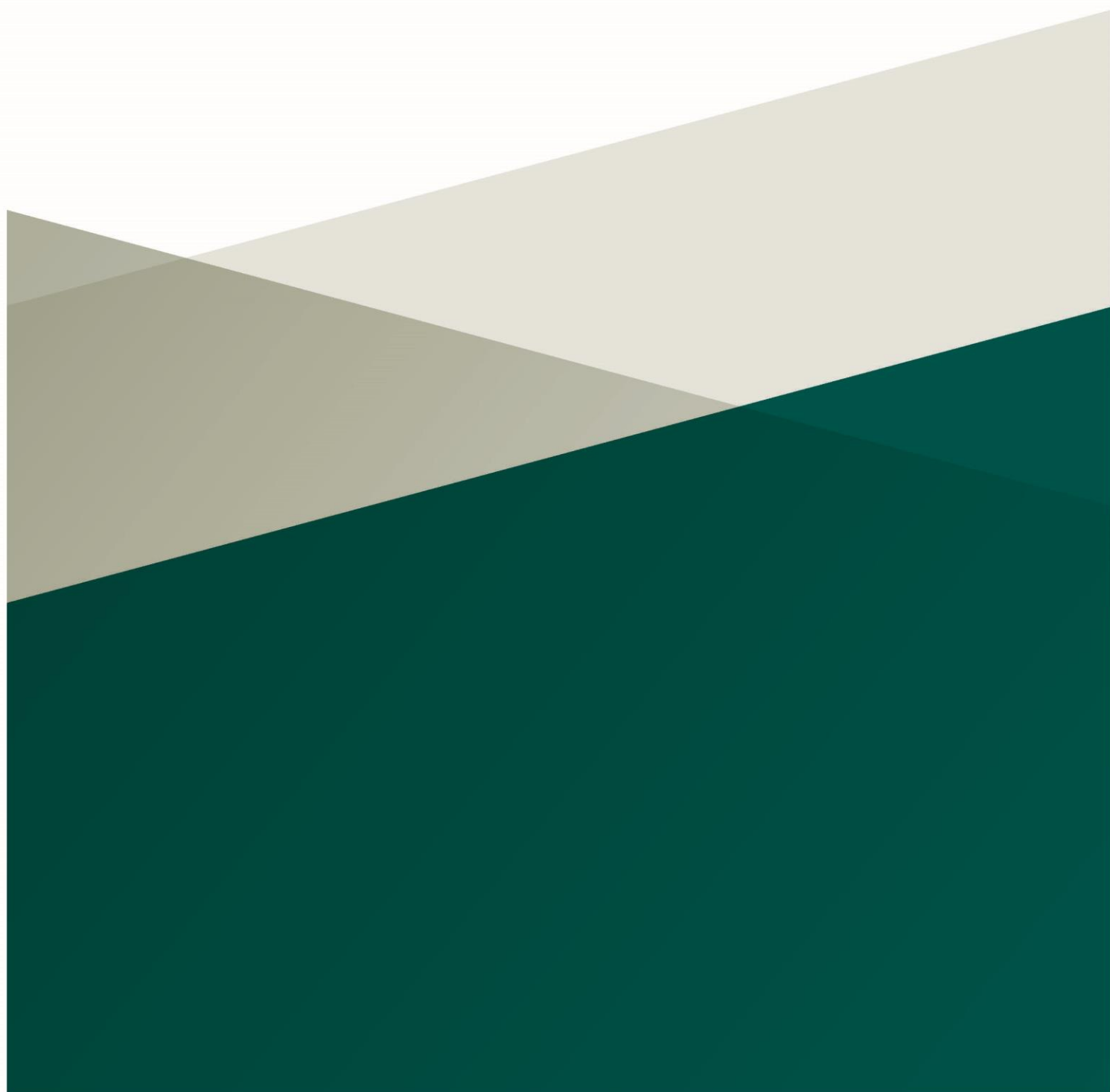




An Roinn Caiteachais Phoiblí
Sheachadadh PFN agus Athchóirithe
Department of Public Expenditure
NDP Delivery and Reform

Interim Guidelines for Use of AI in the Public Service

February 2024



Interim Guidelines for Use of AI in the Public Service

Interim Guidelines for Use of AI in the Public Service	2
1 Introduction	3
2 AI use cases	3
3 Government commitment to ethical AI	4
4 Key Considerations before deciding to adopt an AI tool	5
4.1 Risk Assessment	5
4.2 Is there advice or guidance Available?	6
5 Safeguards and Considerations	6
5.1 Human Oversight	6
5.2 Bias and unfairness	7
5.3 Data Protection and Intellectual Property Implications	7
5.4 Skills and Competencies	7
5.5 Cyber Security	7
6 Supports and Resources for Civil and Public Service Organisations	8

1 Introduction

Artificial Intelligence tools and technologies have significant potential to transform the delivery of public services, and to help in dealing with big societal challenges. It can be an excellent tool for the optimisation of scarce resources or smart scheduling. It can identify patterns and anomalies in large data which could give valuable insights to policymakers and improve the accuracy of forecasting.

