoring for solutions using AI.

3.6 Tooling Choice: Use the right tool for the job



Description: You should ensure you select the most appropriate technology to meet your needs. AI is good at many tasks but has a number of limitations and can be expensive to use, as well as having a potentially negative environmental impact. You should ensure you understand how your choice impacts the end-user.

Impact: This impacts decisions on tool selection and/or solution design

Actions:

- Ensure a good understanding of your use case and organisational goals, and undertake appropriate research, taking appropriate advice and ensuring relevant approvals are sort.
- Undertake feasibility studies and/or pilots to refine business case estimates, identify potential pitfalls, migration efforts, responsible use, decision making, in order to build confidence and trust, and manage change.
- As part of an options assessment, consider the use of the most efficient models and techniques (which may include non-AI methods). Many problems solvable by Gen AI, can also be addressed by more cost and energy efficient alternatives. It should be noted that AI causes significant greenhouse gas emissions, which needs to be considering when deciding whether to use an AI solution.
- Follow the guidance in Section 6.

Governance: Show evidence of options analysis including costs, environmental impact, benefits, and risks, to the appropriate authority such as the Enterprise Design Authority (EDA)

3.7 Collaboration: Collaborate with a variety of others from the start



Description: Across the Department and Government there are many teams and individuals involved in AI, as well as external expertise from commercial colleagues. Leverage these communities for better outcomes and ensure strong collaboration between the solution provider and the requester.

Impact: This impacts decisions on approach, requirements, procurement, design, training, and testing.

Actions:

- Develop a clear engagement plan from the start to align expectations on responsible and ethical use of AI for both in-house and third-party solutions.
- Ensure the end-user is clear on their responsibilities and the impact of their actions.
- Consult widely with industry and technology experts.
- Learn lessons from existing use cases.

Governance: Show evidence of compliance with DSD 3rd Party Onboarding Standard Operating Procedure (AE1/23/1799380) for any 3rd party AI procured.

3.8 Assurance: Use principles alongside other organisation's policies and assurances



Description: This policy is not standalone and must be considered alongside other organisational policies, standards and guidance for ICT, security, data protection, and procurement. Appropriate assurances must be established.

Impact: This impacts decisions on overall approach and governance.

Actions:

- Follow existing guidance and governance procedures and be aware of key risks – see section 7.2.
- Ensure existing guidance and standards are kept up to date with AI needs, if appropriate.
- Ensure compliance with data processing/storage/transmission legislation (including Model provenance, vendor accountability, and data residency). See data governance guidance:

<http://nics.intranet.nigov.net/daera/campaigns/data-governance>

Governance: Show evidence of compliance with existing DAERA assurance processes including Technical Design (TDD) review, Risk Assessment, and Data Protection Impact Assessment for all AI development and integration projects. (This includes obtaining appropriate approvals for new or novel AI uses from the AI Design Authority (AIDA), a subgroup of the Enterprise Design Authority (EDA) and securing Change Approval Board (CAB) approval before releasing any AI development and integration projects to production. AI related incident reports are monitored/reviewed by Security Assurance Group (SAG)).

4 AI Product Usage Policy

All AI products are currently blocked from use on the NICS network, except where Departmental approvals are specifically given in accordance with the security guidance and processes, [NICS Use of Generative Artificial Intelligence \(AI\)](#). With the publication of this DAERA AI policy, opportunities will develop for the managed use of AI products.

In the meantime, where staff have use of AI products (especially generative AI) on their own devices (i.e. personal devices not managed by NICS) that they wish to use for work purposes, caution needs to be exercised to ensure the risks identified in this policy are managed. Official information must never be shared with unapproved AI functionality and the sources used by the product must be

checked for validity. Examples of these types of free AI products include chat (ChatGPT), copywriting (Copy.AI, Notion.ai), image creation (DALL-E, Canva), and many more, including tools that have AI functionality embedded in them. Please note DAERA security policy dictates that users **must not** use home PCs or personal devices such as laptops and smart phones to access, save or store DAERA information.

