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

Guidelines for the Responsible Use of Artificial Intelligence in the Public Service

Department of Public Expenditure,
NDP Delivery and Reform

Revenue
20%
34%
83%
Analysing Dashboard
Total
1,000 %
8,000
6,000
4,000
2,000
100
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sept
Oct
Nov
Dec
New total
Old total
Ai
$ 3,553K | $2,688K
50%
55%
73%
80%

Table of Contents

Glossary of Terms

A

Accountability: One of the seven principles for trustworthy AI. Ensures that responsibility for AI decisions remains with identifiable individuals or entities, and that mechanisms are in place for redress in case of harm or failure.

AI Act (EU AI Act, 2024): Legislation enacted by the European Union to ensure that AI systems are used in a safe, transparent manner that is aligned with fundamental human rights. It categorises AI systems by risk levels and sets compliance principles accordingly.

AI Lifecycle: The series of stages an AI system goes through, from initial design and development to deployment, monitoring, and eventual decommissioning.

AI System: A machine-based system capable of operating autonomously and producing outputs like predictions, recommendations, or decisions based on input data.

B

Better Public Services: 'Better Public Services' is the transformation strategy for the public service aimed at delivering inclusive, high-quality and integrated public services.

C

Compliance: Adhering to relevant legal and ethical guidelines, such as the EU AI Act and GDPR, when designing and using AI systems.

Content Generation: AI's ability to autonomously create text, images, audio, or video.

D

Data Sharing and Governance Act (2019): Legislation that establishes a legal framework for the secure and efficient sharing of data between Public Service Bodies.

Data Collection & Processing: The process of gathering, cleaning, and preparing data for use in AI models in a manner that complies with data protection regulations.

Decision Framework: A structured approach to evaluate whether AI is an appropriate solution for a given problem or potential improvement.

Deployment: A stage in the AI lifecycle where a system is integrated into real-world environments, making it operational for users.

Design, Data & Models: The initial phase of the AI lifecycle, which involves the Planning & Design, Data Collection & Processing, and Model Building elements of developing an AI system.

Design Principles: A set of guiding principles that ensure public services, including digital and AI-driven solutions, are user-centred and inclusive.

Diversity, Non-Discrimination, and Fairness: A key principle of responsible AI that ensures equitable outcomes and prevents bias in AI systems.

E

Ethics Guidelines for Trustworthy AI: A document by the European Commission's High-Level Expert Group (HLEG) outlining principles and principles for responsible AI.

G

GDPR (General Data Protection Regulation): European regulation governing data protection and privacy, critical for AI systems handling personal data.

Generative AI (GenAI): A type of AI designed to generate content, such as text, images, or code, based on input data. Examples include AI chatbots, image generation tools, and content automation systems.

H

HLEG (High-Level Expert Group): The European Commission's advisory body on AI, responsible for the Ethics Guidelines for Trustworthy AI.

Human Agency and Oversight: A principle ensuring human control and intervention in AI systems, prioritising ethical decision-making.

M

Model Building: The process of creating and training an AI model using collected data, refining it to achieve desired outcomes while addressing bias and technical robustness.

O

OECD AI Principles: Guidelines developed by the Organisation for Economic Co-operation and Development (OECD) to guide organisations in their efforts to develop AI in a responsible manner.

Operation & Monitoring: A stage in the AI lifecycle that involves the ongoing management of an AI system to ensure it continues to function effectively, and the continuous monitoring to detect performance issues and security threats.

P

Planning & Design: A stage in the AI Lifecycle where objectives, risks, constraints, and ethical considerations are defined. It involves determining whether AI is the right solution and aligning the project with responsible AI principles.

Privacy and Data Governance: A principle focusing on protecting personal data, ensuring secure data handling and processing in compliance with regulations like GDPR.

Productivity: The ability of AI to improve efficiency and effectiveness in public services by automating tasks and analysing large datasets.

R

Regulatory Compliance: The principle that AI systems align with regulations such as the EU AI Act and GDPR.

Responsible AI Framework: A structured approach developed by the Irish Public Service to ensure ethical, safe, and effective use of AI technologies.

Responsiveness: The capacity of AI systems to adapt to user needs, providing personalised, human-centred public services.

Retire/Decommission: The final stage in the AI lifecycle, where the system is discontinued or replaced.

S

Societal and Environmental Well-Being: A principle that ensures AI systems contribute positively to society and minimise environmental harm.

T

Technical Robustness and Safety: A principle ensuring AI systems are resilient, secure, and reliable under all conditions.

Transparency: The principle that AI systems are explainable, clear, and understandable to users and stakeholders.

V

Verification & Validation: A phase in the AI lifecycle where AI models are tested for compliance with ethical and legal standards.

Ministerial Foreword

Artificial Intelligence is changing how we live, work, and engage with the world around us. Governments worldwide face the dual challenge of meeting the changing digital needs and expectations of their citizens while keeping pace with advancements in technology. As Minister for Public Expenditure, Infrastructure, Public Service Reform and Digitalisation, I am committed to Ireland embracing the full potential of emerging technology to deliver better public services and deliver better outcomes.

As a Department, we are committed to tackling societal issues and driving change with best use of emerging technology. To harness the full potential of AI as an accelerator of this process, we must equip public servants with the tools and framework to design, deploy, and maintain AI solutions responsibly. This framework provides practical guidelines, Irish use cases, and recommendations to achieve this.

"The canvas, seven trustworthy principles, and decision framework set out in this document can help us embrace AI in a responsible way as we endeavour to improve public services and deliver for the public."

These Guidelines will compliment and further inform corporate strategies regarding the adoption of innovative technology and ways of working and to set a high standard for public service transformation and innovation, while prioritising public trust and people's rights. The Guidelines have been developed to actively empower public servants to use AI in the delivery of services. By firmly placing the human in the process, I believe these guidelines will go a long way in enhancing public trust in how Government uses AI.

The effort to produce this framework was one of conscious collaboration. The development of these guidelines included the input of an array of public servants, each with varying experience with AI and multi-disciplinary backgrounds and service responsibilities. This approach reflects the recognition that AI will affect the ways of working for a multitude of stakeholders and must be applicable to a wide set of service providers. The use cases included in this document offer valuable insights into the value generated by AI across a number of fields.

Artificial Intelligence presents Government with opportunities to improve public services. By making it easier for public servants to deploy AI solutions, we can address old problems, generate value for the public, and deliver better public services. I strongly encourage public servants to avail of the resources available in AI, including these Guidelines, in the course of their work, and I look forward to seeing their positive impact on our public services.

Jack Chambers, T.D.,
Minister for Public Expenditure, Infrastructure,
Public Service Reform and Digitalisation

Chapter 1:
Executive Summary

In early 2024 the Irish Government made a commitment that Artificial Intelligence (AI) tools used in the public service must comply with seven key principles for Trustworthy AI [1]. These principles, originally established by European Commission's High-Level Expert Group (HLEG), cover; 'Human agency and oversight', 'Technical robustness and safety', 'Privacy and data governance', 'Transparency', 'Diversity, non-discrimination and fairness', 'Societal and environmental well-being' and 'Accountability' [2]. These commitments inform the overarching principles contained in these Guidelines.

This document outlines what each of these principles means for the Irish Public Service and demonstrates our commitment to each. To support public servants in applying these principles, the guidelines provide practical tools and real-world illustrative examples. The purpose of these Guidelines is to provide practical information and resources for all public servants and Government officials on how to design, develop, deploy and maintain AI solutions responsibly.

The responsible use of AI can transform public services to be more efficient, accessible, and responsive. These guidelines will help:

- Promote a common understanding of why and how to use AI responsibly.
- Make informed decisions on whether AI is the right approach for the problem.
- Design and deploy AI solutions that align with the seven principles.
- Identity and address potential risks throughout an AI system's lifecycle.

The seven principles for Responsible AI

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination, and fairness
6. Societal and environmental well-being
7. Accountability

The Guidelines cover four key areas which seek to support public service workers to adopt and use AI in a responsible way. They are:

1. The Seven Principles for Responsible AI

In this section you can learn about the seven key principles for building and using AI responsibly (Chapter 4).

2. A Decision Framework

In this section you will find a framework that you can use at the very start of your AI project to evaluate whether AI is the right solution, and if so, to ensure responsible use (Chapter 5).

3. A Responsible AI Canvas Tool

This section contains a tool to be used at the planning stage of an AI project. You can use this tool to design and map out how to create and deploy AI solutions that meet the Seven Principles for Responsible AI (Chapter 6).

4. AI Lifecycle Guidance

In this section you will find guidance on how to incorporate the Seven Principles and related practices throughout the entire process of your AI project, from development to deployment, and when considering to review or decommission the project (Chapter 7).

The European Union's AI Act, which came into force in 2024, provides the world's first comprehensive legal framework for the regulation of AI in the public interest and for societal and economic benefit. The AI Act classifies AI systems based on risk, setting specific principles for high-risk AI applications while imposing bans on unacceptable-risk AI practices. Given its wide-reaching implications, the AI Act plays a crucial role in shaping how AI is governed across Europe, including within the Irish Public Service.

These guidelines are aligned with the AI Act to ensure that AI deployed in the public service adheres to European regulatory standards. The AI Act reinforces key principles of responsible AI, such as transparency, accountability, and fairness, which are foundational to this document.

While AI can be viewed as a new technology, the approach to adopting AI, as with any technology, should involve the application of all relevant and existing public sector governance frameworks. As such, it is important to consider these Guidelines as a further support to existing organisational governance relating to the adoption of technology, robust data governance, value for money, and innovative ways of working.

The technology and the associated regulatory frameworks for AI and Gen AI are advancing rapidly. To remain effective and relevant, these Guidelines are designed as a living document, adaptable to ongoing changes and emerging best practices. The Guidelines and related implementation tools and resources will be updated regularly to incorporate new developments in AI technologies, changes in the regulatory environment, and lessons learned from real-world applications. This iterative approach ensures that the Guidelines continue to provide robust, actionable recommendations that align with technological advancements and the evolving expectations of the people we serve.

Chapter 2: Introduction

2.1 What is AI?

The most important definition of AI, for Irish Public Service workers, can be found in the EU AI Act (Regulation (EU) 2024/1689):[3]

> *"An AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".*

However, it is worth mentioning that the definition of AI is much debated. One of the most widely used definitions globally is that by the OECD which can be found below:

> *"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment"* [4].

2.2 Overview of the Guidelines

AI tools and technologies have significant potential to transform the delivery of public services and to help in dealing with big societal challenges. With AI and other digital technologies reshaping the workplace, the Irish Public Service is committed to setting a standard for responsible AI.

While AI offers exciting possibilities and benefits for the public service in terms of efficiency, effectiveness, and responsiveness, it also presents risks that could impact individuals and society. Without the proper safeguards and controls, AI systems can unintentionally amplify unfair or undesirable outcomes for individuals and communities.

By its nature, the work we do as public service workers has an impact on people's lives and wellbeing. It is essential that we fully understand the potential risks involved in using AI and what safeguards are needed. Recognising this, the Department of Public Expenditure, Infrastructure, Public Service Reform and Digitalisation has developed these guidelines to ensure that AI is approached and used responsibly. By prioritising responsible AI, we are committed to building and maintaining public trust.

While AI can be viewed as a new technology, the approach to adopting AI, as with any technology, should involve the application of all relevant and existing public sector governance frameworks. As such, it is important to consider these Guidelines as a further support to existing organisational governance relating to the adoption of technology, robust data governance, value for money, and innovative ways of working.

The recommendations in these guidelines should be applied by public service workers or any third party working on behalf of the public service. It is relevant to any in-house solution developed or the adoption and deployment of one that is acquired from external vendors.

While Responsible AI is our ambition in the Irish Public Service, our legal obligations are largely defined by the EU AI Act (2024) the General Data Protection Regulation (GDPR) (2018) and the Data Sharing and Governance Act (2019).

2.3 Purpose

These Guidelines support the 'Better Public Services' transformation strategy's vision to deliver "inclusive, high quality and integrated public service provision that meets the needs and improves the lives of the people of Ireland" [5].

The Guidelines cover four key areas which seek to support public service workers to adopt and use AI in a responsible way. They are:

1. The Seven Principles for Responsible AI

In this section you can learn about the seven key principles for building and using AI responsibly (Chapter 4).

2. A Decision Framework

In this section you will find a framework which can be used at the very start of your AI project to evaluate whether AI is the right solution, and if so, to ensure responsible use (Chapter 5).

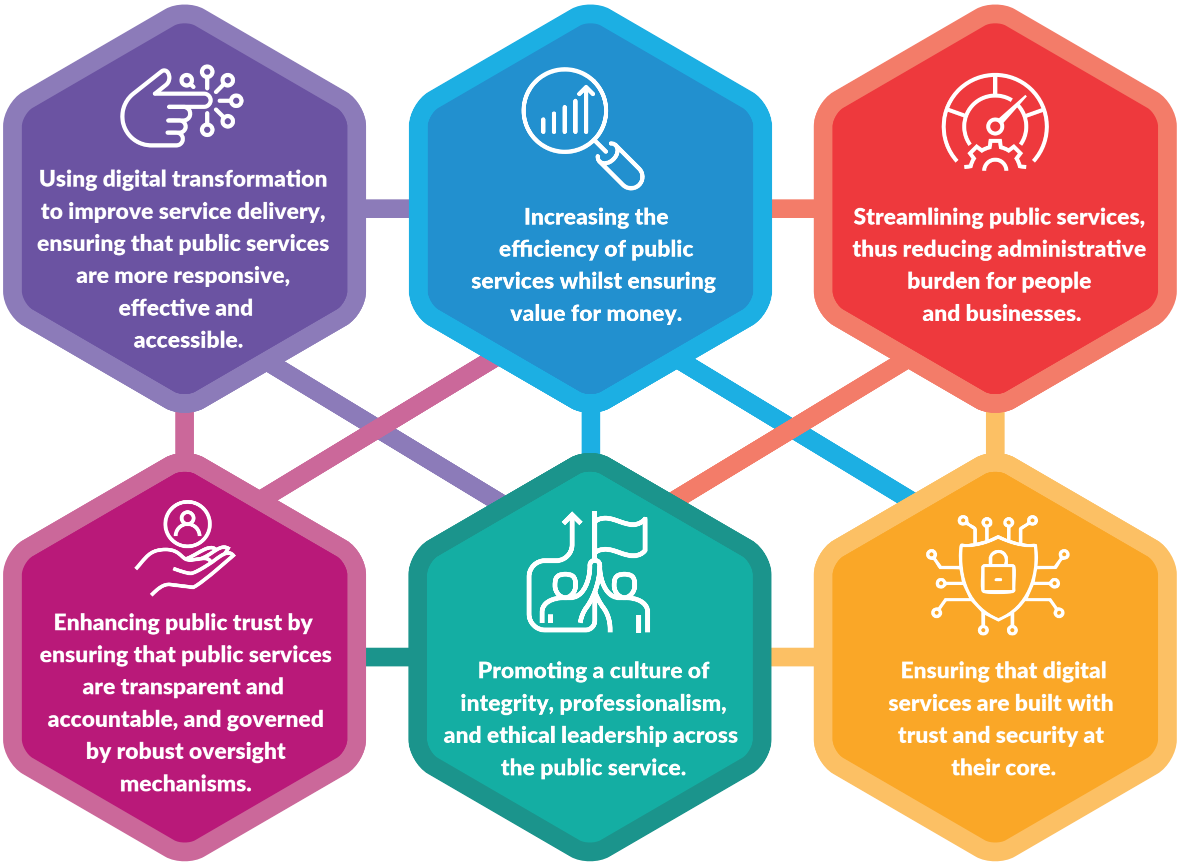
3. A Responsible AI Canvas Tool

This section contains a tool you can use at the planning stage of an AI project. You can use this tool to design and map out how to create and deploy AI solutions that meet the Seven Principles for Responsible AI (Chapter 6).

4. AI Lifecycle Guidance

In this section you will find guidance on how to incorporate the Seven Principles and related practices throughout the entire process of your AI project, from development to deployment, and when considering reviewing or decommissioning the project (Chapter 7).

Implementing of these guidelines will contribute significantly towards achieving a number of the objectives set out in ‘Better Public Services’, namely:



Using digital transformation to improve service delivery, ensuring that public services are more responsive, effective and accessible.

Increasing the efficiency of public services whilst ensuring value for money.

Streamlining public services, thus reducing administrative burden for people and businesses.

Enhancing public trust by ensuring that public services are transparent and accountable, and governed by robust oversight mechanisms.

Promoting a culture of integrity, professionalism, and ethical leadership across the public service.

Ensuring that digital services are built with trust and security at their core.

2.4 Audience

While guidance can be put in place to ensure AI is used safely and responsibly, it is the responsibility of all public service workers to ensure that these measures are followed. Therefore, all public service workers should familiarise themselves with these guidelines to understand how AI can be used responsibly. The following groups may find the Guidelines of particular relevance to their roles and responsibilities:

- **Public Service leaders** are responsible for ensuring that AI systems align with public interests and comply with legal frameworks, including the EU AI Act (Regulation (EU) 2024/1689). Leaders must put in place adequate measures to ensure the responsible use of AI systems, and these guidelines outline the key steps required.
- **IT, data, analytics, and AI professionals:** These workers are responsible for following these guidelines when developing and implementing AI models.
- **Users of IT systems:** These users must ensure that AI system outputs are appropriately critiqued and that the oversight measures outlined in these guidelines are effectively implemented.
- **AI Providers and AI Deployers:** The guidelines offer specific advice for these roles, as defined under the EU AI Act.

2.4.1. ‘Providers’ & ‘Deployers’ of AI

Understanding the difference between an AI Provider and an AI Deployer is crucial because under the EU AI Act each has distinct responsibilities and compliance principles that affect how AI is developed, used, and regulated. This is also relevant to how these Guidelines apply. It is important to note that these two roles are not the only two outlined in the AI Act, but they are the most common for public service projects.

2.4.1.1 AI Providers:

These are defined in the EU AI Act (Regulation (EU) 2024/1689) as *“a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge”* [6].

Illustrative Example

A Public Service Body is developing and deploying an AI system for internal use under its own authority. Thus, in this instance it acts as both the **AI Provider** and the **AI Deployer**.

2.4.1.2 AI Deployers

These are defined in the EU AI Act as *“a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity”* [6].

Illustrative Example

A Public Service Body is looking to implement an AI system. The AI system will be developed by an external vendor who are the AI Providers. In this case the Department will be the **AI Deployer**.