AI tools can be used to enhance workplace safety, and can be deployed to perform hazardous and dangerous tasks. The recently developed general language models can also be a valuable tool in training and education, and in supporting clerical tasks across the public service.

However, AI systems also pose risks that can negatively impact individuals and society. Without proper controls, AI systems can produce or amplify inequitable or undesirable outcomes for individuals and communities.

By its nature, the work we do as public servants has impact on people's lives and wellbeing, and so it is essential that we fully understand the potential risks involved in using AI, and what safeguards are needed.

This document sets out some guidelines and issues for consideration for public sector organisations when considering the use of AI tools.

2 AI use cases

Examples of how AI can be used as a tool include:

- Efficiency: Optimising tasks like resource allocation, planning and research
- Diagnostic: AI's ability to identify trends and spot anomalies make it an excellent diagnostic tool
- Instruction: Providing personalised guidance or learning content
- Creation: Generating or enhancing content. This could have widespread applicability to the creation of first drafts of routine documents such as letters, press releases and briefs

3 Government commitment to ethical AI

The Government has made a commitment that AI tools used in the civil and public service must comply with seven key requirements. These are:

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

A summary of these requirements is attached in Annex 1. The requirements are included in Ethics Guidelines for Trustworthy Artificial Intelligence which were developed by the High-Level Expert Group on Artificial Intelligence established by the European Commission.

The High-Level Expert Group has also delivered an Assessment List for Trustworthy Artificial Intelligence (ALTAI), through which these seven key principles and requirements are translated into an accessible and dynamic checklist that guides developers and deployers of AI in implementing such principles in practice.

AI systems may be developed in house or procured from a third party vendor. In either case, the responsibility to ensure AI is trustworthy and ethical rests with the system user, not the developer. This responsibility cannot be delegated.

Ethics considerations should be built into the Requirement Specification of the system, and specific ethics test cases should be written and signed off by senior management.

Subject matter experts should be part of the design process for an AI system, and have an important role to play in ensuring that all service users that will be affected by the system are treated fairly.

User manuals and user guides must have the ethical considerations outlined clearly with a specified email or person nominated as the contact point for ethical issues.

4 Key Considerations before deciding to adopt an AI tool

4.1 RISK ASSESSMENT

When considering whether AI is the right tool for a particular policy objective, it is essential to think critically about the risk of negative impacts. Many uses of AI are low risk, for example AI enabled inventory management systems, or Chatbots. However, where the proposed use of AI is high risk, more detailed consideration is needed.

High Risk Use Cases:

The use of AI systems in contexts where the system output could impact:

- **fundamental human rights** (e.g. human dignity, privacy, non-discrimination, fair trial, safety, freedom of expression): **or**
- **well-being** (e.g. job quality, health, civic engagement, education, the environment, social interactions)

These scenarios are likely to be **high risk** and will require careful ex-ante consideration.

More specifically, the draft EU AI Act sets out the following contexts as high risk use cases for AI:

- Biometric identification
- AI Systems intended for use as safety components in the management and operation of critical [digital] infrastructure
- Education and vocational training (in particular for determining access, assessing student performance, assessing the appropriate level of education for the individual)
- Employment, workers management and access to self-employment
- Access to and enjoyment of essential private services and public services and benefits
- Risk assessment and pricing for life and health insurance
- Systems intended for use by law enforcement to assess risk of a person reoffending, for emotion recognition purposes, to evaluate the reliability of evidence, to profile natural persons
- Migration, asylum and border control management
- Administration of justice and democratic processes

Under the regulations, high-risk AI systems will be subject to strict obligations before they can be put on the market, including adequate risk assessment and mitigation; appropriate

human oversight measures to minimize risk; and logging activity to ensure traceability of results.

4.2 IS THERE ADVICE OR GUIDANCE AVAILABLE?

If your planned use of AI falls in any of the above high-risk categories, the risks of using AI may outweigh the benefits. However, it is important that you look for expert guidance, so that you can make a fully informed decision.