5 Exceptions Policy

Any exceptions to this policy must be formally requested from the Information Assurance Steerco along with justification for the exception and analysis of potential risks. Exceptions must be approved, and mitigating controls agreed and implemented. The risks associated with approved exceptions will be registered on the Governance, Risk, and Compliance (GRC) tool and monitored for as long as the exception remains. The risk owner will be the person, minimum Grade 7, responsible for raising the exception in the first instance. Regular review of all approved exceptions must be carried out by the risk owner via the GRC tool as part of the exception. Regular reviews ensure that the exception remains valid, and risks are managed effectively. No exception should be in place for more than 12 months.

Any individual who is responsible for creating unapproved exceptions to this policy through either actions or lack of actions on their part will be in breach of this policy and may be subject to disciplinary and/or legal action.

6 Solution and Tooling Governance

The following guidance must be followed when adopting any new AI solution or technology within the Department. There are three possible routes to this, subject to the guidance below:

1. AI technology developed in-house as part of a product by the Digital Services Division on request, subject to business case and relevant approvals through the Repeatable Design process including TDD/TDA/EDA as applicable.
2. Re-use a pre-existing AI technology on the Digital Services Division (DSD) technology catalogue
3. Procurement from a third-party supplier following all relevant third-party onboarding governance.

6.1 Inhouse Development

6.1.1 Business Requester

When engaging Digital Services Division (DSD) for solution development including AI follow this guidance.

Demand for DSD services can be generated in several ways including, but not limited to:

1. Business Group Digital Board (preferred);
2. Contact from DSD representative; or
3. Contact from a Business representative.

Portfolio Management Branch will record all new Demand at this stage. Requests from business areas for a product that does not fall within the scope of an existing business case will then follow the pre-engagement process and be added to the funnel as a “Pre-engagement” product. Further information is provided at AE1/24/1154121.

Further information and support on potential AI applications can also be found on the [Data & Analytics Services Catalogue intranet page](#).

6.1.2 ICT Staff

When developing any solution, including using AI functions, follow the guidance, standards and procedures including development of Technical Design Document (TDD) and approval before starting development including proof of concepts (POCs).

- Information on the [Repeatable Design Process](#) can be found [here](#)
- Also see Standards section of this policy – section 7.1.2
- See: <https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development/guidelines/secure-development>

6.2 AI Products on Service Catalogue

Products on the Service Catalogue may not necessarily comply with the principles of this policy for the specific use case required by the business and should not be seen as a shortcut to compliance. This is especially the case with upgrades.

After ensuring financial approval from your budget holder, make a request via DAERA Service Requests Service.Request@daera-ni.gov.uk. You should consider if the service is the appropriate selection – see tool choice principle in section 3.6

6.3 Third Party Solutions and Procurement

AI is increasingly being used in more products and services, so even if you are not purposely going out for an AI product it may come as part of the solution.

A Controller or Joint-Controllers will determine the purpose and means of processing of personal data and as such will have overall responsibility of ensuring that processing using artificial intelligence is lawful and that the appropriate technical and organisational measures are in place. The Controller(s) will ensure that data protection by design and by default (UKGDPR Article 25) is adhered to and that mandatory data protection impact assessments (UKGDPR Article 35) are completed.

A third-party Processor must provide sufficient guarantees that technical and organisational measures are in place to ensure that any processing using AI on behalf of the Controller(s) is done in accordance with the UK General Data Protection Regulation.

The Gov.uk guidelines for AI Procurement [\(1\)](#) enables Government Departments to engage effectively with innovative suppliers or to develop the right AI-specific criteria and terms and conditions that allow effective and ethical deployment of AI technologies.

Some key points are:

Prepare

- Get expert advice and conduct data discovery
- Ensure data protection by design and by default (UKGDPR Article 25) is adhered to which will integrate the necessary safeguards into the processing of personal data in order to meet the requirements of the UKGDPR and protect the rights of data subjects
- Conduct technical and organisational risk assessment with DSD and engage with DSD Portfolio and Procurement teams for advice
- Consider preliminary market engagement

- Obtain confirmation of data residency, storage locations, and training-data handling in writing before approval
- Consider risks identified in the policy when writing requirements