2.4.1.3 How the same party can be both the Provider and the Deployer

AI Providers are generally subject to more stringent obligations than AI Deployers. However, as in the first example, the same party can be both the Deployer and the Provider. These scenarios include the following: [7]

1. When the Deployer puts their name or trademark on a high-risk AI system.
2. When the Deployer makes a substantial modification to a high-risk AI system.
3. When the Deployer modifies the intended purpose of an AI system.

Given that there are different obligations associated with Providers and Deployers, understanding who will assume these roles is a key step during the planning stage of the AI lifecycle. This will help to ensure compliance with the relevant obligations in the EU AI Act (Regulation (EU) 2024/1689).

Chapter 3:
AI in the Public Service

AI can be applied across various tasks, helping the Irish Public Service address everyday challenges and improve services. This chapter provides examples of the types of tasks that AI can be used for in the delivery and improvement of public services and outlines the key benefits and risks.

3.1 Types of tasks AI can be used for in public services.

AI can support a wide range of tasks within the public service, offering new ways to improve efficiency, enhance service delivery, and address complex challenges. The OECD has classified the opportunities AI could be used for into a list of prospective tasks. This could be used as a source of inspiration for Irish Public Service workers when considering the opportunities that AI presents. The table below outlines common AI tasks, describes what each does, and provides practical examples: [8]

Task	What it does	Examples
Recognition	Using AI to identify and categorise data (e.g. images, video, audio and text) into specific classifications.	Document Analysis: Automatically extracting information from scanned documents like reports or forms. Disease Diagnosis: Recognising patterns in medical images (e.g., X-rays or scans) to detect potential illnesses early.
Event detection	Using AI to detect and monitor patterns, outliers or anomalies, often in real time.	Fraud detection: Continuously analysing financial transaction data to identify anomalies or patterns of activity inconsistent with user behaviour that may indicate fraud. Infrastructure monitoring: Detecting faults or structural weaknesses in public infrastructure. Crime monitoring: Identifying suspicious activity using CCTV feeds or social media analysis.
Forecasting	Using past and existing behaviour and data to predict future outcomes, to inform decision-making.	Revenue and Expenditure Forecasting: Analysing economic trends, tax collection data, and spending patterns to predict future revenues and expenditure needs. Crop yield forecast: Predicting crop output, based on climate patterns, crop types and historical yields.

Task	What it does	Examples
Personalisation	The use of AI to tailor services, content or experiences to the individuals user needs.	User Queries: Adapting portals or Chatbots to provide relevant information to specific users. Education: Providing customised learning experiences for individual learners based on their progress. Healthcare: Providing personalised health plans based on patient history and genetic data.
Interaction support	The use of AI to enhance communication and interaction between users and systems.	Virtual Assistants / Chatbots: Using conversational AI techniques to support users to navigate Government websites. Enhanced web accessibility: Converting text to speech for visually impaired users interacting with public services. Language translation: Providing translation of Government-issued information into multiple languages.
Goal-driven optimisation	The use of AI algorithms to optimise processes or decision-making	Reducing carbon emissions: Optimising energy use in public buildings or transportation networks, reducing overall carbon footprints. Predictive maintenance: Ensuring public transport infrastructure and vehicles are serviced and maintained proactively to avoid breakdowns and service disruptions. Public health management: Predicting patient inflows during seasonal surges to reduce overcrowding and improve care.

Task	What it does	Examples
Reasoning with knowledge structures	The use of structured data to determine new insights, solve problems, or answer complex queries.	Policy Analysis: Evaluating how changes in policy may impact on other sectors using knowledge graphs. Legal Argument: Analysing case law and suggesting relevant precedents. Urban Planning: Identifying optimal locations for new schools or hospitals based on demographic data.
Content generation	The use of AI algorithms and models to generate new content such as text, images, audio or video based on a set of input data, parameters, or prompts.	Report Drafting: Analysing lengthy documentation and generating summaries or draft versions that contain the most important data points and insights. Translation Services: Automatically generating translations of Government publications. Public Awareness Campaigns: Creating videos or infographics to educate the public about relevant initiatives.

3.2 Benefits of AI for Better Public Services

AI has the potential to transform Irish Public Services, making them more efficient, fair, and responsive. According to The Organisation for Economic Co-operation and Development (OECD) study, AI's positive impact on public services can be categorised into three core areas: productivity, responsiveness, and accountability. These categories reflect the ways in which AI can streamline operations, better meet public needs, and maintain trust.

3.2.1 Productivity (Efficiency and Effectiveness)

3.2.1.1 Efficiency

AI can increase productivity by automating complex, repetitive tasks, allowing public servants to focus on higher-value work. According to The Organisation for Economic Co-operation and Development (OECD) study, AI could be used to increase the "efficiency of internal operations, by automating complex but repetitive administrative processes and procedures to support and facilitate the productive work of public officials, free up the time of skilled civil servants and ensure the reliability of the continuous delivery of public services" [8].

Example: A Department is introducing an AI system to assist in the streamlining of a grant application process. The AI provides automation for the initial screening of application to check if the required information and documentation provided meets certain predefined criteria. This automation reduces the time spent by staff on repetitive tasks, allowing them to focus on other things. However, human oversight remains key here and a team member remains responsible for overseeing the performance of the AI system.

3.2.1.2 Effectiveness

AI can improve the effectiveness of policymaking by analysing large datasets to understand user needs and detect patterns. According to the OECD study, AI could be used to improve the "effectiveness of policymaking by using large amounts of data to gain more granular insights into user needs and find patterns. This, in turn, would allow Government to formulate more targeted policies and deliver better outcomes, by better targeting social expenditures, public investments and Government services" [8].

Example: A Department is using machine learning to analyse data on service usage for its Department. By identifying the trends in different regions, the AI system helps the team understand where demand is highest, allowing them to allocate resources more effectively and put the right initiatives in place where they are most needed.

3.2.2 Responsiveness

According to the OECD study, AI could be used to improve "Governments' ability to anticipate societal trends and user needs". This enables the delivery of proactive, personalised, and human-centred public services, making the Irish Public Service more adaptable and attuned to the evolving needs of the public.

Example: An AI-powered chatbot is being implemented by a Government Department to help answer queries from the Irish public. This chatbot will reduce wait time for people and allowing them to conveniently self-serve outside of traditional business hours.

3.2.3 Accountability

According to the OECD study, data analytics and machine learning could be used to "detect fraud and risks to public sector integrity by identifying irregularities or suspicious patterns and raising red flags". This strengthens accountability, ensuring that public resources are managed responsibly.

Example: A Public Service Body is implementing an AI system to review grant applications for potential fraud or irregularities. These cases can then be directed to a human for further investigation. This is an example of how AI can be used to strengthen oversight mechanisms and make savings by identifying more fraudulent behaviour.

3.3 Risks associated with using AI in Public Services

The adoption of AI in public services presents significant benefits but also introduces various risks. The European Commission has identified key challenges related to AI technology adoption in the public sector. [9]

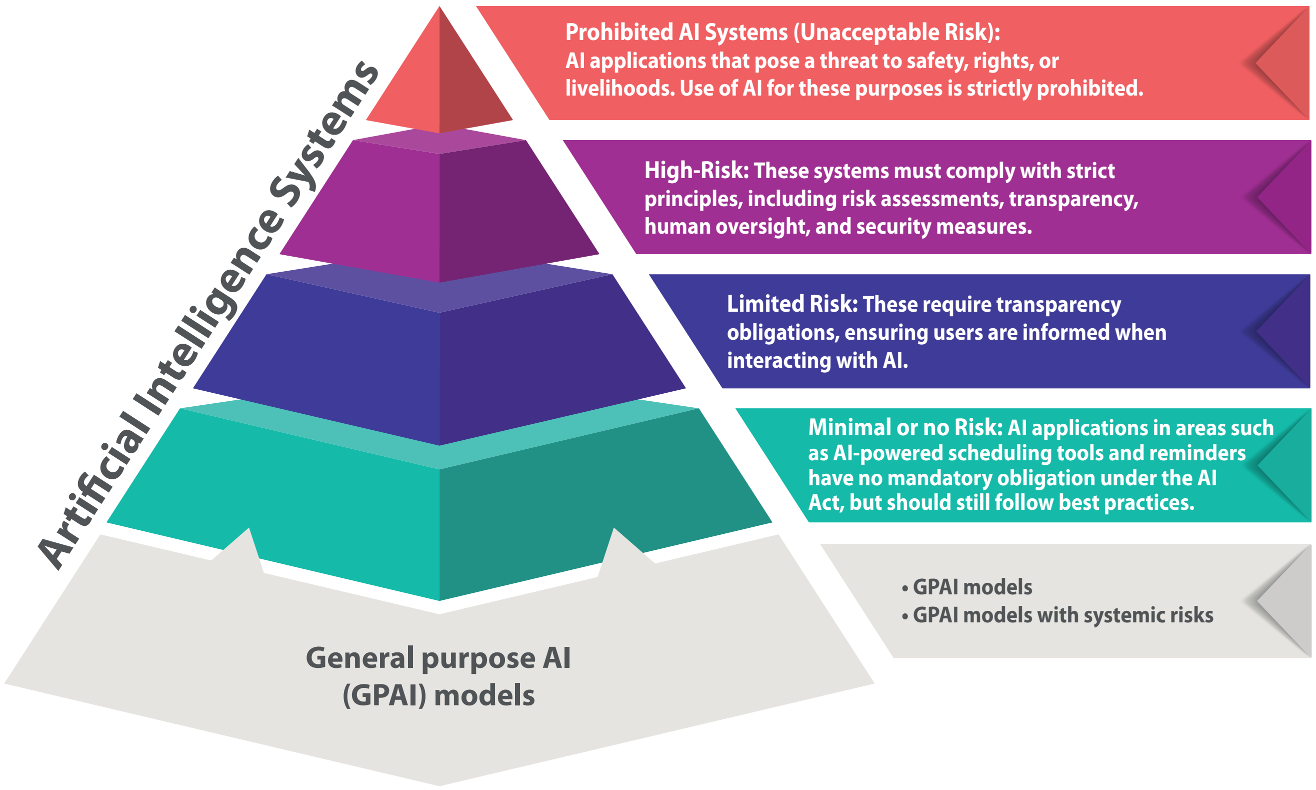
- Data bias and discrimination: AI systems may reinforce biases present in the data, leading to unintended discrimination.
- Transparency and explainability: Where complex "black-box" algorithms can make AI-driven decisions difficult to understand, affecting public trust.
- Dehumanisation of services: Automated AI systems may lack flexibility for a lot of cases, risking a loss of the human touch in public services.

Addressing these challenges is essential to ensure AI is implemented responsibly.

3.3.1 The EU AI Act's Risk-Based Approach

The EU Commission published the Guidelines on prohibited AI practices as defined in the EU AI Act (2024) in February 2025. These offer valuable insights into the Commission's interpretation of the prohibitions. The guidelines provide legal explanations and practical examples to help stakeholders understand and comply with the AI Act's principles. However, they are non-binding and authoritative interpretations is reserved for the Court of Justice of the European Union (CJEU). [10]

For practices that are not defined as prohibited in the EU AI Act (2024), the EU AI Act categorises AI systems into different risk levels, requiring Public Service Bodies to assess AI applications accordingly. This section illustrates the various levels of risk as set out in the EU AI Act and identifies the types of uses that are categorised under each level.


Artificial Intelligence Systems
Prohibited AI Systems (Unacceptable Risk): AI applications that pose a threat to safety, rights, or livelihoods. Use of AI for these purposes is strictly prohibited.
High-Risk: These systems must comply with strict principles, including risk assessments, transparency, human oversight, and security measures.
Limited Risk: These require transparency obligations, ensuring users are informed when interacting with AI.
Minimal or no Risk: AI applications in areas such as AI-powered scheduling tools and reminders have no mandatory obligation under the AI Act, but should still follow best practices.
• GPAI models
• GPAI models with systemic risks
General purpose AI (GPAI) models

Unacceptable Risk (Prohibited Systems)

AI systems that threaten people's safety, rights, or livelihoods are strictly prohibited, except in limited circumstances. The ban on AI systems classified as posing unacceptable risk came into force in February 2025 [11] [12] [7].

Example
A Public Service Body proposes using AI to categorise people by ethnicity based on facial features (biometric categorisation).

EU AI Act Rating: Unacceptable Risk.

Outcome: This use is strictly prohibited under the EU AI Act and cannot be implemented.

Rationale: This prohibition is in place to protect people's rights and freedoms.

High-Risk

These systems are subject to the most stringent regulations. They must comply with "strict principles, including risk-mitigation systems, high quality of datasets, logging of activity, detailed documentation, clear user information, human oversight, and a high level of robustness, accuracy, and cybersecurity" [13].

Example
A Public Service Body is looking to implement an AI use case under what the EU AI Act considers critical infrastructure.

EU AI Act Rating: High Risk. If the system was to fail or malfunction it "may put at risk the life and health of persons at large scale and lead to appreciable disruptions in the ordinary conduct of social and economic activities" [14]. Thus, if this Department did want to implement this AI use case, it has to meet the stringent principles set out by the EU AI Act.

Rationale: Given the negative impact this use case could have on Irish society if it malfunctioned, the obligations associated with high-risk use cases ensures appropriate safeguards and due diligence are carried out to help protect the public.

Limited Risk

These are AI systems that present only limited risks (e.g. chatbots or AI systems that generate content). These limited risk AI systems are subject to transparency obligations, so the end-user is aware that content was generated using AI [11].

Example
An AI-powered chatbot is being implemented by a Public Service Body to help answer queries from the Irish public.

EU AI Act risk classification: Limited Risk. The Department must ensure that people are notified that they are interacting with an AI system when using the chatbot.

Rationale: The public has a right to know when they are interacting with an AI system, rather than a human, so they can judge the interaction appropriately.

Minimal or no Risk

These are AI systems that pose minimal or no risk (e.g. AI in video games). AI systems with this risk level are not regulated or affected by the EU AI Act [11].

> **Example**
> A Public Service Body is looking to use AI to provide language translation, speech-to-text, and text-to-speech services for better accessibility.
>
> **EU AI Act risk classification:** Minimal or no risk. This type of AI system is not legally regulated under the EU AI Act. However, the Department should still follow the guidelines in this document to create the AI system responsibly. The Department should also ensure that the AI system functions as planned and monitor its use to ensure that there are not any unintended consequences.
>
> **Rationale:** In situations like this where there is minimal, or no risk, AI technology can be used freely to improve public services and for the benefit of the public.

Additionally, tools like the **EU AI Act Risk Classifier & Compliance Checker** can assist in determining an AI system's risk level and regulatory obligations. By proactively assessing AI risks and applying necessary safeguards, the Irish Public Service can maximise AI's benefits while ensuring ethical, transparent, and responsible implementation.

Chapter 4: Principles for Responsible Use of AI in the Public Service

The seven principles for Responsible AI are:

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination, and fairness
6. Societal and environmental well-being
7. Accountability

These are the seven guiding principles informing all commitments and recommendations set out in this document. They are aligned with the European Commission's HLEG's seven principles for Trustworthy AI [2] which were specifically adopted by the Irish Public Service in early 2024.[1] They are part of a recommended governance structure which covers both AI and Generative AI, and spans technical and non-technical aspects of AI adoption. Figure 1 below presents a comprehensive framework for the responsible use of AI in the public service, spanning the principles, regulatory compliance, and the Better Public Services strategy's commitment to user-centred service design and delivery.

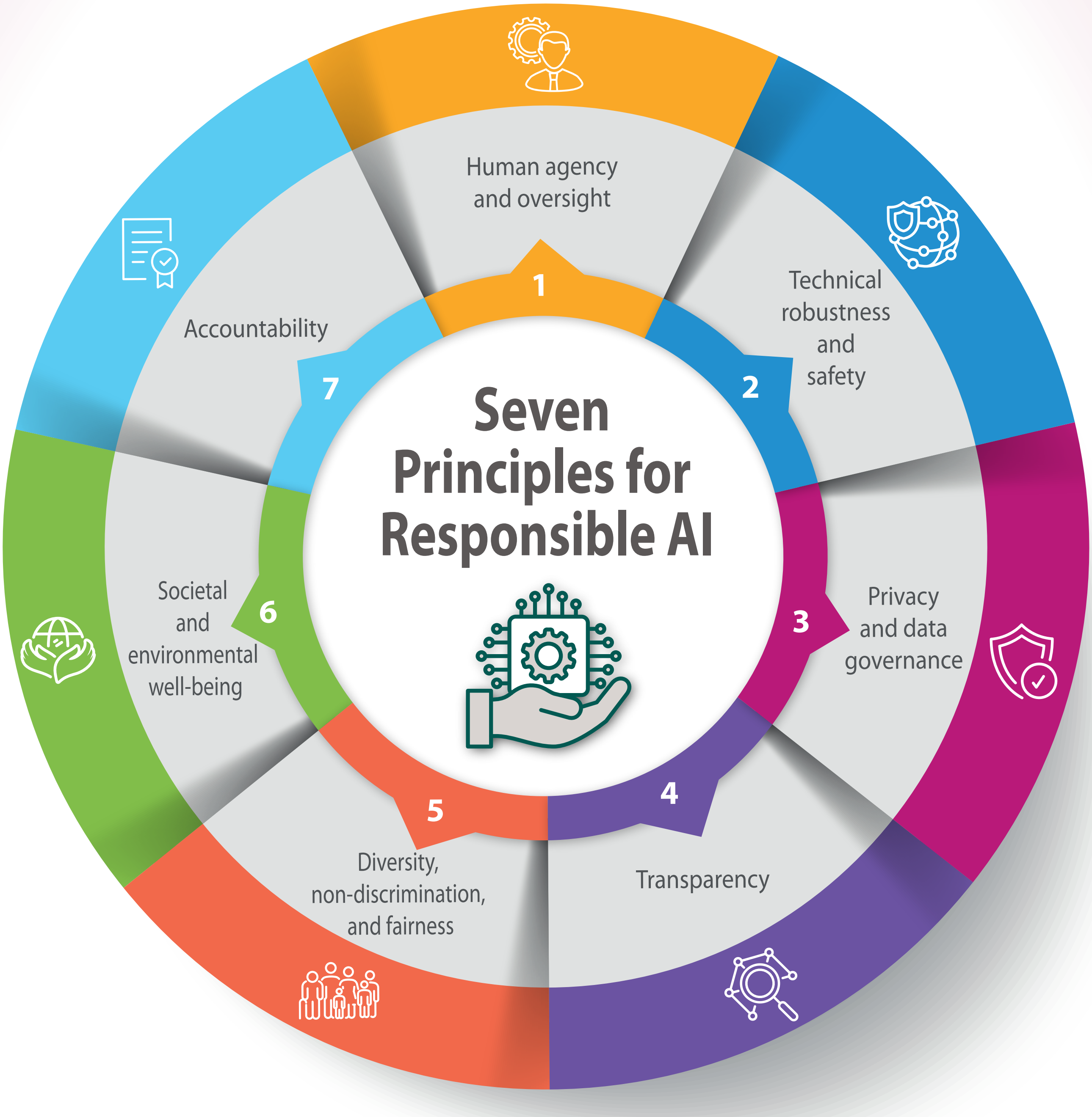

Seven Principles for Responsible AI
1
Human agency and oversight
2
Technical robustness and safety
3
Privacy and data governance
4
Transparency
5
Diversity, non-discrimination, and fairness
6
Societal and environmental well-being
7
Accountability

Figure 1: *The Irish Public Service Responsible AI Framework*

4.1 Human Agency and Oversight

4.1.1 Overview

AI can create efficiencies and empower better decision-making in the public service. However, it is essential that we always stay in control of these systems. When the right safeguards and human oversight is in place, AI systems can be used very effectively to support decision-making rather than replacing it.