Government has established an AI Advisory Council, which is a resource for policymakers who need expert assistance to support them in thinking critically about context and potential or unexpected negative and positive impacts of proposed AI adoption.

To seek advice from the AI Advisory Council, contact the Artificial Intelligence and Future Manufacturing Unit in the Department of Enterprise, Trade and Employment, who provide a Secretariat to the Council at digitaleconomy@enterprise.gov.ie.

5 Safeguards and Considerations

5.1 HUMAN OVERSIGHT

AI tools can assist human capabilities, but they should never replace them. All AI tools used in the public service must be part of a process that has human oversight built into the process.

For example:

- **The Importance of Human Judgement.** AI can generate evidence to support better decisions. However, it cannot substitute for the use of judgement. Where AI is used as a tool in a decision-making process in a high risk context, **a human must make the final decision.**
- **The importance of Human Oversight and Review.** Issues can also arise such as General AI models drawing from unreliable or out of date sources. A human must always critically review material generated by AI systems for sense and accuracy.
- **The importance of well-designed instructions.** The Garbage In Garbage Out rule also applies to AI systems. An AI system will do what it is asked to do – so if it is given poorly thought out instructions, it will produce bad outcomes.

5.2 BIAS AND UNFAIRNESS

When using AI as a tool, it is important to be aware of the potential for AI-based systems to introduce or reinforce a risk of perpetuating inequity and historic bias. The bias can come from the data set, the system programmer, the user, or all three.

Conversely, AI can be used as a tool to reduce human biases, for example by excluding identifiers that can trigger unconscious bias.

A thorough consideration of bias is a prerequisite for the introduction of any AI system in the public service. Robust consultation with a diverse range of stakeholders is recommended, to surface possible biases.

Where the AI system is built in house, testing for accuracy and bias is an essential step. A robust data set that represents the diversity of potential end users in real world conditions and environments should be used for testing.

Where the AI system is procured from a 3rd party vendor, it is essential to evaluate for accuracy and bias before implementing the system.

5.3 DATA PROTECTION AND INTELLECTUAL PROPERTY IMPLICATIONS

Data used in an AI model must comply with GDPR requirements. Under GDPR, permission must be sought to use personally identifiable information. This includes facial images and voice.

Data should not be used in AI models in a way that breaches intellectual property rights.

Where the system is procured from a third party vendor, they must confirm that their data is GDPR compliant and does not breach the intellectual property rights of others.

5.4 SKILLS AND COMPETENCIES

You do not necessarily need programming skills to use AI tools, but there are AI skills and competencies that are necessary.

For example, team members will need to understand how the AI works, the associated risks and opportunities, and be able to determine whether outcomes are as expected or corrective action is needed.

5.5 CYBER SECURITY

Most of the current Generative AI models are available for use on public cloud-based systems. It is difficult to exercise control over the information input into these models, which

may then be seen by others, including cyber criminals. This poses a risk to the cyber security of civil and public service organisations.

It is recommended that relevant risk assessments and safe usage policies are in place for any Public Body considering the use of an AI tool. The National Cyber Security Centre has recently published two relevant guidance documents including a guidance document on the cyber security considerations when procuring ICT products and services¹, and specific guidance on the use of Generative AI.

6 Supports and Resources for Civil and Public Service Organisations

If you decide to proceed with adopting an AI tool, having considered all of the above you will need to ensure there are appropriate controls and safeguards in place to ensure adherence with the seven requirements to which the government has committed.

Further supports and resources including learning and development interventions, funding opportunities and networks of professionals are available via the Public Service Transformation Delivery Unit in the Department of Public Expenditure, NDP Delivery and Reform, and can be accessed at gov.ie/transformation.

¹ National Cyber Security Centre guidance document on [cyber security consideration when procuring ICT products and services](#).

Annex I

Summary of Ethics Guidelines for Trustworthy Artificial Intelligence developed by the High-Level Expert Group on Artificial Intelligence established by the European Commission

The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. A specific assessment list aims to help verify the application of each of the key requirements:

- Human agency and oversight: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches
- Technical Robustness and safety: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.
- Privacy and data governance: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.
- Transparency: the data, system and AI business models should be transparent. Traceability mechanisms can help achieving this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.
- Diversity, non-discrimination and fairness: Unfair bias must be avoided, as it could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.
- Societal and environmental well-being: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment,

including other living beings, and their social and societal impact should be carefully considered.

- **Accountability:** Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate and accessible redress should be ensured.