Implementation & Management

- Testing the AI model on an ongoing basis is necessary to maintain its accuracy
- Operational or service staff must have enough knowledge of, or training on, the AI system to understand how to use it and act on its outputs
- Consider the end-of-life processes for your AI system, what the data should look like, and an Exit Strategy from the 3rd party

Please refer to the DSD 3rd Party Onboarding Standard Operating Procedure
<http://nics.intranet.nigov.net/daera/campaigns/digital-procurement>

7 Appendices

These appendices contain additional information to assist in your own area of work such as in identifying and managing project risk or developing a business case that includes AI. These appendices include:

- [Governance Context](#) provides references and information on legislation, regulation, and standards and how they apply to DAERA
- [Risk Assessments](#) for five key areas but are not intended to be exhaustive. You will need to consider additional risks relevant to your specific situation
- [Benefits Analysis](#) - not exhaustive and not all relevant to every scenario
- [Glossary of Terms](#) relevant to this document
- [Reference table](#) for documents referred to in this document

7.1 Governance Context

This section provides context for where this policy fits into broader guidance and provides further reading should you need it. The governance context section is divided into two areas:

1. Legislation & Regulation
2. Standards

7.1.1 Legislation & Regulation

This is likely to be an evolving area but for now this policy will summarise the AI position in relation to three areas of regulation and legislation.

Name: UK Gov Generative AI Framework	Reference: https://www.gov.uk/government/publications/generative-ai-framework-for-hmg/generative-ai-framework-for-hmg.html
Description: Gov.uk expanded framework, providing practical considerations for anyone planning or developing a generative AI solution. It is necessarily incomplete and dynamic. It is incomplete because the field of generative AI is developing rapidly and best practice in many areas has not yet emerged. It is dynamic because we will update it frequently as we learn more from the experience of using generative AI across government, industry, and society.	

Name: UK GDPR	Reference: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/
Description: The Information Commissioner's Office provide a comprehensive guide to data protection and privacy legislation. Article 23 of the UKGDPR deals with automated individual decision-making, including profiling. Data subjects have the right not to be subject to a decision based solely on automated processing, including profiling which produces legal effects concerning him or her or significantly affects him or her. However above would not apply if: • it was part of a contract between individual and DAERA	

- DAERA must do something by law and there are suitable safeguards in place
- Individuals explicit consent given

Provision of appropriate safeguards in any process is essential.

Decisions based on special category data (e.g. Biometric data, trade union membership, health) should be avoided.

Processing of personal data needs to be covered by at least one of the lawful grounds of processing set out in Article 6 in UK GDPR

Name:	Reference:
EU AI Act	https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
Description:	
The EU Act categorises AI systems into four risk categories, each with different compliance requirements. While as a non-EU government department, DAERA may not be subject to the requirements of the Act in all circumstances, it is nevertheless important to be mindful of how the Act may apply in certain circumstances, particularly when interacting with EU bodies, organisations, and citizens.	

7.1.2 Standards

Wherever relevant, DAERA looks to international guidance and best practice standards to influence and inform how the delivers the Department's services to a level of consistency and quality.

Name:	Reference:
ISO standards	https://cmwebapp/ContentManager/?uri=119323539&t=record
Description:	
As a paying agency DAERA is certified to ISO27001:2022 standard and although this does not make specific/explicit reference to AI controls from a security perspective, it is implicitly addressed as a significant contextual influence that DAERA must address in terms of its deployment as an 'in scope' technology. The AI standards will align with ISO42001 for Artificial Intelligence Management System (AIMS) within organisations.	

Name:	Reference:
NCSC Guidance	https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development/guidelines/secure-development
Description:	
NCSC has a succinct and easy to read guide called "AI and cyber security: what you need to know" that provides clear advice on AI security. The key message is to abide by 'Security by Design' principle for AI as with other systems and technology.	

Name:	Reference:
Coding Standards	https://pr-dae-tfsapp1/tfs/DARD%20ISB/DSD/wiki/wikis/DSD.wiki/2148/DSD-Security-(Coding-Design)-Standards-Guidance

Description:

Use or creation of AI services and systems must be consistent with existing DSD standards.