According to the European Commission, *"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"* [15].

4.1.2 What this principle means for us in the Irish Public Service

The below guidance has been adapted from the European Commission's HLEG's publication [2].

Fundamental rights

Given the reach and capacity of AI systems, they can negatively affect fundamental rights. In situations where such risks exist, a fundamental rights impact assessment may be undertaken. Conducting this prior to development will also help minimise inefficiency. If a fundamental rights impact assessment determines that the AI system could potentially affect people's rights and is still being progressed, they must be able to give feedback on how it impacts them.

Human agency

We will provide users with the necessary knowledge and tools to understand and where feasible, evaluate or contest AI systems. If our fundamental rights impact assessment determines that decisions made by the AI system could significantly affect an end-user, provision may be made so that people can opt-out of decisions made solely by AI, as appropriate.

Human oversight

All AI tools used in the public service must be part of a process that has human oversight built into it. This may include a human directly overseeing decisions in real time ("human-in-the-loop"), allowing the system to operate autonomously with the ability to be overridden or stopped by a human as needed ("human-on-the-loop") or at a higher level ("human-in-command").

The more autonomy the AI system has, the stricter the oversight and testing will need to be. AI tools can assist human capabilities, but they should never replace them.

Important elements to consider:

- **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 Generative 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.

4.1.3 Benefits of this Principle

- **Facilitates accountability:** Having human oversight structures in place ensures human responsibility remains central in decision-making, allowing public trust in AI applications.
- **Enhancing public adoption of AI:** AI systems with proper agency and oversight will help the adoption and effectiveness of AI systems for the Irish public.

4.1.4 Illustrative Example

A Public Service Body is looking to implement an AI system to help process applications faster by screening for basic eligibility criteria. The AI system will create efficiencies in the Public Service Body. However, someone in the organisation must remain accountable for each component of the AI lifecycle.

4.2 Technical Robustness and Safety

4.2.1 Overview

Our AI systems should be dependable and perform as expected. We manage sensitive data on behalf of people, so it is our responsibility to have the best-in-class security practices in place. Every public service worker has a role in protecting this information from intentional or accidental exposure. According to the European Commission, "AI systems need to be resilient and secure. They need to be safe, ensuring a fallback 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 minimised and prevented" [15].

4.2.2 What this principle means for us in the Irish Public Service

The below guidance has been adapted from the European Commission's HLEG's publication [2].

Resilience to attack and security

Ensure third-party AI systems, or those internally developed, comply with our IT security policies and standards such as guidance from the Data Protection Commission and the National Cyber Security Centre. Protecting the integrity of our systems is key in maintaining public trust, especially given the sensitivity of the data we have been trusted with managing.

Fall-back plan and general safety

Ensure that a clear procedure has been created that outlines what to do if an AI system fails or malfunctions. This might involve switching to a simpler system or allowing human controllers to step in. The stakeholder accountable for the deployment stage of the AI lifecycle must ensure this fallback plan can be activated if needed before the AI system is deployed.

Ensure a full risk assessment is conducted. This should cover all application areas the AI system will be subject to.

Accuracy, reliability, and reproducibility

Establish procedures for testing and monitoring throughout the full AI lifecycle. Testing during the 'Verification and Validation' stage of the AI lifecycle must ensure these procedures were followed and comprehensive testing was done on the model accuracy. The testing should simulate all potential conditions that the AI system will be subject to.

Where reproducibility is required, this should be specifically assessed. However, this is not possible to achieve in some use cases where GenAI is used given the nature of the solution. On that basis, users should consider the appropriateness of GenAI for a task where reproducibility is required but may not be possible. When 100% model accuracy is not possible, we need to know the likelihood of errors and clearly document our rationale to still deploy the model and the ways we mitigated risk.

During the 'Deployment' and 'Operation and Monitoring' stages of the AI lifecycle, we must assess if the system is performing in line with testing under real world conditions. Evidence of worse than expected performance must be escalated back to the development team for refinements.

4.2.3 Benefits of this Principle

- **Respects the data we have been trusted with managing:** We have been trusted with managing some of the public's most sensitive data. This principle helps to ensure that we have the best processes in place to protect this data and justify the trust placed in us.
- **Encourages proactive contingency planning:** If something goes wrong, having a clear plan means we can act quickly to minimise any harm to the people affected.
- **Puts a focus on reliability:** By having technically robust systems, we will be better able to provide consistent delivery of vital public services.

4.2.4 Illustrative Example

A Public Service Body is looking to implement a new AI system. The Public Service Body must ensure that the system has best-in-class security. The system could otherwise be attacked, leading it to make different decisions or causing it to shut down altogether.

4.3 Privacy and Data Governance

4.3.1 Overview

Building and maintaining public trust is key. Central to this is ensuring that we have robust policies and procedures in place to protect personal data, as well as setting high standards for data governance. It is essential that all AI systems and applications in the Irish Public Service fully respect data protection laws, including the GDPR (2018).

4.3.1.1 Types of datasets

One of the first things to consider is what collection of data, or dataset, will be used by the proposed AI system application. This will help to determine the level of risk involved and other actions that may need to be taken. The two factors to be considered at this stage are:

1. Does the intended dataset, to be used for AI processing, contain any personal data? The definition of "***personal data***" set out in the GDPR (2018) [25] means *any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.*

Therefore if the dataset contains any data that comes within the above definition, then the GDPR does apply and further steps will have to be taken to determine the level of risk and appropiate actions to follow.

2. However, if the dataset does not fall under the definition of personal data and is data that is entirely non-personal in nature then the GDPR (2018) does not apply. Also, if a dataset is properly anonymised and cannot be reconfigured to either directly or indirectly link back to an individual then it may not be subject to the GDPR (2018) principles. In this instance, under the correct circumstances, there is potential for the use of synthetic data, as outlined in a guidance by The European Data Protection Supervisor. [16]

4.3.1.2 Guidelines for Processing Personal Data by AI

Legal Basis

There are two types of legal basis that could potentially be used and which will depend on the level of risk identified from a Data Protection Impact Assessment (DPIA).

Consent of the individual could be used in low risk scenarios such as a Chatbot. Additional privacy by design features could be included to ensure that the AI machine learning only processes the data that is necessary and relevant to provide correct answers to queries and does not process any personal information.

High risk processing could involve the profiling of individuals on a large scale for the performance of a task in the public interest. In these scenarios the legal basis may require a specific legislative mandate under primary legislation or regulations.

Any processing of data that does entail legislative change will require a mandatory consulation with the Data Protection Commission and completion of a legislative consultation form. [17]

If the AI processing involves any special categories of data as defined in the GDPR (2018) then extra precautions will be needed. It is advisible to contact the Data Protection Commission before commencing any proposal of this nature.

According to the European Commission, "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" [15].

4.3.2 What this principle means for us in the Irish Public Service

The below guidance has been adapted from the European Commission's HLEG's publication [2].

Privacy and data protection

We will ensure that AI systems respect data privacy and protection throughout their entire lifecycle. Data used in an AI model must comply with the GDPR (2018) principles. We must complete processes like Data Protection Impact Assessments (DPIAs) to identify and mitigate risks. We will only collect data for specific purposes, use it fairly and never allow it to discriminate unfairly against an individual. Where the system is procured from a third-party vendor, they must confirm that their data is GDPR (2018) compliant and does not breach the intellectual property rights of others.

Quality and integrity of data

Before training any AI system, the data must be carefully checked to fix any errors and address potential biases. Clear policies and procedures for managing data, including how it is stored, used, and deleted, must be followed in line with organisational policies. Testing and documentation should cover every step of the AI lifecycle. Good quality data is the foundation of any successful AI system.

Access to data

We will ensure that there are clear protocols in place to determine who should have access to data. Access to the user data should be restricted only to those who have a legitimate need in line with organisational policies in place.

4.3.3 Benefits of this Principle

- **Aligns with our GDPR (2018) responsibilities:** By complying with our GDPR (2018) obligations, we can avoid situations that could hamper public trust.
- **Encourages top-class data governance:** Maintaining excellent data governance will help enhance data quality. This will produce better AI systems to benefit the Irish public, as well as creating some operational efficiencies.
- **Puts a focus on data access:** Emphasises the importance of maintaining strict access controls to safeguard our data. Failure to do so could negatively impact public trust.

4.3.4 Illustrative Example

A Public Service Body is looking to introduce a new AI system. They will need to engage with the Data Protection Officer and complete a Data Protection Impact Assessment for the new AI system. The Public Service Body must ensure that the data is handled securely, that only necessary data is used, and that privacy is protected. Should the new AI system not perform as expected, they will need to ensure that there is a contingency plan in place that can be activated at short notice.

4.4 Transparency

4.4.1 Overview

We are committed to being transparent and upfront with end-users on the AI systems being used. This includes notifying end-users when they are interacting with an AI system (e.g. a chatbot) or where an output has been generated by an AI system. We are committed to explaining key aspects of the AI system in terms that the end-users understand. We will maintain high standards of documentation. This will ensure informed oversight, facilitate auditability and corrective action should something go wrong.

According to the European Commission, "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" [15].

4.4.2 What this principle means for us in the Irish Public Service

The below guidance has been adapted from the European Commission's HLEG's publication [2].

Traceability

Clear documentation on AI model development must be maintained. This includes aspects such as the data used, processing that was carried out and algorithms used, the rationale for choices and for compromises made. If something goes wrong, this documentation will help us understand what happened, fix it quickly and prevent the same error in the future.

Explainability

We will be open about the data the AI system uses and how it makes decisions. This is particularly important if the AI system affects people's lives. We will explain our AI systems in terms that end users can understand.

Communication

We must clearly inform users when they are interacting with an AI system or when some outputs have been developed with the aid of AI. This applies especially to systems like chatbots, AI agents and automated decision-makers. The end-user should be made aware of AI involvement to ensure informed engagement and provided alternative service options where appropriate.

4.4.3 Benefits of this Principle

- **Fosters public trust:** Transparency will help to build trust amongst the Irish public by making it apparent where an AI system was used and how a decision was made in terms they understand.
- **Facilitates traceability and auditability:** Having clear documentation demonstrates the careful due diligence we carried out in building our responsible AI solution. The documentation also enables swift corrective action when something goes wrong, as well as facilitating auditing to enhance trust.

4.4.4 Illustrative Example

An AI chatbot is being implemented by a Public Service Body to help answer queries from the Irish public. The organisation must ensure that people are notified that they are interacting with an AI system when using the chatbot.

More complex queries may need human intervention. Public Service Bodies should ensure these queries can be redirected to a human to answer them.

4.5 Diversity, Non-discrimination, and Fairness

4.5.1 Overview

We are committed to making sure our AI systems are fair and inclusive. We play a vital role in access to key services such as health, education, and welfare, so we must make every effort to ensure our AI systems are equitable. Unfair bias in AI could harm individuals, communities, and society. Thus, we must undertake comprehensive bias detection and implement mitigation strategies where required.

The European Commission dictates that "unfair bias must be avoided, as it could have multiple negative implications, from the marginalisation 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" [15].

4.5.2 What this principle means for us in the Irish Public Service

The below guidance has been adapted from the European Commission's HLEG's publication [2].

Avoidance of unfair bias

We will conduct appropriate due diligence to proactively detect any inherent biases both prior to deployment and post deployment. This ensures that there are no unwarranted or unexpected outcomes. When working on an AI project, it is vital to incorporate diverse perspectives. The solution must be inclusive in design and informed by the Design Principles for Government in Ireland as appropriate.

Algorithmic bias can be caused by the use of training data that reflects existing systemic biases or by not being representative of the population that the AI system will be used by. Particular care is needed to ensure that vulnerable groups are not adversely affected. The model itself must also be ethically balanced so proper oversight is needed.

Accessibility and universal design

We will design our AI systems with accessibility and universal design principles in mind, ensuring that they are user-centric and inclusive for all individuals.

Stakeholder participation

In order to develop AI systems that are trustworthy, it is advisable to consult stakeholders who may directly or indirectly be affected by the system throughout its lifecycle. It is beneficial to solicit regular feedback even after deployment.

4.5.3 Benefits of this Principle

- **Promotes fairness and equality:** Designing fair AI systems helps reduce discrimination and improves public service accessibility.
- **Avoids ethical risks:** By mitigating bias as much as possible, we can ensure our desired AI systems meet ethical standards.

4.5.4 Illustrative Example

A Government Department is looking to introduce a new AI system. Before working on the AI system, the Department must assess if there is, for example, unfair bias in the dataset. Issues such as this must then be resolved and mitigated as much as possible.

The training data used should be representative of the population and not favour one group of people over another. The Department should look to adopt the design principles for Government in Ireland and consult with stakeholders throughout the entire AI lifecycle.

4.6 Societal and Environmental Well-being

4.6.1 Overview

We are committed to ensuring that all AI systems in the Irish Public Service have a positive impact on Irish society. According to the European Commission, "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" [15].

4.6.2 What this principle means for us in the Irish Public Service

The below guidance has been adapted from the European Commission's HLEG's publication [2].

Sustainable and environmentally conscious AI deployment

We must ensure that, while benefiting from the power of AI, we do so in a means that is as environmentally conscious as possible. We have committed to reducing carbon emissions in line with Government's Climate Action Plan. We must ensure that any environmental harm related to an AI system is mitigated as much as possible.

Publications such as the OECD's 'Measuring the environmental impacts of Artificial Intelligence Compute and applications' [18] have explained that "reducing the environmental impacts of AI is to some extent linked to reducing the environmental impacts of information and communication technology (ICT) systems more generally". Where possible, efforts should be made to reduce energy consumption during model training. Adopting good practices, such as using pre-trained models where possible or powering data centres with renewable resources where suitable, will help.

Social impact

AI systems can positively impact our social well-being. However we must also consider the potential negative impacts on people's physical and mental health. Ongoing monitoring will be needed to ensure these systems do not negatively impact users. Furthermore, the public service plays a crucial role in maintaining public trust in institutions and supporting democracy. Any AI system that runs the risk of negatively impacting the democratic process should not be deployed.

4.6.3 Benefits of this Principle

- **AI used as a means of good:** This principle puts a focus on the broader impact of AI systems. We should always work to ensure that AI contributes positively to society, outweighing any potential negative effects it may have.

4.6.4 Illustrative Example

A Government Department is looking to introduce a new AI system. They must consider the full societal impact of the AI system, such as the environmental impact or anything that could negatively affect people's physical or mental wellbeing. The Department must try and mitigate these concerns as much as possible. They must evaluate if the benefits to society of the AI.

4.7 Accountability

4.7.1 Overview

For all public service AI systems, we are committed to identifying clear lines of responsibility. At every stage of an AI system's development and operation, there will be a person responsible. We will maintain audit trails to make sure decisions made by AI are clear and traceable. If things go wrong, people will have a way to seek help and raise concerns.

According to the European Commission, "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" [15].

4.7.2 What this principle means for us in the Irish Public Service

The below guidance has been adapted from the European Commission's HLEG's publication[2].

Accountability

We will specify and document who holds responsibility at each stage of the AI system's lifecycle. In situations where someone can no longer fulfil that responsibility, it must be designated to another individual who has the competencies to carry out that role. The documentation must be updated accordingly. 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.

Auditability

This does not necessarily imply that information about business models and intellectual property related to the AI system must always be openly available. Evaluation by internal and external auditors and the availability of such evaluation reports can contribute to the trustworthiness of the technology. In applications affecting fundamental rights, including safety-critical applications, AI systems should be able to be independently audited.

Minimisation and reporting of negative impacts

It is recommended that relevant risk assessments and safe usage policies are in place for any Public Service Body considering the use of an AI tool. We have a duty to minimise the negative effects associated with AI systems. If someone has concerns about an AI system, they will be protected if they raise those concerns.

If someone is negatively affected by an AI decision, we will provide them with clear information about how to raise the issue.

4.7.3 Benefits of this Principle

- **Fosters public trust:** The public can be assured that there is a human accountable for an AI solution.
- **Improves governance:** Clear accountability structures promote better governance of AI systems to ensure they are developed and used responsibly.

4.7.4 Illustrative Example

They must consider the full societal impact of the AI system, such as the environmental impact or anything that could negatively affect people's physical or mental wellbeing. The Department must try and mitigate these concerns as much as possible. They must evaluate if the benefits to society of the AI system outweigh potential negative consequences and document the rationale applied.

Chapter 5: Decision Framework when working with AI

This chapter introduces a Decision Framework which can be used as a guide for public service workers when considering using AI to solve a problem or improve a service. This framework will help evaluate if AI is the most suitable solution for our needs. The Decision Framework contains several key components.

5.1 Is AI the best solution?

As per the National AI Strategy, we want Ireland's public service to become a showcase of AI adoption [19]. However, that does not mean that AI is always the best solution. We should start by considering the problem we are trying to solve and what the end-user's needs are. Then we can evaluate what approaches are available to us.

When considering if AI is the best solution, we should work with a cross-functional team of experts (including specialists who have good knowledge of the data and domain experts who know the environment where the model will be deployed) and consider factors such as the following:

- What alternative solutions are available to us to solve this problem and what are the advantages and disadvantages of each solution? We must also consider how the anticipated cost of each solution compares to our budget. For example, can easier methods be used that can generate the same quality results in less time or at a lesser expense?
- What data do we have, and will this data be accurate, representative and complete enough to be used? Do we have a large enough dataset to train an AI model?
- Can we use this data responsibly and are we allowed to use it under the guidelines of the GDPR (2018)?
- Will the benefits of the AI system outweigh any prospective negative outcomes?
- Will it equally benefit all users or just disproportionately help some, at the cost to others?
- Do we have the required skillsets at our disposal to be able to deliver the AI solution?
- Will it solve the problem? What metrics are important to assess this hypothesis and how will we measure them?

5.2 What type of AI solution is the best fit?

To help determine what type of AI solution is the best fit, our cross-functional team must ask themselves questions such as the following:

- What way will the end-user be looking to use or interact with the AI system?
- Do we need to be able to achieve the same outcome every time we run the model?
- How accurate do we need to be?
- To what length will we need to explain the model to the end-user?
- How long does the model need to generate results?
- Is the training data reflective of the real-world situations that it will address?
- What risk category would this fall under for the AI Act (2024)?

5.3 Using 'free-of-charge' or enterprise versus licensed AI offerings

When adopting AI, a well-planned approach will help to balance the risks and benefits of AI adoption. Public Service Bodies should evaluate whether to build, buy, or avail of 'free-of-charge' solutions. Factors such as implementation, speed, and compatibility need to be considered.

We want to deliver the best possible solutions, whilst also meeting our obligations to deliver value for the public, delivering benefits, and measuring impact. This includes post-project reviews to analyse if the intended benefits were achieved, the impact of those benefits, and the lessons learned.

5.3.1 Adopting GenAI in the Irish Public Service

It is advised against incorporating GenAI into business processes unless based on an approved business case in accordance with the principles and practices set out in this document. It is also recommended that more general access to such tools by staff should not be permitted until Departments have conducted the relevant business assessments, have appropriate usage policies in place and have implemented staff awareness programmes on safe and appropriate usage of these tools.

"The NCSC recommends that access is restricted by default to GenAI tools and platforms and allowed only as an exception based on an appropriate approved business case and needs. It is also recommended that its use by any staff should not be permitted until such time as Departments have conducted the relevant risk assessments, have appropriate usage policies in place and staff awareness on safe usage has been implemented" [20].

5.3.2 Open-source versus Enterprise

When determining whether to buy or build an AI solution, an evaluation of both options will likely be needed to determine the best approach. As part of this evaluation, we can consider factors such as the following:

- What are the advantages and disadvantages of both solutions? Which solution will perform better or lead to a better solution for the end-user?
- What are the costs and return on investment of building vs. buying over the full AI lifecycle, for example will one cost more during the development phase? Will the bought solution have an ongoing license fee?
- How quickly do we need the AI system to be implemented?
- Could the built or bought solution be used for other use cases in the public service?
- Do we have the required skillsets at our disposal to build a solution or indeed run a bought solution? This includes implementing the solution, operating it and maintaining it after deployment.
- Does one option offer better data security?
- Does one option offer better compatibility with existing systems?
- What training and support would be needed for both options?

5.3.3. Value for Money and Evaluating Impact – Key Considerations

When compared with previous generations of Information Technology, AI can be significantly more costly to deploy and manage, and requires considerably more computing power, data and organisational capability. The associated costs of building, deploying, and supporting AI solutions can increase significantly once rolled out. From an expenditure perspective, AI is primarily reliant on cloud computing technology, which can change the spending and risk profile of the organisation, which in turn can affect business continuity, and increase critical infrastructure dependency. This underscores the importance of value for money analysis from the start, the adoption of regular cost analysis and monitoring across the lifecycle of adoption, and agile evaluation processes across the AI deployment lifecycle. This approach ensures that decisions to adopt AI are informed, that the value of the deployment is maximised, and that the AI deployment is sustainable and beneficial over time.

There are several considerations to AI cost analysis, which can be made at the start, and included in any business case. These continue to be relevant across the life cycle of deployment and can contribute to the demonstration of value for money in terms of efficiency. These considerations include:

- Infrastructure costs: the cost of hardware, cloud computing resources, and storage needed to train and run AI models.
- Data costs: Acquiring, preparing, and labelling data for AI models can be a significant expense, especially for complex projects requiring large datasets.
- Development costs: This covers the salaries of AI engineers, data scientists, and other specialists involved in building and refining AI models. It also includes the cost of software tools and licenses.
- Operational costs: These are ongoing expenses associated with running and maintaining AI systems, such as energy consumption, monitoring, and updates.
- Unexpected costs: AI projects can encounter unforeseen challenges, such as model retraining or debugging, which can add to the overall cost.

As with any business case, the starting option should be 'do nothing'. This option, as well as the timing of adopting AI, needs to carefully considered given that AI solutions, as with any new or emerging technology, tend to be expensive for early adopters, with price declining over time as new providers enter market.

Evaluation of AI use in public services (including process, impact and value for money questions) is also necessary. We need to understand the impact of AI systems compared to the status quo, improve current interventions, inform future adoption, deployment and development, and ensure accountability for public spending. While the key principles of robust impact evaluation are no different for AI adoption or deployment than for any other type of Government programme or initiative, AI interventions can present additional opportunities and challenges for evaluation. These include the iterative and evolving nature of deploying AI. In this respect, emerging best practice in evaluating the impact of AI in public services encompasses: [20]

- Considering the evaluation as early as possible in the process of deploying AI and be clear on the purpose of the deployment.
- Develop a full understanding of the relationship between the proposed inputs and outputs, and the intended outcomes of the AI deployment (commonly referred to as a Logic Model and Theory of Change).
- Document and log all the steps planned and undertaken in the development and deployment of the AI solution, and note any difference between what is planned and what takes place.
- Be prepared to adapt and adopt further appropriate evaluation methods to reflect the evolving nature of the AI deployment over time.
- Consider and document any differences in outcomes and impact for different groups, when planning the evaluation and during the course of the deployment where this becomes known.
- Think early on about how to establish a clearly defined baseline to support the evaluation, considering what data already exists and what may need to be collected.

5.4 Inclusion and diversity in AI from the start

When working in an AI project, it is vital to incorporate diverse perspectives. The solution must be inclusive by design. Working in a multi-background team of different genders, races, cultural backgrounds, disabilities, ages, socio-economic statuses, education can assist in the promotion of inclusive and equitable systems. This diversity of thought brings richer perspectives, mitigates biases, and designs systems to benefit all segments of the Irish Public. This is a necessary approach to designing and building effective AI solutions.

5.5 Next steps after the Decision Framework

After completing the Decision Framework and deciding to proceed with a given AI solution, the Responsible AI Canvas should be completed as a collaborative planning exercise.

Chapter 6:
The Responsible AI Canvas

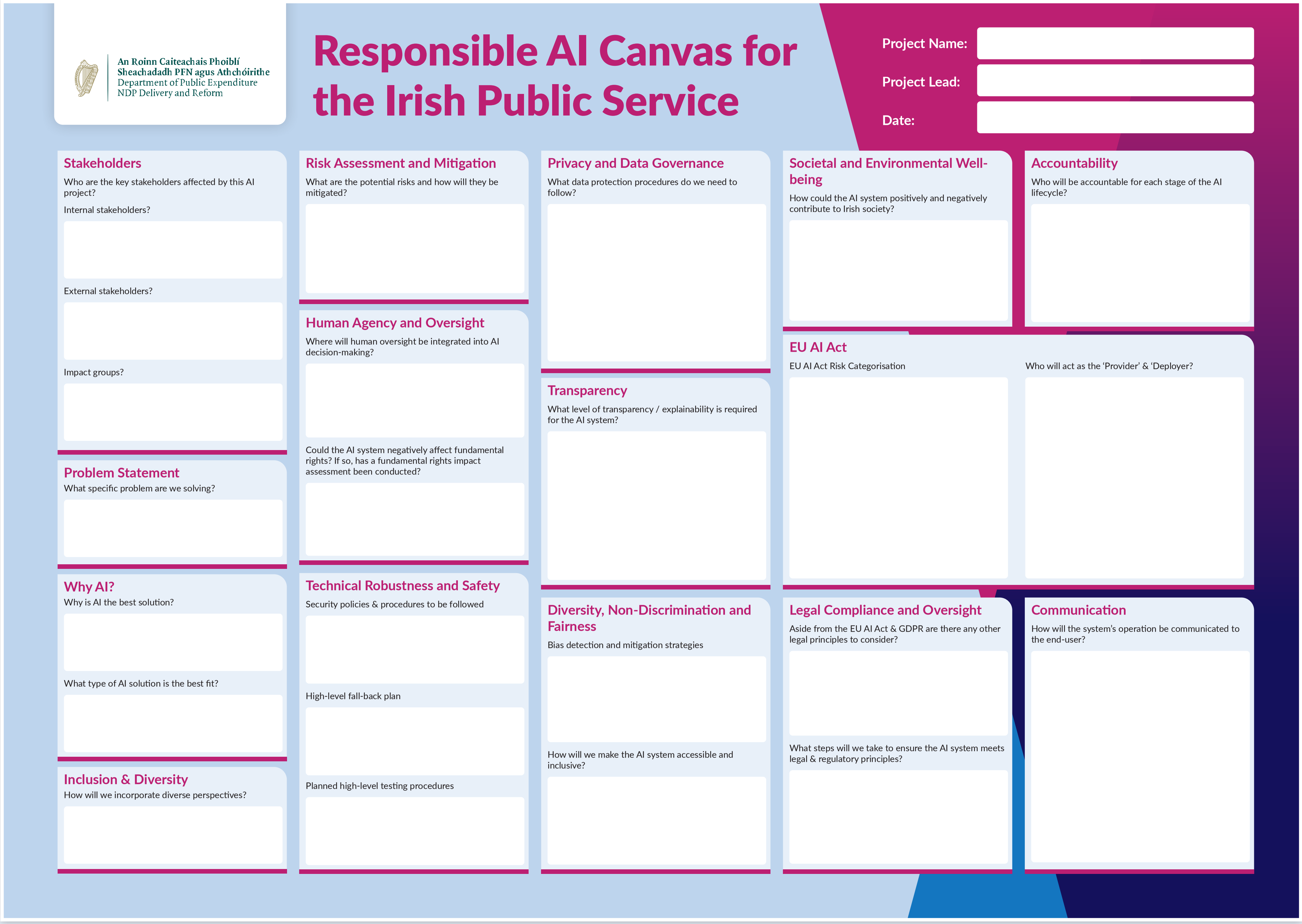
6.1 The Responsible AI Canvas Overview

The Responsible AI Canvas is a simple, structured tool, designed to help public service workers develop, implement and oversee responsible AI solutions that meet the seven Principles for Responsible AI. It is recommended that this tool is used at the planning stage of an AI project.

The canvas should be completed by a cross-functional team of experts including but not limited to, service providers, technical teams, product owners, legal teams and external partners. This process should align with the Design principles for Government in Ireland.

The canvas aims to facilitate structured conversations on implementing the Seven Principles. The Responsible AI Canvas will pose some key questions in line with the Guidelines. It will guide public service workers through certain elements such as helping us identify stakeholders and defining the problem statement. It will also help initiate conversations about compliance principles for GDPR (2018) and the EU AI Act (2024). The canvas encourages proactive risk management to assess and mitigate potential risks from the outset.

The Responsible AI Canvas can be accessed at gov.ie/transformation.



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

Responsible AI Canvas for the Irish Public Service

Project Name:

Project Lead:

Date:

Stakeholders

Who are the key stakeholders affected by this AI project?

Internal stakeholders?

External stakeholders?

Impact groups?

Problem Statement

What specific problem are we solving?

Why AI?

Why is AI the best solution?

What type of AI solution is the best fit?

Inclusion & Diversity

How will we incorporate diverse perspectives?

Risk Assessment and Mitigation

What are the potential risks and how will they be mitigated?

Human Agency and Oversight

Where will human oversight be integrated into AI decision-making?

Could the AI system negatively affect fundamental rights? If so, has a fundamental rights impact assessment been conducted?

Technical Robustness and Safety

Security policies & procedures to be followed

High-level fall-back plan

Planned high-level testing procedures

Privacy and Data Governance

What data protection procedures do we need to follow?

Transparency

What level of transparency / explainability is required for the AI system?

Diversity, Non-Discrimination and Fairness

Bias detection and mitigation strategies

How will we make the AI system accessible and inclusive?

Societal and Environmental Well-being

How could the AI system positively and negatively contribute to Irish society?

Accountability

Who will be accountable for each stage of the AI lifecycle?

EU AI Act

EU AI Act Risk Categorisation

Who will act as the 'Provider' & 'Deployer'?

Legal Compliance and Oversight

Aside from the EU AI Act & GDPR are there any other legal principles to consider?

What steps will we take to ensure the AI system meets legal & regulatory principles?

Communication

How will the system's operation be communicated to the end-user?

6.2 How to use the Responsible AI Canvas

The Responsible AI Canvas is used during the initial stages of any AI project. This could be part of a planning or a design thinking workshop to facilitate teams working through responsible AI principles. Below are the key steps on how to use the canvas effectively:

1. **Assemble a cross-functional team:** This could include, but is not limited to, technical teams, product owners and legal teams. This may also feature relevant external stakeholders. This provides a broad range of perspectives and potential risks, and implementation challenges are identified at the outset.
2. **Work through each section of the canvas:** Each question should be discussed but the canvas should not be viewed as a complete list. Teams are actively encouraged to add additional questions and adapt the canvas to fit the specific use case.
3. **Risk management:** The canvas encourages proactive risk management to assess and mitigate potential risks from the outset. However, plans should be initiated to allow for ongoing risk assessments to be carried out throughout the full AI lifecycle.
4. **Communication strategy:** It is worth considering, from the beginning, the communication strategy on how the AI system will be used, as this may have an influence on the solution itself.
5. **Ongoing monitoring:** The canvas is designed to be used during the planning stage of the AI lifecycle. However, teams need to establish governance mechanisms for the full AI lifecycle.

Scope and limitations of the Responsible AI Canvas

'The Responsible AI Canvas is a valuable tool to align AI projects to these guidelines. However, using the canvas does not guarantee responsible AI solutions or regulatory compliance. The Responsible AI Canvas will pose some key questions to consider in line with these Guidelines. The accountable stakeholders will have to ensure that the end-to-end AI solution is lawful and was developed, deployed, and maintained responsibly.'

6.3 Next steps after the Responsible AI Canvas

After completing the Responsible AI Canvas, Chapter 7 provides information on responsible actions that should be taken throughout the AI lifecycle.

Chapter 7:
AI lifecycle for responsible adoption

7.1 AI lifecycle Introduction

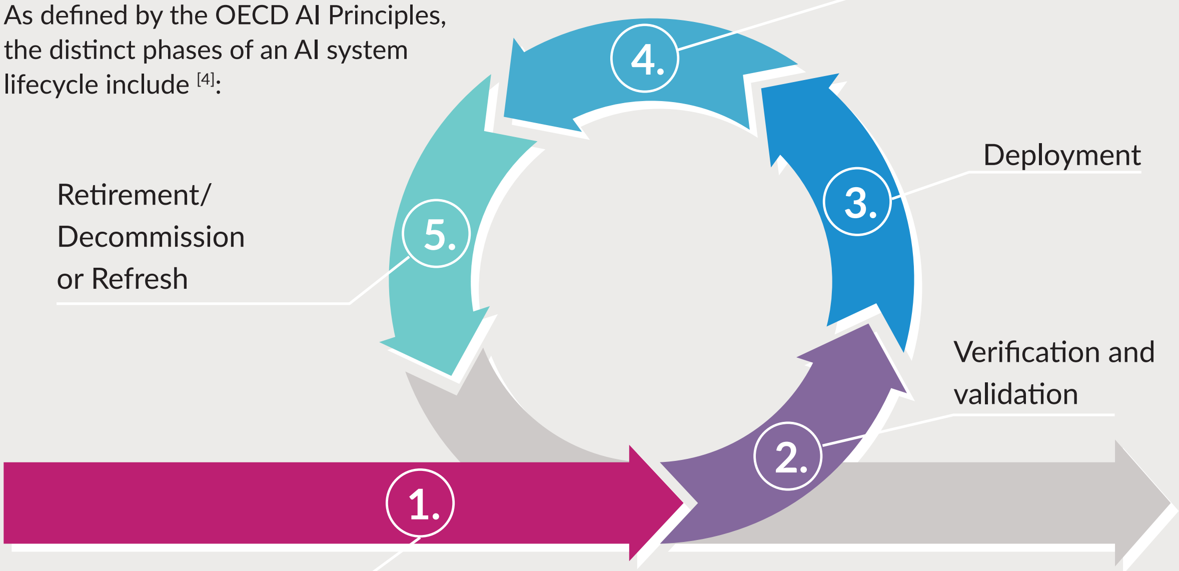
As we have outlined, our commitment to responsible AI extends beyond legal obligations and compliance. While adhering to the EU AI Act and other regulations provides a crucial foundation, we believe that integrating the seven principles for the responsible use of AI at each stage of the lifecycle is fundamental to achieving responsible AI.

This chapter offers practical guidance to public service workers on essential actions and considerations to follow during each stage of the AI lifecycle. It uses the phases of an AI lifecycle as defined by the OECD [4] and aligns each step with these guidelines.

By considering these best practices, alongside regulatory principles, we can ensure our approach to AI aligns to a high standard for responsible and transparent AI in the Irish Public Service, for the benefit of the public.