Name:

Information Commissioner's Office
AI Guidance

Reference:

<https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/>

Description:

The Information Commissioner's Office has developed AI guidance and its relevance to data protection.

Name:

NICS Use of Generative Artificial
Intelligence (AI)

Reference:

<http://nics.intranet.nigov.net/nics/documents/nics-use-generative-artificial-intelligence-ai>

Description:

A policy is to ensure a consistent approach on the use of Generative AI across the NICS, providing strategic context from the National AI strategy, with high level view of implications and benefits as well as a list of approval required to utilise a Gen-AI technology

7.2 Risk Assessment

Undertaking a risk assessment and being aware of potential issues in advance enables us to take proactive decisions to reduce potential risks and ensure regulatory compliance. We have an obligation to safeguard citizens, staff, and other stakeholders; and to ensure data is secure. This also helps to maintain transparency and accountability while taking advantage of new opportunities. It may also allay staff concerns of losing jobs.

Below are five general risk statements that have been identified for AI use in DAERA; individual risk assessments per product or use case should be carried out before implementing any AI solution to address these statements, as a minimum. This high-level risk assessment covers: Ethics, Data Protection, Effectiveness, Opportunity, and Data Poisoning. Other relevant risks may and should also be considered.

7.2.1 Unethical Risk

AI can be unethical when it is deployed in ways that infringe upon privacy, perpetuate biases, or harm individuals or society without proper consideration of ethical implications, potentially leading to unintended consequences and societal harm.

Potential Impact (without mitigation): Very high reputational damage and loss of public confidence

Causes: Very high likelihood

- Inherent bias present in training data e.g. race, gender
- Lack of transparency in AI model predictions and actions
- Used in unethical ways or for unethical reasons
- Lack of Governance / Review / Approval gates
- Mis-prompted question wording used as input to Gen AI
- Accidental misuse and understanding of limitations

Potential Mitigation Actions:

- Identify Accountability – identify who is responsible for the decisions and actions taken by the AI system (including design, development, deployment, and outcomes).
- Ensure appropriate user training is undertaken – AI actors need awareness and to acquire appropriate skills whether designing, procuring, or using an AI solution.
- Documentation - provide/obtain documentation on AI product accuracy, limitations, and how to override results.
- Ensure monitoring and auditing – implement systems and processes to track behaviour and performance of AI, to evaluate the impact on individuals and society and make changes if data or decisions become unrepresentative.
- Provide transparency – make the working of AI understandable to stakeholders, including factors used to make decisions and where there is a human in the loop.
- Use high-quality training data – the training data should be accurate and representative as possible, and that biases and distortions are understood and addressed both on a technical and business level.
- Carry out appropriate testing – thoroughly assess the ethical implications of the AI system with simulated scenarios to ensure outcomes are not unfair or discriminatory (e.g. factors such as race, gender, etc)

7.2.2 Data Breach Risk

AI systems are particularly at risk of causing a data breach compared to other IT systems due to their reliance on vast amounts of data, which may include sensitive or personal information, coupled with the complexity of algorithms and models, increasing the likelihood of inadvertent exposure or exploitation of vulnerabilities, necessitating heightened attention to security protocols and data protection measures.

Potential Impact (without mitigation): Very High reputational damage and financial penalty

Causes: Very high likelihood

- Individuals unintentionally sharing inappropriate data to AI explicitly or indirectly
- Datasets used for AI training/learning are shared outside the organisation
- AI solution does not apply suitable access controls
- Adherence to industry standards/best practice on the use of AI is not maintained
- Threat actors use AI technology against NICS e.g. data poisoning and attacks on the model itself
- Insufficient guardrails around the use of datasets
- Nonadherence to DAERA policy / standards / processes
- Hidden permission and use of DAERA data by 3rd parties for training their AI models

Potential Mitigation Actions:

- Develop appropriate guidance - Develop policies, standards and acceptable use guidance and training on AI
- Keep current - Keep staff up-to-date with the latest guidance and findings
- Carry out risk assessment and regular reviews - Undertake risk assessments per use case or new product
- Ensure security by design and access control are built-in
- Make use of AI powered tooling for defence
- Ensure awareness and compliance needs for Data processing/storage/transmission locations

7.2.3 Ineffective Use Risk

The embryonic nature of AI technology can lead to delays or ineffectual results due to challenges such as data quality issues, limited understanding of AI capabilities, and the need for iterative development processes, which may prolong implementation timelines and hinder initial effectiveness until further advancements and expertise are gained.

Potential Impact (without mitigation): High financial cost and project delays

Causes: Very high likelihood

- Underestimate of effort
- Lack of forecasting or monitoring of in-life run costs
- Overestimate of usefulness
- Lack of Training/Competency
- Vendor Lock-In and Dependency
- Technical Constraints
- Hallucinations (a response generated by AI which contains false or misleading information presented as fact)

Potential Mitigation Actions:

- Ensure appropriate user training is undertaken, and documentation is provided so that users understand the limitations of the system, and how to use and interpret the results of the system correctly
- Ensure that users can override the output, action or recommendation made by the system and can make an informed judgement on when it is best to do so
- Address project optimism bias for business case cost estimates to improve accuracy through a feasibility study
- Consult widely with industry and technology experts
- Study lessons learnt from existing use cases for new projects
- Establish and maintain an AI Community of Practice and/or Centre of Excellence

7.2.4 Missed Opportunity Risk

The opportunity risk of not leveraging AI includes missing out on potential efficiency gains, competitive advantages, and innovation opportunities that AI can offer, as well as falling behind in industries where AI adoption is becoming the norm or expected by citizens, ultimately limiting growth potential and risking obsolescence in an increasingly AI-driven world.

Potential Impact (without mitigation): High financial e.g. duplication of effort; unnecessary manual effort

Causes: High likelihood

- Lack of trust in AI/too risk averse
- AI limitations are not fully understood
- Lack of coordination/consistency in adoption across UK government and Departments
- Failure to identify/measure benefits

Potential Mitigation Actions:

- Carry out research into current processes and practises to identify use cases where efficiencies could be made through the use of AI
- Research available solutions against organisational goals and objectives
- Implement pilot projects to test feasibility and impact of new technologies
- Keep up to date on advancement and progress of new technology and solutions available for use
- Take a feasibility-based approach to assess potential use cases or new ideas to identify potential pitfalls, migration efforts, responsible use, decision making, to build confidence and trust, and manage change.

7.2.5 Data Poisoning Risk

Data poisoning is a type of cyberattack in which an adversary intentionally compromises a training dataset used by an AI or machine learning (ML) model to influence or manipulate the operation of that model.

Data poisoning can be done in several ways:

- Intentionally injecting false or misleading information within the training dataset
- Modifying the existing dataset
- Deleting a portion of the dataset

By manipulating the dataset during the training phase, the adversary can introduce biases, create erroneous outputs, introduce vulnerabilities, or otherwise influence the decision-making or predictive capabilities of the model. Data poisoning falls into a category of cyberattacks known as adversarial AI. Adversarial AI or adversarial ML is any activity that seeks to inhibit the performance of AI/ML systems by manipulating or misleading them.

Potential Impact (without mitigation): Very high operational impact, reputational damage, and loss of public confidence

Causes: Very high likelihood

- **Direct Attacks** - also known as targeted attacks, they occur when threat actors manipulate the ML model to behave in a specific way for a particular targeted input, while leaving the model's overall performance unaffected.
- **Indirect Attacks** – these are nontargeted attacks that aim to affect the overall performance of the ML model, not just a specific function or feature.
- **Mislabeling Attack** - a threat actor deliberately mislabels portions of the AI model's training data set, leading the model to learn incorrect patterns and thus give inaccurate results after deployment.
Data injection - threat actors inject malicious data samples into ML training data sets to make the AI system behave according to the attacker's objectives.
- **Data manipulation** - involves altering data within an ML model's training set to cause the model to misclassify data or behave in a predefined malicious manner in response to specific inputs. Techniques for manipulating training data include adding incorrect data, removing correct data, or injecting adversarial samples. The end goal of a data manipulation attack is to exploit ML security vulnerabilities, resulting in biased or harmful outputs.
- **Backdoors** - Threat actors can also plant a hidden vulnerability -- known as a backdoor -- in the training data or the ML algorithm itself. The backdoor is then triggered automatically when certain conditions are met. Typically, for AI model backdoors, this means that the model produces malicious results aligned with the attacker's intentions when the attacker feeds it specific input. This type of attack is a severe risk in AI and ML systems, as an affected model will still appear to behave normally after deployment and might not show signs of being compromised as the attacker waits for their moment.
- **ML supply chain attack** - ML models often rely on third-party data sources and tooling. These external components can introduce security vulnerabilities, such as backdoors, into the AI system. These attacks are not limited to ML training models; they can occur at any stage of the ML system development lifecycle.