The AI Lifecycle

As defined by the OECD AI Principles, the distinct phases of an AI system lifecycle include [4]:



Operation and Monitoring

Deployment

Retirement/
Decommission
or Refresh

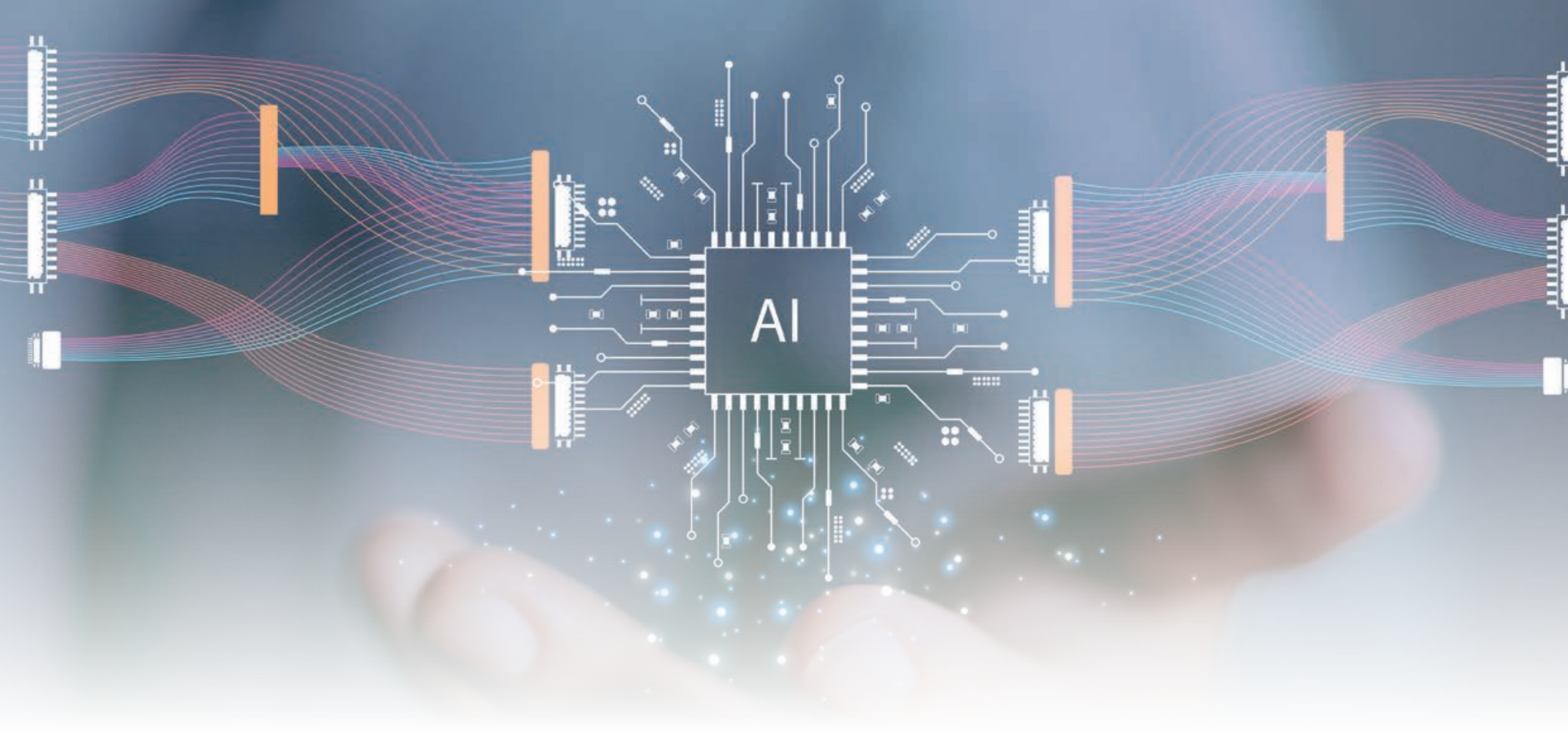
Verification and
validation

Design, Data & Models

i. Planning & Design

ii. Data Collection & Processing

iii. Model Building



According to the OECD, the phases of the AI lifecycle "often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase." [4]. A new use case of the process may lead to a refresh of the cycle.

This iterative and agile nature is essential to consider, as AI projects can evolve significantly over time, potentially altering which best practices are relevant and necessary for each stage. By taking a flexible approach and adapting these practices to each phase, public service teams can address the unique challenges of AI development.

7.2 Implementing the seven principes across an AI lifecycle

In this section, you will find a structured approach and suggested actions that public service teams can take to apply the seven principles at each stage of the AI lifecycle.

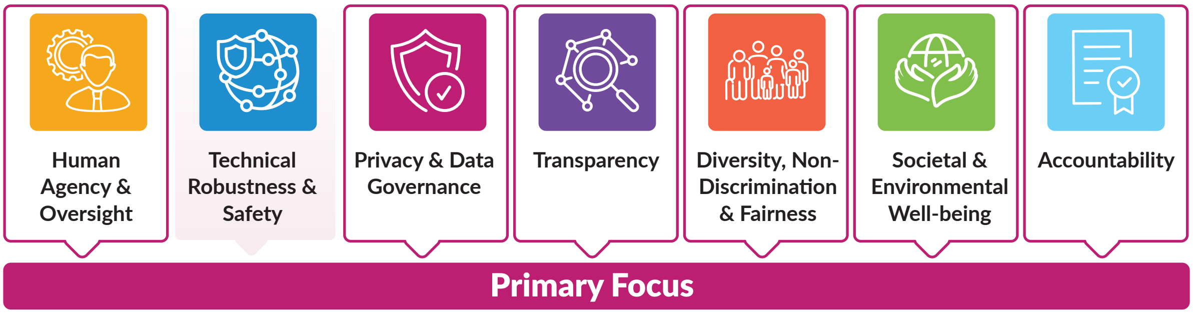
It is important to note that these practices are not exhaustive or mandatory. They should be applied as appropriate for the AI system's context and risk level. Different use cases present different risks with some requiring a higher standard of assurance than others. Therefore, not all AI use cases will require the application of all available practices to be considered safe and responsible.

Public service teams are encouraged to add additional actions at each stage of the lifecycle based on each specific context. This will help to ensure that the right safeguards are in place to address the unique demands and risks associated with AI system delivery.

The use of these tables are not in place of Corporate Governance ensuring compliance with existing data and digital regulation, but will aid in asking the relevant questions. However, these actions, when combined with existing Corporate Governance, strengthen our commitment to responsible AI for better public services.

At each stage of the AI lifecycle, the HLEG requires that heightened attention should be given to areas of "primary focus" due to their relevance and impact during those phases. This emphasis is intended to guide practitioners in prioritising their efforts. However, it does not reduce the importance of the other HLEG principles, which remain essential considerations across all stages of the lifecycle [2]. The next section provides a guide on the relevant primary focuses through the AI lifecycle.

7.2.1 Planning & Design



Human Agency & Oversight

Technical Robustness & Safety

Privacy & Data Governance

Transparency

Diversity, Non-Discrimination & Fairness

Societal & Environmental Well-being

Accountability

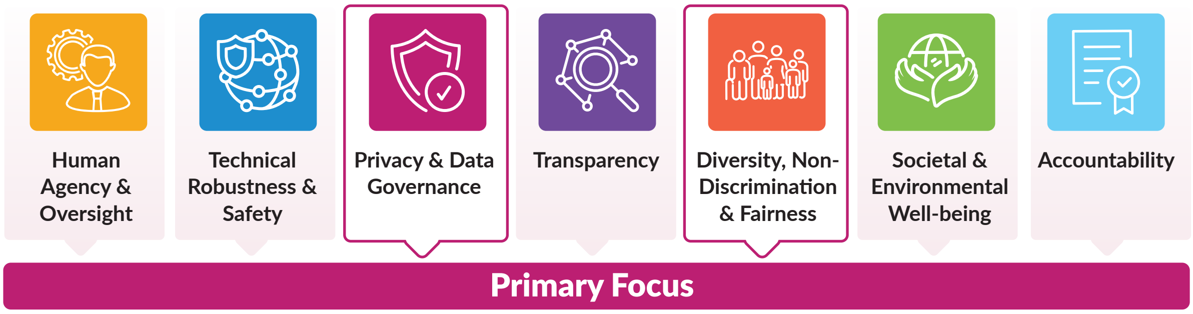
Primary Focus

This first phase sets up the baseline of the AI system and project. The ethical considerations and alignment with human values are fundamental at this stage, as good design sets out good practice. This phase involves defining the objectives, goals and evaluating if AI is a suitable approach, by using the Decision Framework for AI. The key stakeholders, accountability, risks, assumptions, constraints, and project plan are all decided in this phase.

Principle	Planning and Design Actions to take
1. Human Agency & Oversight	○ Define where human oversight and control will be integrated, especially for decision-making (i.e. reviewing the process in which AI will be embedded) ○ Establish clear guidelines on human intervention points to ensure output does not override human decision-making. ○ Set up role-specific responsibilities for oversight, from the beginning to allow for intervention and correction, as necessary (i.e. clear RACI matrix). ○ Identify and consult with stakeholders, including subject matter and legal experts and impacted groups and their representatives. ○ An assessment should be made if the AI system could produce legal effects on users or similarly significantly affect them. In these cases, users have a right not to be subjected to a decision based solely on automated processing so we must account for this [2]. ○ Utilise the Decision Framework when working with AI.
2. Technical Robustness & Safety	○ Ensure the prospective AI system is in line with data security policies and is robust and secure. ○ Conduct pre-deployment risk assessments [2]. ○ Establish a fallback plan in case something goes wrong with the AI system[2]. This might involve switching to a simpler system or allowing human controllers to step in. However, we should also consider factors such as incident response planning, data recovery planning, rollback planning and cybersecurity incident planning. ○ Engage with the IT security team on security procedures to be followed throughout the AI lifecycle.

Principle	Planning and Design Actions to take
3. Privacy & Data Governance	○ Ensure processes around privacy and data protection are set up and are being followed. Conduct a Data Protection Impact Assessment (DPIA) if required. ○ Review data access procedures to prevent unauthorised access[2]. ○ Integrate data minimisation and security protocols into the design and ensure that sensitive information is encrypted. ○ Plan for data pseudonymisation and anonymisation when possible. ○ Identify Personally Identifiable Information (PII) data fields that the AI system should not use.
4. Transparency	○ Establish what level of transparency / explainability is required for the AI system to plan accordingly. ○ If the AI system could affect human rights, an alternative process should be defined where users can decide to avail of a human interaction [2].
5. Diversity, Non-Discrimination & Fairness	○ Where applicable, establish a plan for engaging with stakeholders who may directly or indirectly be affected by the system throughout its lifecycle [2]. ○ A fundamental rights impact assessment should be undertaken if the AI system could negatively affect human rights. ○ An evaluation should be done on how to make the AI system as accessible as possible, especially for vulnerable groups. Universal design principles should be adopted [2]. ○ Where possible, the team being assembled to work on the AI system should be diverse, to allow for diverse opinions [2]. ○ Assess if the right people are in the room to be discussing the impacts of the AI approach. Diversity of mind is crucial from the beginning.
6. Societal & Environmental Well-Being	○ Conduct social impact assessments to ensure AI systems contribute positively to Irish society [2].
7. Accountability	○ Establish accountability frameworks early in design for each stage of the AI lifecycle, covering key risk areas such as, design risk, data risk, algorithmic risk, performance risk, technology risk, third-party risk, conduct/compliance/legal risk and business process risk [2]. ○ Use impact assessments (e.g. red teaming or forms of Algorithmic Impact Assessment) to try and minimise the potential negative impact of the AI system [2]. ○ Assess future scenarios: ○ When something deviates from the intended output or behaviour, who is responsible for noticing and correcting this? ○ Is someone responsible for making sure that every step is not just done, but done correctly?

7.2.2 Data Collection & Processing

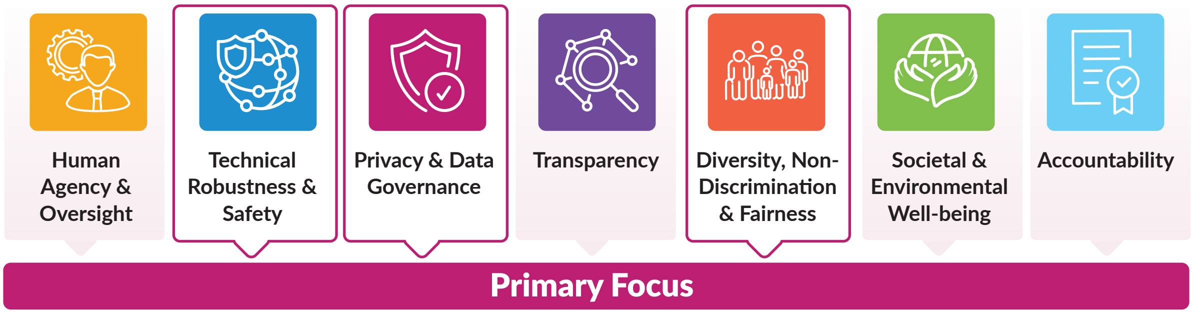

Human Agency & Oversight
Technical Robustness & Safety
Privacy & Data Governance
Transparency
Diversity, Non-Discrimination & Fairness
Societal & Environmental Well-being
Accountability
Primary Focus

As data is the core ingredient for AI systems, the emphasis of responsibility is focused on collecting, managing, and protecting data correctly. Privacy and data protection are paramount. Ensure the data collected for the problem is representative and has diversity. In this stage of the lifecycle, the required data is collected, cleaned, and prepared for model building. It is crucial that the data is of high quality, unbiased and respects privacy.

Principle	Data Collection and Processing Actions to take
1. Human Agency & Oversight	○ Validate data processing to allow for human intervention and oversight during critical decision points. ○ Ensure monitoring is in place for unauthorised entry.
2. Technical Robustness & Safety	○ Ensure data used during this phase is kept secure and procedures are followed in line with security policies. ○ The data collected should be accurate, representative, authenticated, and reliable and suitable for the specific use case. ○ Regularly monitor and audit data sources and data collection processes to ensure data security and integrity.
3. Privacy & Data Governance	○ Ensure processes around privacy and data protection are established and being followed. ○ Limit the collection of data to comply with GDPR with only necessary fields and avoid using Personally Identifiable Information (PII) data unless strictly required. ○ Ensure data collected is not being used to unlawfully or unfairly discriminate against individuals [2]. ○ Address data quality issues as this is paramount to the performance of the AI system. Bias, inaccuracies, and errors must be addressed before the model is trained [2]. ○ Ensure that the data used for development meets the data quality standards defined by your Department or agency. ○ Document the data preparation activities (preprocessing or transformations performed on a dataset prior to training or development). In practice, this may include but is not limited to, featuring engineering, normalisation, or labelling target variables.

Principle	Data Collection and Processing Actions to take
4. Transparency	○ Document to the best possible standard the datasets and the processes that yield the AI system's decision, including those of data gathering, data transformation and data labelling [2]. ○ These should include the record of data origination, intended use and data retention policies.
5. Diversity, Non-Discrimination & Fairness	○ Ensure the dataset is representative and any discriminatory bias is removed or mitigated. Any steps taken should be documented accordingly. ○ Where applicable, engage with stakeholders who may directly or indirectly be affected by the system as appropriate for this stage of the AI lifecycle [2].
6. Societal & Environmental Well-Being	○ Ensure that data collection and processing is done responsibly avoiding negative social impacts.
7. Accountability	○ Ensure the accountability framework established is being followed and audit trails for data collection and processing decisions are being maintained. ○ Continue to use impact assessments (e.g. red teaming or forms of Algorithmic Impact Assessment) to try and minimise the potential negative impact of the AI system [2].

7.2.3 Model Building

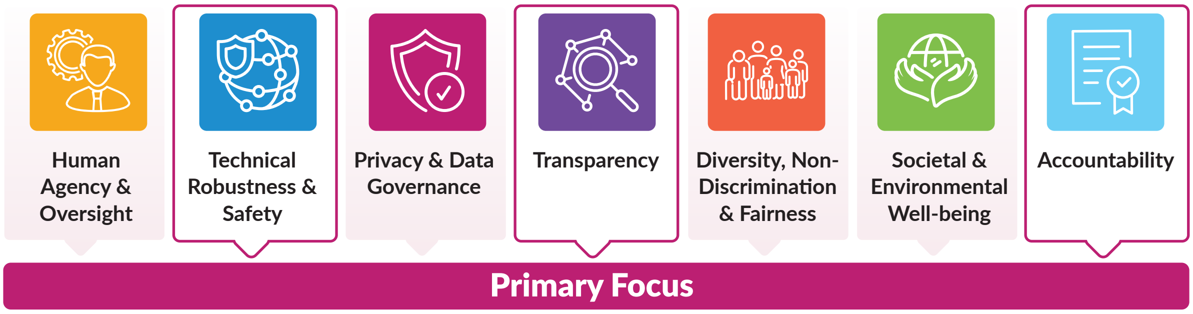

Human Agency & Oversight
Technical Robustness & Safety
Privacy & Data Governance
Transparency
Diversity, Non-Discrimination & Fairness
Societal & Environmental Well-being
Accountability
Primary Focus

This stage of the lifecycle involves training the model on the data prepared. During this training, technical robustness is key so that the model functions reliably. Fairness testing must be carried out during the model development, to avoid discriminatory results or the spread of data errors. Privacy needs to be maintained by following the data-handling practices for use of training. Typically, the training process is iterative and may involve different testing approaches to find the most effective model. These different approaches may require additional measures.

Principle	Model Building Actions to take
1. Human Agency & Oversight	○ Build models while considering their ability to include human involvement to assess the model outputs, even during training and tuning.
2. Technical Robustness & Safety	○ Ensure the model is kept secure and procedures are followed in line with security policies. ○ Design the models to handle diverse conditions, ensuring stability across scenarios. ○ Develop the model in a secure environment with the appropriate access levels.
3. Privacy & Data Governance	○ Ensure processes around privacy and data protection are established and being followed.
4. Transparency	○ Ensure the model can be explained to the level defined in the Planning & Design stage. This could include explaining the technical processes of an AI system and the related human decisions (e.g. application areas of a system) [2]. ○ When explainability is limited, the benefits of the AI system need to be weighed against the explainability limitations. Reasons for progressing need to be documented and human oversight controls increased.

Principle	Model Building Actions to take
5. Diversity, Non-Discrimination & Fairness	○ Continuously test for bias in model outputs during development. Put oversight processes in place to analyse and address the system's purpose, constraints, principles, and decisions in a clear and transparent manner [2].
6. Societal & Environmental Well-Being	○ Consider the environmental impact of AI models and if any mitigation or better choices can be made. ○ Assess how the model could negatively impact people's physical and mental wellbeing. ○ Assess the potential societal impact of the model.
7. Accountability	○ Ensure the accountability framework established is being followed and audit trails for processing decisions and model build are being maintained. ○ Continue to use impact assessments (e.g. red teaming or forms of Algorithmic Impact Assessment) to try and minimise the potential negative impact of the AI system [2]. ○ Ensure there is an ability to report on actions or decisions that contribute to a certain system outcome. Ensure there is also an ability to respond to the consequences of such an outcome [2].