- **Insider Attacks** - perpetrated by individuals within an organization -- such as employees or contractors -- who misuse their authorized access privileges to the ML model's training data, algorithms, and physical infrastructure. These attackers have the ability to directly manipulate the model's data and architecture in different ways to degrade its performance or bias its results. Insider attacks are particularly dangerous and difficult to defend against because internal actors can often bypass external security controls that would stop an outside hacker.

Potential Mitigation Actions:

- **Training data validation.** Prior to starting model training, all data should be validated to detect and filter out any suspicious or potentially malicious data points. This helps safeguard against the risk of threat actors inserting and later exploiting such data.
- **Continuous monitoring and auditing.** Like all information systems, AI systems need strict access controls to prevent unauthorised users from accessing them. Apply the principle of least privilege and set logical and physical access controls to mitigate risks associated with unauthorised access. Continuous monitoring and auditing should also focus on the model's performance, outputs, and behaviour to detect potential signs of data poisoning.
- **Adversarial sample training.** Introducing adversarial samples during the model's training phase is a vital proactive security defence measure to stop many data poisoning attacks. This enables the ML model to correctly classify and flag such inputs as inappropriate.
- **Diversity in data sources.** Using multiple data sources enables an organisation to diversify its ML model training data sets, significantly reducing the efficiency of many data poisoning attacks.
- **Data and access tracking.** Keeping a record of all training data sources is essential to stop many poisoning attacks. Also, consider keeping a record of all users and systems that gain access to the model -- and what they do each time -- to help identify potential threat actors.

7.3 Benefits Analysis

AI has the potential to help:

- **Engage** our citizens/stakeholders
- **Enhance** our services
- **Empower** our staff
- **Optimise** our operations

Advancements in AI technology is revolutionising all sectors in numerous ways, leading to enhanced **efficiency** and productivity. By enabling time saving and improved consistency in certain tasks, AI has the potential to streamline operations, while **improving satisfaction** levels for both employees and consumers.

AI in software development can optimise output while reducing vulnerabilities in code and errors in applications, while also **improving accuracy** in data.

Additionally, AI can be used to facilitate **improvements in the environment** such as a reduction in carbon emissions, lowers costs, and **enhances regulatory compliance**, thereby fostering a more sustainable agricultural sector and a healthier natural environment, as long as the appropriate tool is chosen.

Predictive modelling can be used to enable **greater resilience** to external shocks, such as climate changes, and bolstering yield in industries like farming and forestry.

Overall, AI's multifaceted benefits could span from reducing disease and fraud to enhancing food provenance and emissions tracking, positioning NI for a **more sustainable and prosperous future** in keeping with DAERA's strategic vision.

Through the use of appropriate data and choice of AI tools the following benefit might be possible.

7.3.1 Efficiency

AI can be deployed to search, research, number crunch, and carry large volumes of repetitive tasks resulting in:

- Reduction of staff and customer time – able to efficiently answer questions and provide the information required without significant searching
- Increased throughput – AI doing the repetitive and dull tasks, so staff can instead focus on value-add tasks
- Improved staff capability – through targeted training or in finding the exact information required
- Cost Reduction – as a result of efficiency saving through choosing the right solution for the problem (sometimes this may not be an AI solution)

7.3.2 Quality

Here are some examples of how AI can improve quality:

- Improved customer and staff satisfaction – greater efficiency and reduced waiting times or inefficiencies e.g. meeting or appointment efficiency
- Improved quality of documentation – GenAI assistance

- Reduction of code vulnerabilities, fewer defects/errors, and less platform performance issues – automated detection and pattern recognition
- Improved accuracy - the volume of records and data assessed can improve accuracy that surpasses human ability
- Reduced fraud including Food fraud reduction – pattern recognition
- Disease reduction – predictive analytics and modelling

7.3.3 Environmental*

The improvements in efficiency and quality can be leveraged to deliver environmental benefits such as:

- Sustainable Agri Sector – track and manage carbon emissions and promote carbon-friendly alternatives
- Healthier Natural Environment – air and water quality monitoring, prediction, and planning
- Reduced Carbon Emissions – optimised transportation routes
- Improved yield in industry - demand forecasting to reduce surplus/waste
- Greater resistance to external shocks - predictive modelling of the climate
- Improved forest planting – geospatial analysis for site selection, models for predicting suitable tree species for the soil and location, growth prediction and tracking, fire and pest prediction, disaster response planning
- Food provenance - blockchain integration to ensure tamperproof records, real-time tracking with IoT, pattern recognition to detect fraud, image recognition, and reporting
- Reduction in use of chemicals - precision insights for watering, fertiliser, and pesticides

*The use of AI technology can itself cause significant carbon emissions due to the volumes of data it can consume and that it can run 24-hours a day.

7.3.4 Compliance

A large part of government responsibility pertains to compliance with regulation and the complexity that can surround it, AI can help to identify areas of weakness, or predict areas to focus on, therefore resulting in:

- Increased ability to meet regulatory requirements
- Identification of weakness or inconsistency between regulations
- Ability to summarise for different consumers how best to comply with regulation

7.3.5 Security

As Artificial Intelligence (AI) systems become more advanced, public institutions face growing risks from increasingly sophisticated cyber threats. (see glossary terms for Cyber Threat) AI is increasing the speed and efficiency of pre-existing cyber-attack techniques, while accelerating the development of new tactics and making them increasingly difficult to detect, monitor and remediate.

AI dramatically speeds up the intensity and scale of hostile activity and, when used by malicious actors, allows attackers to quickly adapt to defeat defences and find new vulnerabilities to exploit. AI also allows attackers to adapt their software tools to make them harder to detect in order to carry out more stealthy attacks.

Traditional methods of monitoring, threat hunting, incident response, and other duties are often manual and time-intensive, which can delay remediation activities, increase exposure, and heighten vulnerability to cyber adversaries using AI.

Fighting fire with fire - AI can bring substantial benefits to cyber defensive operations by automating these traditional methods, can transform cyber workflows into streamlined, autonomous, continuous processes that speed remediation and maximise protection.

Immediate and long-term benefits of integrating AI into an organisation's cybersecurity ecosystem include that it:

- Improves protection and remediation due to AI's ability to detect nuanced attacks, heighten security, and enhance incident response
- Increases time savings as AI expedites the detection and response cycle time, rapidly quantifying risks and accelerating analyst decision making with data-driven mitigation measures
- Fortifies protection of reputation and trust in an organization's security systems and protocols, especially important to a public service-oriented department like DAERA
- Improves workforce satisfaction, as cybersecurity professionals can focus on higher level tasks as opposed to time-consuming, manual actions

7.4 Glossary of Terms

This is a list of terms used in this document, for other AI related terms see UK Parliament "Artificial intelligence (AI) glossary" [\(2\)](#)

Term	Description
Artificial Intelligence (AI)	A term coined by Standard Professor John McCarthy in 1955, was defined as <i>"the science and engineering of making intelligent machines."</i> Previous research has included programming machines to behave in a clever way, like playing chess, but, today, we emphasize machines that can learn, at least somewhat like humans do.
AI Actor	Someone who play an active role in the AI system lifecycle, including organisations and individuals that design, create, maintain, deploy, or operate AI.
AI Bias	Refers to a machine learnt bias perpetuated by human biases and learnt from the training material
AI Model	In the context of AI or statistics, a model refers to a mathematical or computational representation that captures patterns, relationships, or underlying structures in data. These can range from relatively simple statistical models, e.g. average fuel consumption per mile based on factors such as speed and engine size, to deep learning and large language models, which may include billions of parameters fine-tuned to interpret or generate complex input/output such as images, video, or natural language.
AI User	Someone who plays an active role in using an AI system, or tools or services with AI embedded in them. This includes organisations and individuals that use or operate AI.
Cyber Threat	AI-generated misinformation and cyber-attacks represent #2 and #5 of the top 10 global risks in 2024, per the World Economic Global Risk Report.