7.2.4 Verification and Validation

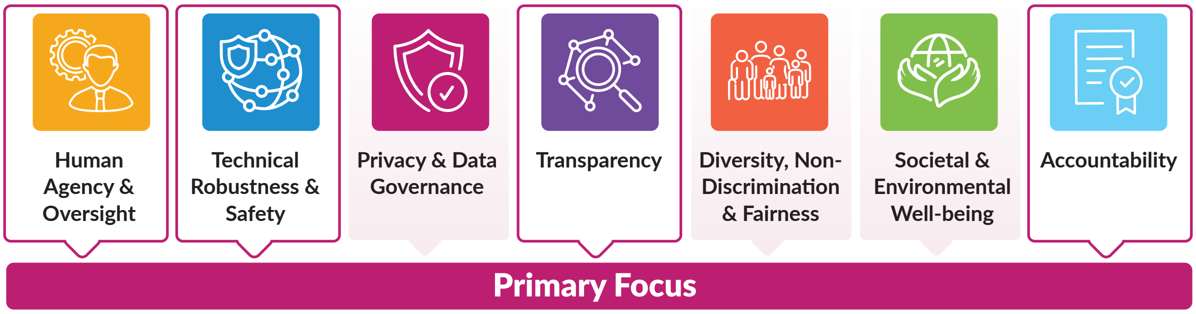

Human Agency & Oversight
Technical Robustness & Safety
Privacy & Data Governance
Transparency
Diversity, Non-Discrimination & Fairness
Societal & Environmental Well-being
Accountability
Primary Focus

The Verification and Validation stage ensures that the model is technically sound, safe, and designed for the intentions and purposes set out in the Planning & Design phase. The team verifies that outputs from the AI system are accurate, and that the system performs consistently without unintended consequences. The model should be tested in environments that resemble real world conditions to confirm that it operates responsibly and as per the design. The ability to understand the model decisions are critical here (for high-risk models which heighten the focus on transparency). This phase can revert back to the model building phase for amendments in the modelling.

Principle	Verification and Validation Actions to take
1. Human Agency & Oversight	○ Models should be tested for their ability to include human involvement and control over decision-making.
2. Technical Robustness & Safety	○ Consideration should be given to possible unintended applications of the AI system and potential abuse of the system by malicious actors. Steps should be taken to prevent and mitigate these. ○ Assess if the system will do what it is supposed to do without harming living beings or the environment. ○ If it is accepted that occasional inaccurate predictions cannot be avoided, it is important that the system can indicate how likely these errors are to occur. ○ Testing should be done in small-scale pilot environments to identify and mitigate problems. ○ The system should be tested for reliability and to be reproducible, where applicable. ○ An explicit and well-formed development and evaluation process should be conducted. This can help in mitigating and correcting unintended risks from inaccurate predictions. ○ Ensure the procedures directed by the IT security team during the ‘Planning & Design’ stage have been followed.

Principle	Verification and Validation Actions to take
3. Privacy & Data Governance	○ Ensure processes around privacy and data protection are established and being followed.
4. Transparency	○ Test if the model outputs can be explained where required. ○ Effective model documentation should include details on how the model was tested for both performance and relevant risks, enabling downstream stakeholders to assess the relevance and durability in new contexts.
5. Diversity, Non-Discrimination & Fairness	○ Continuously test for bias in model outputs. ○ Where applicable, engage with stakeholders who may directly or indirectly be affected by the system as appropriate for this stage of the AI lifecycle [2]. ○ The AI system should be tested to be as accessible as possible. Ensure Universal Design principles are adopted [2].
6. Societal & Environmental Well-Being	○ Assess how the model could negatively impact people's physical and mental well-being. ○ Assess the potential societal impact of the model.
7. Accountability	○ Ensure the accountability framework established is being followed and audit trails for testing and validation are being maintained. ○ Continue to use impact assessments (e.g. red teaming or forms of Algorithmic Impact Assessment) to try and minimise the potential negative impact of the AI system [2]. ○ Ensure there is an ability to report on actions or decisions that contribute to a certain system outcome. Ensure there is also an ability to respond to the consequences of such an outcome [2].

7.2.5 Deployment

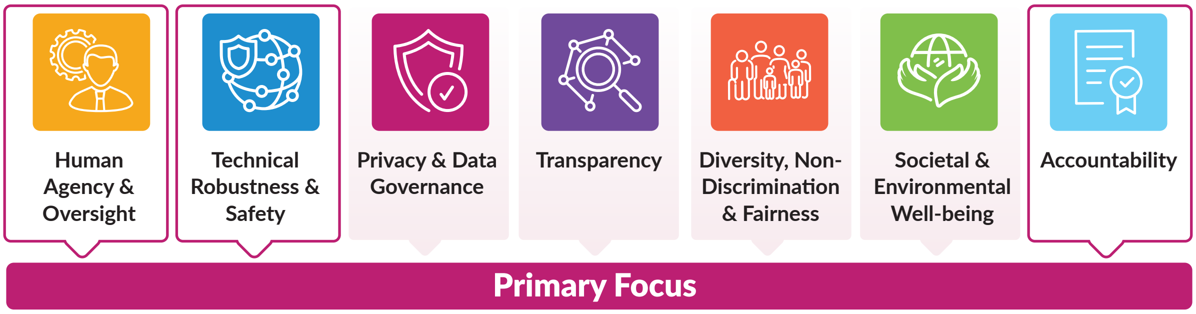

Human Agency & Oversight
Technical Robustness & Safety
Privacy & Data Governance
Transparency
Diversity, Non-Discrimination & Fairness
Societal & Environmental Well-being
Accountability
Primary Focus

Deployment involves an AI system being introduced into its intended operational and production environment, where the relevant end-users have accessibility to the outputs. During this process, the system might be integrated with existing technology or data domains. This involves additional considerations like user access, security controls and change management initiatives in preparation. Clear documentation, communication and change management can determine the success and/or failure of the model entering an operational environment. This point emphasises the human oversight, transparency, and accountability of the solution.

Principle	Deployment Actions to take
1. Human Agency & Oversight	○ The appropriate level of human oversight established in the Planning & Design phase should be implemented. ○ Mechanisms should be put into place to receive external feedback regarding AI systems that potentially infringe on fundamental rights [2]. ○ Users should be supplied with the required information to make informed autonomous decisions regarding AI systems. ○ As per the Planning & Design stage, if the AI system could produce legal effects on users or similarly significantly affect them, an additional process should be made available to them. This will ensure that they have the right not to be subject to a decision based solely on automated processing [2]. ○ Appropriate training and support are provided to the end-users to explain the AI System, outcomes, and user guidance.
2. Technical Robustness & Safety	○ Monitor real-world AI performance to ensure system robustness. ○ Monitor for data contamination risks, concept drift and data drift. ○ Ensure a fallback plan can be activated in case something goes wrong with the AI system [2]. ○ Documentation on the deployment processes and subcomponents. ○ A review of the model documentation (i.e. if such meta-data about the process is not documented throughout the lifecycle, it can be difficult to refactor or reconstruct later).

Principle	Deployment Actions to take
3. Privacy & Data Governance	- Ensure processes around privacy and data protection are established and being followed. - Ensure the AI system is securely deployed and only the required users can access the system.
4. Transparency	- Decisions made by the AI system should be documented to the best possible standard to allow for traceability and an increase in transparency [2]. - Explanations of the degree to which an AI system influences and shapes the organisational decision-making process, design choices of the system and the rationale for deploying it, should be available [2]. - Clearly inform humans when they are interacting with an AI system or that some outputs have been developed with the aid of an AI system. - The AI system's capabilities and limitations should be communicated to AI practitioners or end-users in a manner appropriate to the use case at hand. - If the AI system could impact human rights, an alternative process should be made available where users can decide to avail of a human interaction [2].
5. Diversity, Non-Discrimination & Fairness	- Monitor AI decision-making during deployment, to ensure fairness and equity. - Where applicable, engage with stakeholders who may directly or indirectly be affected by the system as appropriate for this stage of the AI lifecycle [2].
6. Societal & Environmental Well-Being	- Before deployment, assess if the AI system contributes positively to society, outweighing any potential negative effects it may have.
7. Accountability	- Ensure the accountability framework established is being followed. This means having clear accountability for oversight during deployment, ensuring AI decisions can be traced back to humans. - Continue to use impact assessments (e.g. red teaming or forms of Algorithmic Impact Assessment) to try and minimise the potential negative impact of the AI system [2]. - Ensure that there is an ability to report on actions or decisions that contribute to a certain system outcome. Ensure there is also an ability to respond to the consequences of such an outcome [2]. - Ensure a user guide has been created to tell users how to use the AI system once it is deployed. This should be transparent of the model's strengths and weaknesses. The user guides must have the ethical considerations outlined clearly, with a specified email or person nominated as the contact point for ethical issues.

7.2.6 Operation & Monitoring



Human Agency & Oversight

Technical Robustness & Safety

Privacy & Data Governance

Transparency

Diversity, Non-Discrimination & Fairness

Societal & Environmental Well-being

Accountability

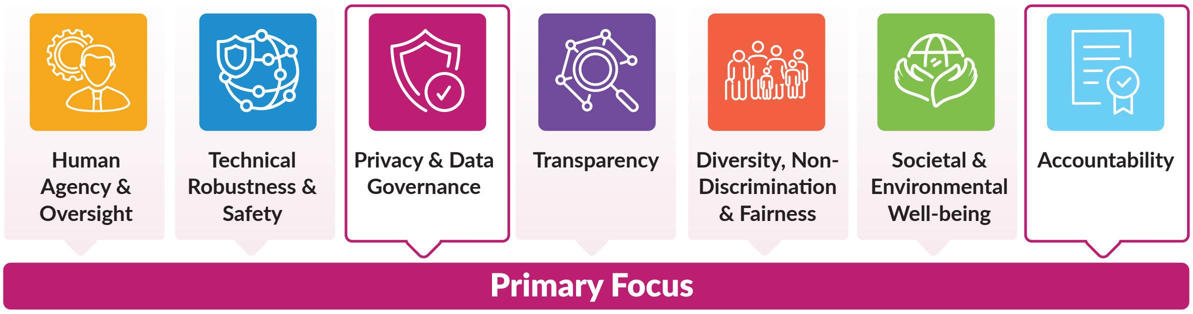
Primary Focus

Once deployed, the AI system should meet its intended purpose. This requires continuous monitoring and oversight with the right controls in place to ensure it continues to meet standards and objectives. Ongoing monitoring can help detect model degradation, errors, or issues. Consistent engagement with end-users and accountable owners should be a continued responsibility. AI is not a 'set and forget' technology. This continued engagement can enhance the model for future iterations.

Principle	Operation and Monitoring Actions to take
1. Human Agency & Oversight	○ Set up continuous monitoring systems ensuring AI systems remain under human control. ○ Ongoing feedback received on the AI system should be reviewed, especially regarding AI systems that potentially infringe on fundamental rights.
2. Technical Robustness & Safety	○ Monitor real-world AI performance to ensure system robustness. ○ Monitor for data contamination risks, concept drift and data drift. ○ Assess if the system will do what it is supposed to do without harming living beings or the environment.
3. Privacy & Data Governance	○ Ensure processes around privacy and data protection are established and being followed.
4. Transparency	○ Monitor feedback to see if users are struggling to comprehend that they are interacting with an AI system or that some outputs have been developed with the aid of an AI system. Communication should be adapted as appropriate.

Principle	Operation and Monitoring Actions to take
5. Diversity, Non-Discrimination & Fairness	○ Conduct post-deployment testing to ensure the system continues to produce fair and non-discriminatory results. ○ Where applicable, engage with stakeholders who may directly or indirectly be affected by the system as appropriate for this stage of the AI lifecycle [2]. ○ Monitor the AI system to ensure it is as accessible as possible.
6. Societal & Environmental Well-Being	○ Monitor the societal impact of the model to gauge ongoing impact.
7. Accountability	○ Ensure the accountability framework established is being followed and proper oversight and monitoring is in place. ○ In applications affecting fundamental rights, including safety-critical applications, AI systems should be able to be independently audited [2].

7.2.7 Retirement/Decommission or Refresh of AI Systems


Human Agency & Oversight
Technical Robustness & Safety
Privacy & Data Governance
Transparency
Diversity, Non-Discrimination & Fairness
Societal & Environmental Well-being
Accountability
Primary Focus

The retiring or decommissioning of an AI system can occur at any point during the operation and monitoring phase of the lifecycle. When the decision is made to decommission the AI system, secure data handling and deletion practices must be followed. Accountability is important here to ensure that best practices for retirement are carried out.

Principle	Retiring/Decommissioning or Refreshing Actions to take
1. Human Agency & Oversight	- Humans should review the decommissioning process.
2. Technical Robustness & Safety	- Follow data security policies related to decommissioning. - Be prepared to quickly and safely disengage an AI system when an unresolvable problem is identified.
3. Privacy & Data Governance	- Ensure safe handling of personal data post-retirement. Follow GDPR guidance.
4. Transparency	- Rationale for decommissioning should be clearly documented. - Key stakeholders should be informed that the system is being decommissioned and other applicable supports should be highlighted to them where relevant.
5. Diversity, Non-Discrimination & Fairness	- Consider fairness to all stakeholder groups with decisions during decommissioning.
6. Societal & Environmental Well-Being	- Decommissioning should be done in a means that is as environmentally friendly as possible, especially for any physical components.
7. Accountability	- An accountable stakeholder should be appointed for the decommissioning and ensure that all documentation is updated.

These guidelines serve as a flexible framework that public service teams can adapt based on the unique needs and risk levels of each AI project. While not mandatory, these practices offer a valuable roadmap for aligning our AI efforts with responsible principles and fostering trust. These practices will help us manage AI responsibly, keeping public welfare and service excellence at the forefront of our agenda.

Chapter 8: Guidance for End-Users of GenAI

It is advised against incorporating GenAI into business processes unless based on an approved business case in accordance with the principles and practices set out in this document. It is also recommended that more general access to such tools by staff should not be permitted until Departments have conducted the relevant business assessments, have appropriate usage policies in place and have implemented staff awareness programmes on safe and appropriate usage of these tools.

While AI has the capacity to transform public services, it is essential that the data underpinning AI is of high quality, as low-quality data cannot be relied upon for its veracity. In effect, to get accurate results from data, high-quality data is required.

In addition to ensuring that all policy and procedures as well as legislative requirements around data are adhered to, data in this context must also:

1. be the correct data to be relied upon in a given circumstance and;
2. be data that is right, or accurate.

1. Regarding the reliance upon the correct data, considerations should include:

- The completeness and relevance of the data
- That certain data that should be excluded for legal, intellectual property, ethical and regulatory reasons
- The age of the data and how important the age of the data is for the purposes for which it is being used
- Clarity as regards the definitions of data attributes, units of measurement, and sources
- The comprehensiveness of the data and its adequacy in terms of those to whom the data relates
- Freedom from bias within the data, especially historical data

2. Regarding the reliance upon data that is right, or accurate, considerations should include:

- Whether duplication has been avoided
- That the data is labelled correctly
- That the data is accurate and that the data values are correct
- That identifiers for the data are consistent and that there is no variation for data that relates to the same subject or party.

Ensuring that high-quality data is present from the outset and maintained is an important responsibility. Therefore, a structured process and clear accountability system should be adopted to ensure that those with the necessary skills at the highest level on a given project or programme rigorously satisfy themselves on an ongoing basis as to as to the quality of the data. Senior leaders and Management Boards should also require quality assurance, using independent reviews to ensure that any matters or issues arising in relation to data quality are comprehended, acknowledged and acted upon quickly and in an appropriate manner.

General-Purpose AI (GPAI)

"General-purpose AI (GPAI) models power AI systems that are capable of performing a wide range of tasks, such as text generation and image recognition, across different applications" [11]. Within the AI Act, "GPAI models that do not pose systemic risks are subject to limited principles, such as transparency obligations, while those with systemic risks must comply with stricter rules" [11].

Free GenAI Tools

Free GenAI tools are very accessible but because they lack suitable management and oversight pose significant risks for use in the Irish Public Sector. Any information given to a public GenAI tool could be used in training the model. **Thus, we advise against their use in the public service.** The National Cyber Security Centre elaborated on this perspective by saying it is "imperative that data that your organisation does not want in the public domain is never entered into a public GenAI model. This includes classified information, personal data, commercially sensitive data, private Government business etc." [20].

Recommendations for GenAI End-Users

Assuming the National Cyber Security Centre Guidance on Generative AI for Public Sector Bodies [21] has already been followed, listed below are best practice recommendations for public sector workers using GenAI tools which have been made available to them, within their organisation.

Transparency and disclosure with outputs:

- Users must disclose when content is generated by an AI system, particularly in communications, content generation, or interactions with the public or other stakeholders.
- Ensure that it is clear to recipients that the content has been generated by AI to maintain trust and transparency.
- Clearly communicate the limitations of GenAI applications' outputs which may include factual inaccuracies or a lack of context-specific insights. As outlined below, 'validate and verify' is crucial when using AI generated outputs.

Data privacy and sensitivity:

- Avoid inputting sensitive, proprietary, or personal data into generative AI systems to prevent unintended storage, misuse, or leakage of this data.
- Comply with all data protection regulations.

Validate and verify:

- Always validate and verify the output generated by AI systems to ensure its accuracy and reliability before using it in decision-making or communications.

Human agency and oversight:

- AI tools are here to complement not replace human judgement. Users must critically assess all outputs. This ensures the accountability remains with the human.
- Implement human oversight to review and approve AI-generated content, especially in critical or sensitive areas.

Mitigating bias and promoting fairness:

- Users should be aware that AI-generated content can reflect biases present in training data. Review and adjust outputs to ensure they are inclusive, fair, and free of potentially harmful stereotypes.
- Post-processing filters can be applied to the model's responses to detect and modify biased language, ensuring more equitable and fair responses before they reach the end-user.

National Cyber Security Centre

A review to check if the use case is in line with the guidance from National Cyber Security Centre should be conducted. When considering GenAI use cases, the NCSC's Cyber Security Guidance on Generative AI for Public Sector Bodies [20] should be consulted.