	Data poisoning will become a top concern as AI supply chain attacks gain momentum. Data poisoning is a type of cyberattack in which an adversary intentionally compromises a training dataset used by an AI or machine learning (ML) model to influence or manipulate the operation of that model. AI companies as well as their training models and downstream suppliers will be increasingly targeted by malicious actors. The OWASP Top 10 for LLM Applications highlights training data poisoning and supply chain attacks as significant risks, running the risk of compromising the security, reliability, and performance of AI applications. Simultaneously, vulnerabilities in AI application supply chains— including technology partners, third-party data sets, and AI tool plugins or APIs—are ripe for exploitation.
Deep Learning	A subset of machine learning that uses artificial neural networks to recognise patterns in data and provide a suitable output, for example, a prediction. Deep learning is suitable for complex learning tasks and has improved AI capabilities in tasks such as voice and image recognition, object detection and autonomous driving.
General Artificial Intelligence (AGI) & Super AI	Refer to hypothetical future AI techniques that approach or surpass human intelligence on a wide range of problems.
Generative AI (GenAI)	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.
Grounding	Provides specific context to more general datasets to improve the AI model for the specific user or organisation or purpose it is used for. It links abstract knowledge to real-world examples.
Hallucination	A phenomenon wherein a large language model (LLM)—often a generative AI chatbot or computer vision tool—perceives patterns or objects that are non-existent or imperceptible to human observers, creating outputs that are nonsensical or altogether inaccurate. Hallucinations can produce outputs that seem plausible, while at the same time being partially or wholly inaccurate and misleading. Care must therefore be taken to review and validate Gen AI outputs to guard against inappropriate interpretation or use of potentially inaccurate results.
Intelligence	The ability to learn and perform suitable techniques to solve problems and achieve goals, appropriate to the context in an uncertain, ever-varying world.
Large Language Models (LLM)	A type of foundation model that is trained on vast amounts of text to carry out natural language processing tasks. During training phases, large language models learn parameters from factors such as the model size and training datasets. Parameters are then used by large language models to infer new content.
Machine Learning	ML is a branch of AI that allows a system to learn and improve from examples without all its instructions being explicitly programmed. An ML system is trained to carry out a task by analysing large amounts of training data and building a model that it can use to process future data, extrapolating its knowledge to unfamiliar situations.

Mis-prompting	Also see prompting. Mis-prompting is the wilful use of inputs to the AI to force a negative result. Data poisoning can taint the training data.
Narrow Artificial Intelligence	Also known as Weak AI. All current AI techniques are examples of Narrow AI, capable of solving specific targeted tasks.
Predictive Statistics	Forecasting and time series analysis.
Prompting	Any form of text, question, information, or coding that communicates to AI what response you're looking for. Adjusting how you phrase your prompt; AI could provide varying responses. Used with Chatbot, LLM, and GenAI.
Training Data	The sets of data used to train an AI system. Training involves presenting a machine learning algorithm with example inputs and expected output from the training data (e.g. pictures of apples and oranges labelled correctly), from which the algorithm produces a model that is able to correctly predict the correct label or output for new input that the model has not previously seen. (e.g. correctly label a picture of an apple or orange not in the training set).

7.5 References

1. **Gov.uk.** Guidelines for AI procurement. *GOV.UK*. [Online] 8 June 2020. <https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement#ai-specific-considerations-within-the-procurement-process>.
2. **UK Parliament.** Artificial intelligence (AI) glossary. *UK Parliament*. [Online] 23 January 2024. <https://post.parliament.uk/artificial-intelligence-ai-glossary/>.
3. **Gov.uk.** A pro-innovation approach to AI regulation - Part 3: An innovative and iterative approach. *GOV.UK*. [Online] 3rd August 2023. <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#part-3-an-innovative-and-iterative-approach>.