Chapter 9:

AI Use Cases in the Public Service

AI presents valuable opportunities for improving the deliv ery of public services in Ireland. By automating tasks, enhancing decision-making, and supporting oversight, AI can make public services more efficient, responsive, and accountable. The OECD carried out extensive research on governing with AI and classified AI use cases in the public sector in four key types of functions:

Internal Operations
Efficiency improvements of internal public sector operations

Service Delivery
Enhancing the delivery of public services by improving responsiveness

Internal and External Oversight
Enhancing oversight and risk detection to facilitate more accountability

Policy-making
Improve the decision-making process

The following sections provide examples from around the world that illustrate how different Governments are successfully applying AI across these four functions.

9.1 Internal Operations - Efficiency improvements of internal public sector operations

According to a G7 report referring to the same OECD research, "About half of the reported AI use cases in the public sector in G7 members are set to increase the efficiency of public sector operation" [22]. This statistic emphasises the readily available power that AI can unlock in increasing internal public service efficiency. By automating repetitive or low value tasks we can free up time for our people to carry out higher value work.

9.1.1 Irish Examples

“The National Transport Authority (NTA) has used an AI Large Language Model (LLM) to assist in the management of questions from Government representatives. Researching and collating the answers to these questions often takes a lot of time and resources, and are answered using different, disconnected sources of information.

The PQ Responder App uses AI to gather and organise important relevant information, making it easier to quickly provide accurate and up-to-date answers. By organising and updating the information regularly, the App ensures that answers are based on the latest information, avoiding outdated or incorrect responses.

The App was designed with a key feature of responsible AI, in that it keeps human oversight in the process. While the AI helps generate answers, humans still ensure that the responses are accurate, ethical, and follow privacy rules.

This process has become more efficient and responsive. It has helped staff to manage the large number of questions they receive and has resulted in a 54% reduction in time spent responding to questions. This has enabled staff to allocate their time to higher-priority tasks.”

“The Central Statistics Office (CSO) is using AI to convert code written in one language (SAS) into another language (R). This saves time, reduces errors and ensures consistency.”

9.1.2 International Examples

“In Italy, the Corte dei Conti (Court of Auditors) uses a custom-AI model called GiusBERTo to automatically de-identify and anonymise court decisions without sacrificing any important information, a process previously done manually. This solution helps to balance the public’s right to access information with the need to protect the privacy of people. The anonymised documents are then subject to human review to ensure their accuracy and completeness” [8].

“In the United States, the US Patent and Trademark Office uses AI to enhance the processing of patent applications by assisting examiners in identifying relevant documents and suggesting additional areas of existing knowledge to search” [22].

“In Sweden, the Companies Registration Office developed an AI model that sorts approximatively 60% of incoming emails. The model reads their content, detects specific key phrases, and forwards it to the right recipient within the Office. In the case that an email does not contain one of the predefined key phrases, it reviews its entire content and makes an assessment based on employees’ previous behaviours” [8].

9.2 Service Delivery - Enhancing the delivery of public services by improving responsiveness

This function is associated with the 'Responsiveness' benefit discussed in section 3.2. According to the G7 report, "most of the use cases across the G7 impacting responsiveness are chatbots that facilitate access to information for citizens and empower public servants to provide faster and more accurate information in response to inquiries" [22]. Using chatbots to aid in the delivery of public services can reduce costs and boost efficiency in the public sector. They could also provide significant benefits to the public by reducing wait times and making some services available outside of traditional business hours.

9.2.1 Irish Examples

"St. Vincent's University Hospital (SVUH) is running a study to test if artificial intelligence (AI) can assist when performing high-quality heart ultrasound scans. This project, called the 'AI-Guided Echo Project,' tests whether AI can help capture accurate images in real-world medical settings.

Currently, most heart scans are performed by cardiac physiologists, but due to a shortage, the waiting time is often more than six months. To help with this issue, SVUH is testing Caption AI, a tool that uses AI to guide non-specialist healthcare workers, like nurses and clinical staff, in capturing high-quality heart images. The AI gives real-time instructions, helping them get the best possible images for cardiologists to analyse later.

By using Caption AI, SVUH hopes to improve access to heart scans, shorten wait times, and reduce pressure on specialist staff. This technology could make heart screenings more efficient, leading to faster diagnoses and earlier treatment, which could save lives and improve patient care."

ST. VINCENT'S UNIVERSITY HOSPITAL
Elm Park

"The Revenue Commissioners is using Large Language Models (LLMs) to route taxpayer queries more efficiently, ensuring faster and more accurate responses."

Revenue

"The Department of Justice has launched a Digital Contact Centre for Irish Immigration using Chatbots & Co-Pilot, improving response times and customer service."

An Roinn Dlí agus Cirt
Department of Justice

9.2.2 International Examples

“Austrian Digitalisation and E-Government Directorate of the Federal Ministry of Finance developed Mona, a conversational chatbot to provide information to entrepreneurs about business-related services and help them navigate the most relevant web content, increasing service quality and relieving civil servants from excessive workload. The system improves responsiveness in public services and performs principally interaction support tasks” [22].

“Canada’s Business Assistant Chatbot, part of the Canada Business App, is a mobile application to support small and medium business owners in navigating Government programs and services, while providing tailored recommendations and personalised notifications on funding” [22].

“In the United States, the Aidan Chat-bot is the Federal Student Aid’s virtual assistant that uses natural language processing to answer common financial aid questions and help customers get information about their federal aid on StudentAid.gov” [22].

9.3 Internal and External Oversight - Enhancing oversight and risk detection to facilitate more accountability

AI enhances accountability by identifying irregularities and flagging potential risks, thereby strengthening oversight mechanisms. This function is associated with the 'Accountability' benefit discussed in section 3.2. According to the G7 report, "AI can enhance Government accountability by improving the capacity, efficiency, and effectiveness of oversight, and supporting independent oversight institutions. By deploying algorithms to analyse massive volumes of data, AI can detect irregularities and potential fraud in processes that are traditionally vulnerable to errors and corruption" [22]. This function of AI can therefore make great savings in the public service and be of great value in highlighting potential cases of fraud and corruption.

9.3.1 Irish Examples

"The Department of Agriculture, Food, and the Marine (DAFM) processes 30,000 to 40,000 grant applications annually, which are submitted online on behalf of farmers, in line with the European Union (EU)'s General Data Protection Regulation (GDPR). However, errors in document submissions could lead to data breaches. As a result, sensitive data could be exposed to employees or even other farmers, creating compliance risks.

To mitigate the risk of major GDPR breaches, DAFM developed an intelligent solution that would help correctly identify personal sensitive information and automate the detection of breaches. SmartText, a machine learning (ML) text analysis platform categorises documents while protecting back-end systems. SmartText uses real-time artificial intelligence (AI) and ML capabilities to extract metadata and other contextual information from unstructured grant application documents, by analysis scans of handwritten or typed letters for sentiment, semantically similar words, topics, and entity names. When SmartText identifies a potential data privacy breach, the document is isolated and only authorised individuals can review the contents and redirect the document as appropriate.

As a result, DAFM have drastically reduced the number of breaches, as well as reduced application processing times from weeks to days. The solution has also helped DAFM significantly reduce manual administration and management."

An Roinn Talmhaíochta,
Bia agus Mara
Department of Agriculture,
Food and the Marine

The Revenue Commissioners are using Machine Learning to implement AI-powered fraud detection, which identifies suspicious transactions, thus improving tax compliance.

Revenue

9.3.2 International Examples

"Spain's Comptroller General has used AI to identify high-risk instances of potential fraud in grant and subsidies programmes" [22].

"In Estonia, the Tax and Customs Board (MTA) has been testing AI to identify incorrectly submitted VAT refund claims and to identify companies or persons in need for inspection" [8].

"Italy reports an AI use case for the detection of defects in banknote production" [22].

9.4 Policy-making – Improve the decision-making process

AI can support more effective decision-making in policy-making by analysing vast amounts of data to produce actionable insights. This function associated with the 'Productivity' benefit discussed in section 3.2. According to the G7 report, "more effective policy-making remains the less explored category of AI application among G7 members" [22]. This use of AI will not be applicable in all cases of policy development. However, there is potential in some cases where the use of AI could be of great benefit.

9.4.1 Irish Examples

"The Office of Public Works (OPW), like many public sector organisations, handles extensive and complex documents. Before GenAI, policy analysts had to read entire documents to understand their contents and impact, often having to analyse multiple documents at once. This process was slow and inefficient, especially when quick summaries were needed to allow for urgent decisions, ministerial briefs, or press releases.

The pilot system, called Policlear, analyses data and information and generates quick, concise summaries for policymakers in a matter of minutes. The system, which can provide summaries of multiple complex reports, was developed by a Dublin City University (DCU) campus company specialising in practical AI solutions for the public sector. Over the course of the pilot, new features were added, including a language translation tool and the ability to engage in interactive conversations with single or multiple documents. Recently, the system was enhanced to allow interaction with archived Oireachtas debates.

The system has significantly improved efficiency, workflow and decision-making, speeding up document summarisation and allowing interactive engagement (or "chat") with documents. As a result, staff have expressed an interest in exploring additional AI solutions for other business challenges."

OPW Oifig na nOibreacha Poiblí Office of Public Works

"The Central Statistics Office (CSO) is coordinating a four-year Eurostat project on AI/ML in official statistics, leveraging predictive analytics to enhance policy decisions."

An Phríomh-Oifig Staidrimh Central Statistics Office

9.3.2 International Examples

"Korea's Disease Control and Prevention Agency addresses situations of emerging infectious diseases. The system performs forecasting tasks by analysing medical data, quarantine data, and spatial data to develop policy responses to infectious diseases" [8].

"In France, the Paris-Saclay agglomeration of municipalities is using AI to simulate different energy management scenarios through a digital twin of their territory, allowing officials to more effectively evaluate the environmental and financial impacts of projects and improve long-term planning capabilities" [8].

Each of these examples illustrates the transformative impact AI can have on public services, from improving operational efficiency to supporting evidence-based policy-making. By exploring these use cases, the Irish Public Service can adopt best practices set out in these guidelines to responsibly harness AI's potential and provide high-quality, accessible services to the public.

Appendices

Appendix 1: Related national and EU Strategies

This document builds upon four complementary strategies / guidelines for Irish public sector workers. The key points from each that relate to this document are discussed below.

‘Better Public Services’, The Public Service Transformation 2030 Strategy

Better Public Services
A transformation strategy to deliver for the public and build trust

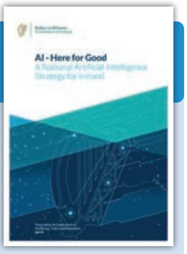
Published in 2023, the ‘Better Public Services – Public Service Transformation 2030 Strategy’ is a strategy with the ambition “to collaboratively deliver impactful outcomes” for the public and to build trust [5].

Trust is a constant theme throughout this paper and the two quotes below illustrate both the importance of maintaining trust and the unique position the public service can play.

- “Trust in Government and in public institutions is an emerging issue in many countries. Maintaining trust is crucial in ensuring the success of a wide range of public policies that depend on acceptance from the public” [5].
- “The public service is uniquely positioned to enhance trust in Government and public institutions and to ensure transparency and integrity in the conduct of public administration that will further increase trust” [5].

Within the Better Public Services Strategy, AI is positioned as a key technology in helping to solve complex issues in public service policy and delivery and to drive improvements. This is part of an overarching commitment to the innovative use of emerging technologies and digital transformation, ensuring a workplace and workforce fit for the future, and better use of evidence and data for policy-making.

'AI - Here For Good': The National Artificial Intelligence Strategy for Ireland



Originally published in 2021, and updated in 2024, Ireland's National Artificial Intelligence Strategy, "AI – Here for Good", provides a roadmap for the responsible, inclusive, and person-centred development and use of AI. It envisions Ireland as a global leader in AI, and emphasises its benefits to businesses, public services, and people. The strategy focuses on three key areas:

- Building public trust in AI by ensuring transparency, accountability, and compliance with the EU AI Act. This includes promoting public understanding and implementing standards for ethical AI use.
- Leveraging AI for economic and societal benefits by supporting AI adoption to enhance businesses and public services.
- Developing enablers for AI through support systems, skills development, investing in research, and ensuring access to quality data.

Strand 4 of the strategy focuses on leveraging the transformative potential of AI to deliver better public services. It sets out a strategic approach to integrating the responsible use of AI into public service delivery in accordance with the aims of Pillar 1 (Digital and Innovation at Scale) of the Public Service Transformation Strategy, 'Better Public Services'. For more information and to read the full document see the original text [23].

Enhancing the European Administrative Space (ComPAct)

Published in 2023, "Enhancing the European Administrative Space (ComPAct)" emphasises how "the public sector needs to be action-oriented, tackle emerging challenges, while strengthening public trust" [24]. ComPAct identifies the key qualities needed by public administrators as "high standards of integrity, transparency, accountability [24]". These are qualities that will help to deliver ethical AI solutions whilst helping to create more "seamless, secure and interoperable digital public services" [24].

Pillar 1 identifies how "Digital transformation also requires a substantial increase of the participation of civil servants in adult learning activities" [24] and how "reskilling and upskilling are massive tasks in public administration" [24]. Resources such as e-learning modules "will enable the direct access of all civil servants across the Member States and will also facilitate self-paced learning" [24] and help to provide the necessary upskilling opportunities.

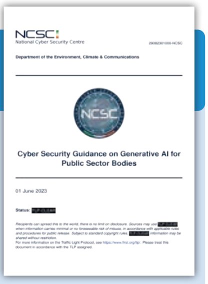
Pillar 2 builds on this by saying "further to digital upskilling and reskilling, public administrations need to embrace interoperability, leverage the increased availability of large amount of data, digitalise administrative procedures, and become AI-ready" [24]. The Commission has committed to supporting "public administrations in implementing digital and data-related legislation and increasing their readiness to integrate AI technologies into their operations in a safe and trustworthy way, supervising AI technologies, strengthening cybersecurity and designing and implementing public policies, including to support convergence of public procurement practices" [24].

One of the ways the Commission is encouraging member states to increase their digital readiness is through "technical support and participation in communities of practice" [24].

Public administrators should seek to avail of the supports and opportunities outlined in the ComPAct paper in getting themselves and colleagues "AI-ready" but follow the guidance of this report to ensure they do so in a responsible manner.

For more information and to read the full document see the original text [24].

National Cyber Security Centre - Cyber Security Guidance on Generative AI for Public Sector Bodies



NCSC
National Cyber Security Centre

29082301000-NCSC

Department of the Environment, Climate & Communications

Cyber Security Guidance on Generative AI for Public Sector Bodies

01 June 2023

Status: TLP:CLEAR

Recipients can spread this to the world, there is no limit on disclosure. Sources may use TLP:CLEAR when information carries minimal or no foreseeable risk of misuse, in accordance with applicable rules and procedures for public release. Subject to standard copyright rules, TLP:CLEAR information may be shared without restriction.
For more information on the Traffic Light Protocol, see https://www.first.org/tlp/. Please treat this document in accordance with the TLP assigned.

Published in 2023, the National Cyber Security Centre explains how “each Department will likely have a different view in terms of the business use case of GenAI tools and platforms, as well as the risk appetite for such use” [20]. However, their recommendation is “that access is restricted by default to GenAI tools and platforms and allowed only as an exception based on an appropriate approved business case and needs” [20]. They continued by saying that “it is also recommended that its use by any staff should not be permitted until such time as Departments have conducted the relevant risk assessments, have appropriate usage policies in place and staff awareness on safe usage has been implemented” [20].

The paper identified a number of key risks involved when using GenAI and provides guidance and some mitigations on managing them. Public sector workers should familiarise themselves with the guidance issued around these risks and the suggested do and don’ts lists.

There is an acknowledgement that with the appropriate measures in place to address the limitations, “there are clear benefits to productivity and effectiveness and GenAI will likely be built into many future product offerings” [20].

For more information and to read the full document see the original text [20].

Appendix 2: The EU AI Act – Key concepts and definition

EU AI Act overview as it relates to these guidelines

The EU AI Act became "the world's first comprehensive regulation on artificial intelligence" when it came into force on the 1st of August 2024 [13]. "The AI Act is designed to ensure that AI developed and used in the EU is trustworthy, with safeguards to protect people's fundamental rights. The regulation aims to establish a harmonised internal market for AI in the EU, encouraging the uptake of this technology and creating a supportive environment for innovation and investment" [13].

In the Irish Public Service, we want to lead the way with responsible AI solutions that go beyond the legal obligations of the EU AI Act. However, the EU AI Act establishes the minimum standards that we are legally obliged to comply with. Failure to meet these principles could lead to negative ramifications for the public service, such as hampering public trust and significant fines. Thus, it is imperative that Public Service Bodies and those working on AI solutions understand the EU AI Act and what is expected of them.

Breaking down the definition of AI according to the EU AI Act

Crucial to the definition of AI in the EU AI Act is the concept of inference, meaning that a model can reach a conclusion without being directly coded to do so. Some of the main algorithms considered are below:

1. Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning.
2. Logic and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference, and deductive engines, (symbolic) reasoning and expert systems.
3. Statistical approaches, Bayesian estimation, search, and optimisation methods.

Risk-Based Approach

The EU AI Act has approached the use of AI using a tiered compliance system with different principles for each tier. All public sector AI systems will need to be evaluated to determine their level of risk. As shown in the Figure 2, the EU AI Act implements a risk-based framework, which categorises AI systems into four risk levels:

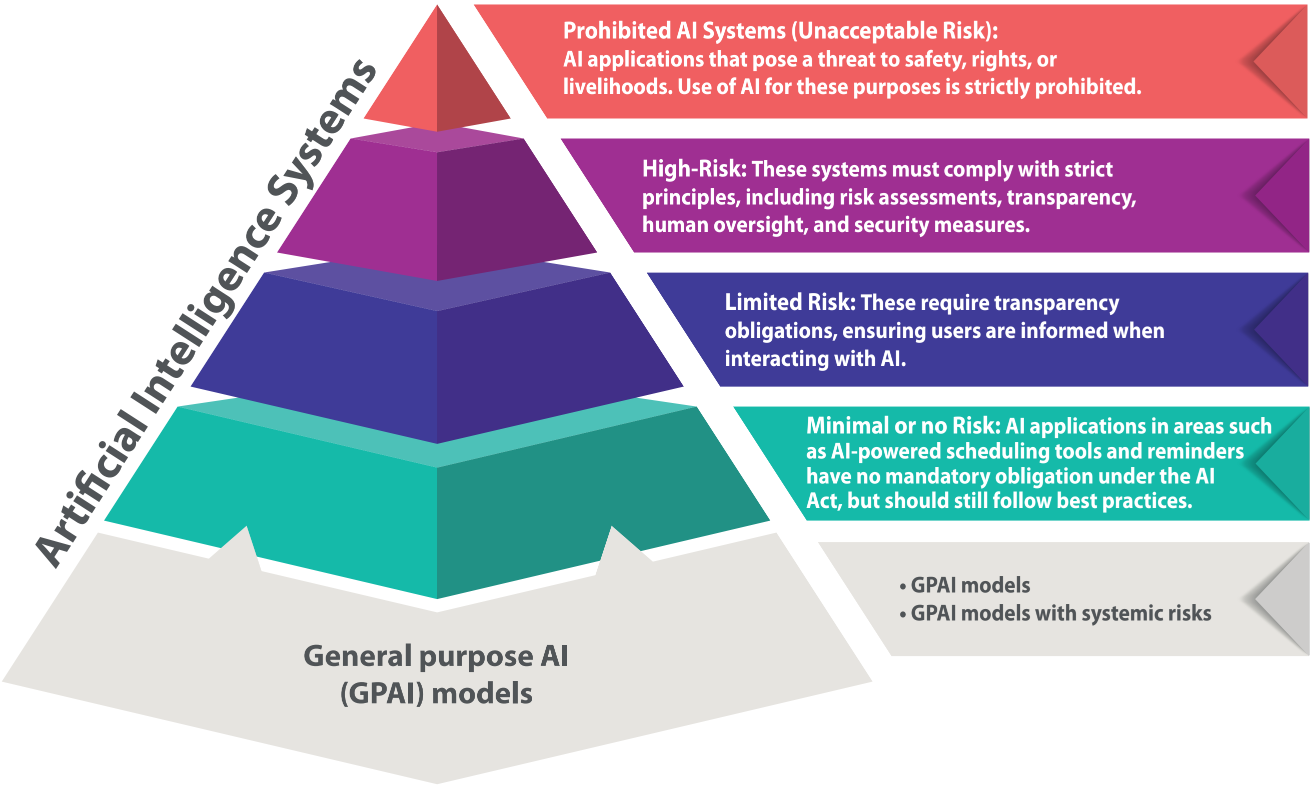

Artificial Intelligence Systems
Prohibited AI Systems (Unacceptable Risk): AI applications that pose a threat to safety, rights, or livelihoods. Use of AI for these purposes is strictly prohibited.
High-Risk: These systems must comply with strict principles, including risk assessments, transparency, human oversight, and security measures.
Limited Risk: These require transparency obligations, ensuring users are informed when interacting with AI.
Minimal or no Risk: AI applications in areas such as AI-powered scheduling tools and reminders have no mandatory obligation under the AI Act, but should still follow best practices.
• GPAI models
• GPAI models with systemic risks
General purpose AI (GPAI) models

Figure 2: *EU AI Act Risk-Based Approach*

A focus on prohibited practices (unacceptable risk)

The EU AI Act outlines specific practices that are prohibited due to their potential harm and the European Commission has published Guidelines on the prohibited AI practices as defined in the EU AI Act (2024). By way of example, these include: [30]

- "Cognitive behavioural manipulation of people or specific vulnerable groups: for example, voice-activated toys that encourage dangerous behaviour in children.
- Social scoring: classifying people based on behaviour, socio-economic status, or personal characteristics.
- Biometric identification and categorisation of people.
- Real-time and remote biometric identification systems, such as facial recognition."

Note: Some exceptions may be allowed, for law enforcement purposes.

A focus on high-risk AI systems

High-risk AI systems face stringent principles under the EU AI Act. However, this should not be an automatic deterrent. The AI systems could deliver significant benefits to the Irish Public Service or Irish society. To do this, we must ensure that proper safeguards are in place that respect the significant impact that this application has on people. Given the nature of public services, many of our AI prospective systems could be classified as high risk.

High-risk AI systems can be divided into two categories:

- Category 1 "AI systems that are used in products falling under the EU's product safety legislation. This includes toys, aviation, cars, medical devices, and lifts." [30]
- Category 2 "AI systems falling into specific areas that will have to be registered in an EU database:
  - Management and operation of critical infrastructure
  - Education and vocational training
  - Employment, worker management and access to self-employment
  - Access to and enjoyment of essential private services and public services and benefits
  - Law enforcement
  - Migration, asylum, and border control management
  - Assistance in legal interpretation and application of the law." [30]

High-risk AI systems are subject to strict obligations before they can be put on the market:

- Adequate risk assessment and mitigation systems
- High quality of the datasets feeding the system to minimise risks and discriminatory outcomes
- Logging of activity to ensure traceability of results
- Detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance
- Clear and adequate information to the deployer
- Appropriate human oversight measures to minimise risk
- High level of robustness, security, and accuracy" [12]

Appendix 3: Other Regulations relevant to these Guidelines

In addition to the EU AI Act, below are some of the most common regulations which may need to be considered for lawful AI. However, the list is not exhaustive. Direction is provided for each on where more information can be found.

General Data Protection Regulation (GDPR)

This regulation came in force in May 2016 and became applicable from May 2018 [25]. As outlined in Section 3.3, data privacy and protection are a key component in responsible AI systems and GDPR is an important regulation in determining our legal obligations. While GDPR does not explicitly mention AI, many of its provisions are relevant to AI applications [26]. It is imperative that accountable stakeholders ensure their AI system complies with this regulation.

Data Governance Act

This act came into force in June 2022 and has been applicable since September 2023. "The EU Data Governance Act provides a framework to enhance trust in voluntary data sharing for the benefit of businesses and citizens" [27]. The Act "aims to regulate the reuse of publicly/held, protected data, by boosting data sharing through the regulation of novel data intermediaries and by encouraging the sharing of data for altruistic purposes. Both personal and non-personal data are in scope of the DGA, and wherever personal data is concerned, the General Data Protection Regulation (GDPR) applies". Where this act could be applicable, accountable stakeholders must ensure they understand the regulation and what is expected of them.

Digital Services Act

This act came into force in November 2022. The Digital Services Act "regulates online intermediaries and platforms such as marketplaces, social networks, content-sharing platforms, app stores, and online travel and accommodation platforms. Its main goal is to prevent illegal and harmful activities online and the spread of disinformation" [28]. Where this act could be applicable, accountable stakeholders must ensure they understand the regulation and what is expected of them.

Digital Markets Act

This act also came into force in November 2022 and became applicable in May 2023. The act "establishes a set of clearly defined objective criteria to qualify a large online platform as a "gatekeeper" and ensures that they behave in a fair way online and leave room for contestability" [29]. Where this act could be applicable, accountable stakeholders must ensure they understand the regulation and what is expected of them.

AI Liability Directive

The purpose of the AI Liability Directive proposal is to improve the functioning of the internal market by setting uniform rules for certain aspects of non-contractual civil liability for damage caused with the involvement of AI systems. The proposal addresses the specific difficulties of proof linked with AI and ensures that justified claims are not hindered.

Further Resources

Online tools such as the ‘EU AI Act Risk Classifier & Compliance Checker’, developed by the European Commission, can be used. This can be a good initial basis for helping to determine how the EU AI Act is applicable to use cases. It will also assist in determining risk classification and provider and deployer obligations. See the European Commission webpage for more information and to try using the tool here [31].

The AI Pact is a European Commission voluntary initiative to help organisations prepare for compliance with the AI Act. [32]

Article 4 of the AI Act requires providers and deployers of AI systems to ensure a sufficient level of AI literacy to their staff and anyone using the systems on their behalf. Shaping Europe’s digital future is a living repository to foster learning and exchange on AI literacy. [33]

The European AI Office, which was established within the European Commission, plays a key role in implementing the AI Act. [34]

The European Commission has published guidelines on AI system definition to facilitate the first AI Act’s rules application. The guidelines explain the practical application of the legal concept, as anchored in the AI Act. [35]

The Commission has also published Guidelines on prohibited AI practices, as defined by the AI Act. [36]

References

[1] Department of Public Expenditure, NDP Delivery and Reform, “Government commits to using trustworthy AI in the Public Service,” gov.ie, 2 May 2024. [Online]. Available: https://www.gov.ie/en/press-release/a5c3e-Government-commits-to-using-trustworthy-ai-in-the-public-service/. [Accessed 15 October 2024].

[2] High-Level Expert Group on AI, “Ethics Guidelines for Trustworthy AI,” European Commission, 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. [Accessed 3 March 2025].

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence ,” OJ L, 2024/1689, 12 July 2024.

[4] Organisation for Economic Co-operation and Development (OECD), “OECD AI Principles Overview,”, May 2024. [Online]. Available: https://oecd.ai/en/ai-principles. [Accessed 8 October 2024].

[5] Department of Public Expenditure, NDP Delivery and Reform, “Better Public Services: A Transformation Strategy to Deliver for the Public and Build Trust,” Government of Ireland, 2023. [Online]. Available: https://www.gov.ie/en/campaigns/1cde2-better-public-services/?referrer=/. [Accessed 3 March 2025].

[6] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, OJ L, 2024/1689, 12 July 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng. [Accessed 3 March 2025].

[7] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, OJ L, 2024/1689, 12 July 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng. [Accessed 3 March 2025].

[8] B.-C. Ubaldi and R. Zapata, “Governing with Artificial Intelligence: Are Governments Ready?,” OECD Publishing, 2024. [Online]. Available: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/governing-with-artificial-intelligence_f0e316f5/26324bc2-en.pdf. [Accessed 3 March 2025]

[9] European Commission, “Adopt AI Study,” Publications Office of the European Union, Luxembourg, 2024. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/86e48333-8499-11ef-a67d-01aa75ed71a1/language-en?WT.mc_id=Selectedpublications&WT.ria_c=41957&WT.ria_f=7490&WT.ria_ev=search&WT.URL=https%3A%2F%2Fop.europa.eu%2Fen%2Fweb%2Fgeneral-publications%2Fai. [Accessed 3 March 2025].

[10] European Commission, “Commission publishes the Guidelines on prohibited artificial intelligence (AI) practices, as defined by the AI Act.,” 04 February 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act. [Accessed 28 February 2025].

[11] European Council, “Artificial intelligence act,” European Union, 14 October 2024. [Online]. Available: https://www.consilium.europa.eu/en/policies/artificial-intelligence/. [Accessed 21 February 2025].

[12] European Commission, “AI Act,” European Union, 14 October 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. [Accessed 21 January 2025].

[13] European Commission, “European Artificial Intelligence Act comes into force,” 1 August 2024. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4123. [Accessed 15 December 2024].

[14] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, OJ L, 2024/1689, 12 July 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng. [Accessed 3 March 2025].

[15] European Commission, “Ethics guidelines for trustworthy AI,” 08 April 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. [Accessed 04 November 2024].

[16] European Data Protection Supervisor, “Synthetic Data,” [Online]. Available: https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en. [Accessed 28 February 2025].

[17] Data Protection Commission, “The Legislative Consultative Process,” Data Protection Commission, [Online]. Available: https://www.dataprotection.ie/en/dpc-guidance/The-Legislative-Consultative-Process. [Accessed 28 February 2025].

[18] OECD, “Measuring the environmental impacts of artificial intelligence compute and applications,” 15 November 2022. [Online]. Available: https://www.oecd.org/en/publications/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications_7babf571-en.html. [Accessed 04 November 2024].

[19] Department of Enterprise, Trade and Employment, “AI - Here for Good: A National Artificial Intelligence Strategy for Ireland Refresh Executive Summary,” AI - Here for Good: National Artificial Intelligence Strategy for Ireland - DETE, 2024. [Online]. Available: https://enterprise.gov.ie/en/publications/national-ai-strategy.html. [Accessed 3 March 2025].

[20] Department of the Environment, Climate & Communications, “Cyber Security Guidance on Generative AI for Public Sector Bodies,” National Cyber Security Centre, 2023. [Online]. Available: https://www.ncsc.gov.ie/pdfs/Cybersecurity_Guidance_on_Generative_AI_for_PSBs.pdf. [Accessed 3 March 2025].

[21] HM Treasury, “Guidance on the Impact Evaluation of AI Interventions. Frontier Economics,” 2024. [Online]. Available: https://assets.publishing.service.gov.uk/media/672c84ebbd79990dfa67cab4/2024-11-05_Guidance_on_the_impact_evaluation_of_AI_interventions_FINAL_PDF_WITH_ACCESSIBILITY_CHANGES.pdf.

[22] OECD & UNESCO, “G7 Toolkit for Artificial Intelligence in the Public Sector,” 15 October 2024. [Online]. Available: https://www.oecd.org/en/publications/g7-toolkit-for-artificial-intelligence-in-the-public-sector_421c1244-en.html. [Accessed 20 January 2025].

[23] Department of Enterprise, Trade and Employment, ‘National AI Strategy Refresh 2024’. [Online]. Available: National AI Strategy Refresh 2024 - DETE. [Accessed 3 March 2025]

[24] European Commission, “Enhancing the European Administrative Space (ComPAct),” European Union, Luxembourg, 2023. [Online]. Available: https://commission.europa.eu/document/download/9ecd5276-df34-41ec-a328-ad51d1190300_en?filename=2023.4890%20HT0423966ENN-final.pdf. [Accessed 3 March 2025].

[25] European Data Protection Supervisor, “The History of the General Data Protection Regulation,” 05 November 2024. [Online]. Available: https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en. [Accessed 05 February 2025].

[26] European Parliamentary Research Service, “The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence,” European Parliament, 2020. [Online]. Available: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530. [Accessed 3 March 2025].

[27] European Commission, “Data Governance Act explained,” 11 October 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained. [Accessed 05 November 2024].

[28] European Commission, “The Digital Services Act,” 27 October 2022. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en. [Accessed 5 November 2024].

[29] European Commission, “The Digital Markets Act: ensuring fair and open digital markets,” 12 October 2022. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en. [Accessed 05 November 2024].

[30] European Parliament, “EU AI Act: first regulation on artificial intelligence,” 08 June 2023. [Online]. Available: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence. [Accessed 21 October 2024].

[31] European Commission, “EU AI Act Risk Classifier & Compliance Checker,” 20 June 2024. [Online]. Available: https://futurium.ec.europa.eu/en/european-ai-alliance/best-practices/eu-ai-act-risk-classifier-compliance-checker. [Accessed 24 October 2024].

[32] European Commission, “AI Pact,” [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/ai-pact. [Accessed 2 March 2025].

[33] European Commission, “Living repository to foster learning and exchange on AI literacy,” [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/living-repository-foster-learning-and-exchange-ai-literacy. [Accessed 2 March 2025].

[34] European Commission, “European AI Office,” [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/ai-office. [Accessed 2 March 2025].

[35] European Commission, “The Commission publishes guidelines on AI system definition to facilitate the first AI Act’s rules application,” [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application. [Accessed 2 March 2025].

[36] European Commission, “Commission publishes the Guidelines on prohibited artificial intelligence (AI) practices, as defined by the AI Act.,” [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act. [Accessed March 2025].

[37] Department of Public Expenditure, NDP Delivery and Reform, “Interim Guidelines for Use of AI in the Public Service,” Government of Ireland, 2024. [Online]. Available: https://www.gov.ie/en/publication/2127d-interim-guidelines-for-use-of-ai/. [Accessed 3 March 2025].

[38] OECD, “OECD updates AI Principles to stay abreast of rapid technological developments,” 03 May 2024. [Online]. Available: https://www.oecd.org/en/about/news/press-releases/2024/05/oecd-updates-ai-principles-to-stay-abreast-of-rapid-technological-developments.html. [Accessed 01 November 2024].

[39] European Commission, “Welcome to the ALTAI portal!,” [Online]. Available: https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal. [Accessed 24 October 2024].

[40] Department of Public Expenditure, NDP Delivery and Reform, “Value For Money Framework,” 21 December 2023. [Online]. Available: https://www.gov.ie/en/collection/c72a9-value-for-money-framework/#:~:text=All%20Irish%20Public%20Bodies%20are,is%20being%20spent%20or%20invested. [Accessed 25 October 2024].

[41] European Commission, “AI Act: Participate in the drawing-up of the first General-Purpose AI Code of Practice,” 30 July 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/news/ai-act-participate-drawing-first-general-purpose-ai-code-practice. [Accessed 01 November 2024].

[42] OECD, “AI principles,” 2024. [Online]. Available: https://www.oecd.org/en/topics/sub-issues/ai-principles.html. [Accessed 22 October 2024].

[43] European Commission, “European AI Office,” 04 November 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/ai-office. [Accessed 04 February 2025].

[44] European Commission, “Artificial Intelligence – Questions and Answers,” 01 August 2024. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683. [Accessed 04 November 2024].

[45] OECD, “Why we need a catalogue of tools and metrics for trustworthy AI,” 04 November 2024. [Online]. Available: https://oecd.ai/en/catalogue/tools. [Accessed 04 February 2025].

[46] National Institute of Standards and Technology, “NIST AI RMF Playbook,” 04 November 2024. [Online]. Available: https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook. [Accessed 04 November 2024].

[47] ODI, “Data assurance: Building trust in data,” 04 November 2024. [Online]. Available: https://theodi.org/insights/projects/data-assurance-building-trust-in-data/. [Accessed 04 January 2025].

[48] Massachusetts Institute of Technology, “What are the risks from Artificial Intelligence?,” 04 November 2024. [Online]. Available: https://airisk.mit.edu/. [Accessed 1 March 2025].

[49] OECD, “OECD AI Incidents Monitor (AIM),” 04 November 2024. [Online]. Available: https://oecd.ai/en/incidents. [Accessed 04 November 2024].

[50] United Nations, “Governing AI for Humanity,” September 2024. [Online]. Available: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf. [Accessed 04 November 2024].

[51] European Commission, “EU-U.S. Terminology and Taxonomy for Artificial Intelligence - Second Edition,” 05 April 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence-second-edition. [Accessed 04 March 2025].

[52] European Commission, “Shaping Europe’s digital future: AI Act,” 14 October 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. [Accessed 15 October 2024].

[53] European Parliament, “Artificial Intelligence and public service,” July 2021. [Online]. Available: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/662936/IPOL_BRI(2021)662936_EN.pdf. [Accessed 31 March 2025].

[54] Department of the Taoiseach, “National Risk Assessment 2024,” [Online]. Available: https://assets.gov.ie/305237/617914f2-e461-4bd9-a35a-6fdb670de27c.pdf. [Accessed February 2025].

Serbhísí Poiblí Níos Fearr
Better Public Services